

# **Learning From RNA sequencing Data The Role of Estrogen Receptor $\alpha$ In The Control Of Alternative Splicing In Breast Cancer**

Dissertation:

To obtain the academic degree

Doctor in Philosophy

PhD Programme in Complex Systems for Life Sciences

Submitted to the Department of Clinical and Biological Sciences of

University of Turin

By

Jamal ELHASNAOUI

2021

## Declaration:

This work has been carried out under the supervision of Prof. Michele De Bortoli at the Clinical and Biological Sciences department, University of Turin, Italy.

First reviewer:

Prof. Juan Valcárcel Juárez

Center for Genomic regulation- CRG: Barcelona, Catalunya, (SPAIN)

Email: [juan.valcarcel@crg.eu](mailto:juan.valcarcel@crg.eu)

Second reviewer:

Prof. Graziano Pesole

Dipartimento Di Bioscienze,

Biotechnologie E Biofarmaceutica

Professor of Molecular Biology,

Università degli Studi di Bari and National Research Council, Italy

Email: [graziano.pesole@uniba.it](mailto:graziano.pesole@uniba.it)

Date of defense: October, 04<sup>th</sup>, 2021

## **Acknowledgment:**

First of all, I want to thank Prof. Michele De Bortoli. Thank you so much for giving me the opportunity to perform my research in your laboratory. Thank you very much for giving me the freedom to grow as a scientist and the advice and impetus needed to succeed.

Secondly, many thanks to Dr. Giulio Ferrero for the insightful and extremely encouraging talks, advice and the unconditioned help he provided me throughout my thesis trajectory. Without your help Giulio, I don't know if some works of my thesis could have taken place.

My thanks also go to Dr. Valentina Miano for the insightful and encouraging comments she has always given me during my presentations and talks. Specifically, Valentina has always provided me with extremely important questions that when I look at them my results become more and more insightful.

My special thanks also goes to Prof. Graziano Pesole and Prof. Juan Valcárcel Juárez for their acceptance to revise my thesis.

I thank all the former and current members of De Bortoli's lab. We have had a fantastic working environment consisting of hard work, but without neglecting having relaxing time afterwards.

Lastly, my family has always given me enormous support, financially and emotionally. My great appreciation goes to my parents Fatima ELKAMILI and Lahcen ELHASNAOUI. My gratitude goes to my brothers and sisters for always being there for me, and finally to Hanaa Malloul Hanaa, the best thing that has ever happened to me. The best thing was the enormous amount of support and love you gave me throughout the ten years we have known each other. I will always remember this for the rest of my life. Your unlimited support is the reason that makes me stay up until this moment.

## **Abstract (English version):**

Breast cancer (BC) is a heterogeneous disease and the luminal Estrogen Receptor-alpha-positive (ER $\alpha$ +) tumor is the most frequent BC subtype. As for other complex phenotypes, the BC gene expression profiles are tightly regulated at both transcriptional and posttranscriptional levels. Notably, alternative splicing (AS) is an important molecular mechanism which regulates gene expression and governs many aspects of cellular proliferation, differentiation and development. The regulation of AS is driven by different RNA binding proteins (RBPs) mainly splicing factors (SFs), but emerging evidence revealed the essential role of noncoding RNAs as well as transcription factors in AS regulation.

In this thesis, using MCF-7 cells as cellular model of the luminal B subtype, I investigated through RNA-seq data analysis, the effect on both gene/isoform expression and AS in four different knockout (KO) experiments: i) the KO of the well-known SFs, hnRNPL and ESRP1/2; ii) KO of the ER $\alpha$ -controlled lncRNA DSCAM-AS1; and iii) the KO of ER $\alpha$  in hormone depleted cells. For this purpose, I developed a comprehensive bioinformatic analysis pipeline which starting from RNA-seq data is able to characterize the gene and isoform expression levels as well as detecting AS events, isoform switching, and predicting enriched RBPs binding motifs among AS events between two biological conditions.

The application of this bioinformatic pipeline on hnRNPL and ESRP1/2 KO experiments revealed distinct biological processes regulated by these SFs and a nonoverlapping set of AS events affected by each treatment. Interestingly, the hnRNPL KO extensively impairs the MCF-7 proliferation and induces a more mesenchymal phenotype in MCF-7 cells, while ESRP1/2 KO mainly affects the expression of immune-related genes with little effects on cell proliferation. Furthermore, in hnRNPL KO experiment, ER $\alpha$  was downregulated at both gene and isoform levels and DSCAM-AS1 was the most significant gene characterized by an isoform switching event, supporting its known interaction with this SF.

The KO of the ER $\alpha$ -controlled lncRNA DSCAM-AS1 in BC was also investigated. As for hnRNPL KO, DSCAM-AS1 silencing strongly hampered the expression of cell cycle related genes and induced an extensive change in AS, enriched in 3'UTR shortening and exon skipping events. By an RBP binding motif enrichment approach hnRNPL was the most enriched SF for DSCAM-AS1-regulated exon skipping and 3'UTR events. This thesis reveals the complexity in regulation of AS by a network of SFs, lncRNA, and ER $\alpha$  in luminal BC and provides a comprehensive computational approach to analyse these molecular events from RNA-seq data.

# Table of Contents

chapter 1. Introduction .....	1
1.1 The central dogma of molecular biology and its fails face to the human genome complexity .....	1
1.2 Post-transcriptional control of gene expression .....	5
1.2.1 Constitutive pre-mRNA splicing .....	6
1.2.2 Alternative splicing of pre-mRNA.....	8
1.2.3 non-coding RNAs as regulators of gene expression at transcriptional and posttranscriptional levels .....	13
1.2.4 Alternative splicing process as a regulator of development, differentiation, and tissue identity definition.....	22
1.3 High throughput techniques as tool for alternative splicing characterization on a genome wide scale	27
1.3.1 Overview of a typical RNA-seq data analysis pipeline .....	28
1.3.2 Bioinformatic approaches to exploring and quantifying alternative splicing .....	29
1.4 Breast cancer and Estrogen Receptors .....	32
1.4.1 ER-alpha signaling pathway in mammary gland morphogenesis and development .....	32
1.4.2 ER-alpha signaling pathway in breast cancer development and progression .....	34
1.4.3 Estrogen receptors and their transcriptional activities.....	36
1.5 Aims of the Thesis:.....	54
chapter 2. Results .....	55
2.1 The functional investigation of the hormone-independent Estrogen Receptor alpha signaling pathway on transcriptional and post-transcriptional levels .....	56
2.1.1 The hormone independent activity of ER $\alpha$ is crucial for cell proliferation and for maintaining the epithelial phenotype.....	57
2.1.2 ER $\alpha$ depletion in MCF-7 BC cells induces a differential expression of gene isoforms resulting in differential functional consequences .....	62
2.1.3 ER $\alpha$ depletion in MCF-7 induces internal alternative splicing events .....	67
2.1.4 Discussion.....	80
2.2 The epithelial ESRP1 and ESRP2 factors control AS in ER $\alpha$ + BCs and their expression is tightly controlled by hormone-independent activity of ER $\alpha$ .....	87
2.2.1 ESRP1 and ESRP2 expression is altered in ER $\alpha$ + BCs and is regulated by ER $\alpha$ .....	87

2.2.2 ESRP1 and ESRP2 regulate metabolism, cell proliferation, in addition to EMT pathways in MCF-7 cells .....	90
2.2.3 ESRP1 and ESRP2 splicing factors control a set of ASEs in ER $\alpha$ + BCs .....	105
2.2.4 A set of ESRP1/2 regulated ASEs occurs also upon silencing ER $\alpha$ in hormone-deprived MCF-7 cells.....	109
2.2.5 Discussion:.....	114
2.3 DSCAM-AS1-driven proliferation of breast cancer cells involves regulation of alternative exon splicing and 3'UTR usage. ....	124
2.4 Deciphering the role of the splicing factor hnRNPL in the control of luminal epithelial BC transcriptome .....	125
2.4.1 HNRNPL knock-down strongly hampers the proliferation of MCF-7 BC and induces a mesenchymal-like phenotype .....	126
2.4.2 HNRNPL sustains cell proliferation of MCF-7 BC cells by stabilizing the expression of DSCAM-AS1 and ESR1 isoforms .....	130
2.4.3 Analysis of alternative splicing changes induced by <i>HNRNPL</i> silencing reveals important functions of hnRNPL in MCF-7 BC cells .....	132
2.4.4 hnRNPL, ER $\alpha$ , and DSCAM-AS1 are three factors synergistically acting on proliferation in ER $\alpha$ + BCs .....	140
2.4.5 Discussion:.....	142
chapter 3. General conclusions .....	144
chapter 4. Materials and Methods.....	146
4.1 Computational methods: .....	146
4.1.1 Analysis of ESRP1, ESRP2, ESR1, and HNRNPL expression in TCGA clinical data .....	146
4.1.2 Overlap with ER $\alpha$ ChIP-Seq data .....	146
4.1.3 RNA-sequencing dataset analysis .....	147

---

## chapter 1. Introduction

### Chapter organization

Gene expression (GE) is an integrated process that involves complex regulation at *genomic*, *epigenomic*, and *transcriptomic* levels. In this context, a specific layer of complexity is further introduced by the generation of different transcript isoforms due to co-transcriptional modification events. As a mechanism enabling the synthesis of different RNA isoforms from the same gene locus, alternative splicing (AS) is considered the most contributing mechanism to the complexity and diversity of the proteome content of the cell. Importantly, AS was found to be mis-regulated under disease states, especially in cancer, where this alteration might be evaluated as a hallmark of cancer cell biology.

In this introduction, a general overview on the human genome complexity and its integrative regulatory mechanisms is provided, with a special focus on the mechanisms of AS and its impact on GE under both physiological and pathological conditions.

The first section of this introduction is focused on the posttranscriptional regulation of GE process, followed by a detailed overview of the AS process including its mechanisms, functional consequences, regulation by RNA-binding proteins (RBPs) or splicing factors (SFs) as well as by the non-protein coding RNAs (ncRNAs). Finally, its role under physiological conditions and its aberrant regulation under cancer context is described.

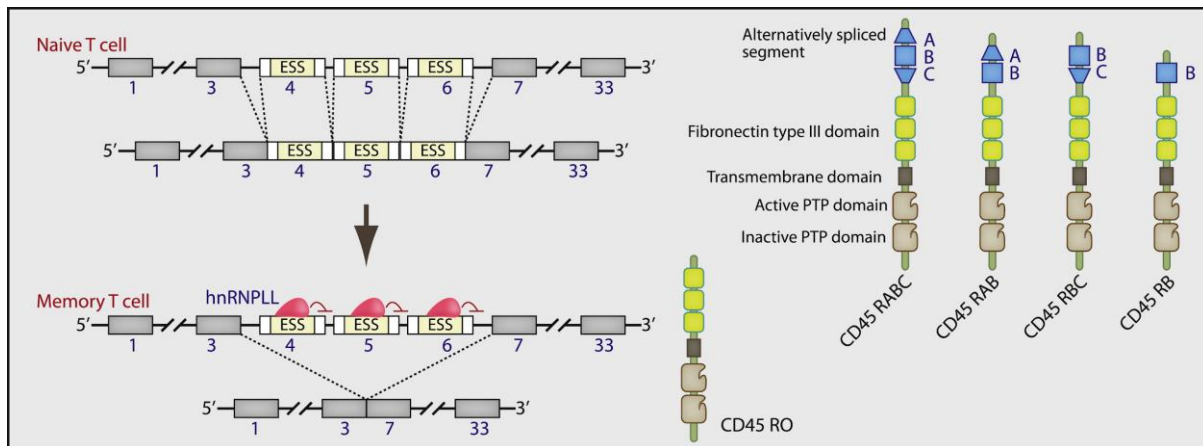
The second section of this introduction provides an overview on the main experimental methods, especially High-throughput RNA sequencing techniques (RNA-seq) and debate on how the continuous reduction of its cost has enabled the generation of high amounts of data that could be analysed for AS regulation. In addition, this section also provides a focus on the different computational approaches for RNA-seq data analysis which can help in deciphering/constructing the pattern of splicing events in cells.

Finally, the last section highlights the role of the estrogen receptor-alpha ( $ER\alpha$ ) signaling pathway in breast cancer (BC) as a model of complex network regulation of epithelial gene expression. More specifically, this section provides evidence on how  $ER\alpha$  pathways control the expression of specific RBPs and ncRNAs as main regulators of AS.

### 1.1 The central dogma of molecular biology and its fails face to the human genome complexity

Over decades, the sequence information in a cell was thought to follow a linear flow from genomic (DNA) to transcriptomic (RNA) to proteomic (protein) layers through the process of transcription and translation, respectively (Hartman 1959). However, the completion of the human genome project (ENCODE Project Consortium 2012) revealed that the human genome produces around 90,000 different proteins, while the number of human genes discovered was fewer than 25,000 genes, even lesser than the number of genes present in less complex genomes (e.g. 40,000 genes in *Zea mays*) (Valdivia 2007). This non-proportionality between the genome complexity, the number of discovered genes, and the number of encoded proteins was mainly attributed to the mechanisms of AS (**Figure 1.1-1**), and it was acknowledged as the mismatch in the “one gene to one protein formation” paradigm of the central dogma of molecular biology. Thus, the gene-centric view

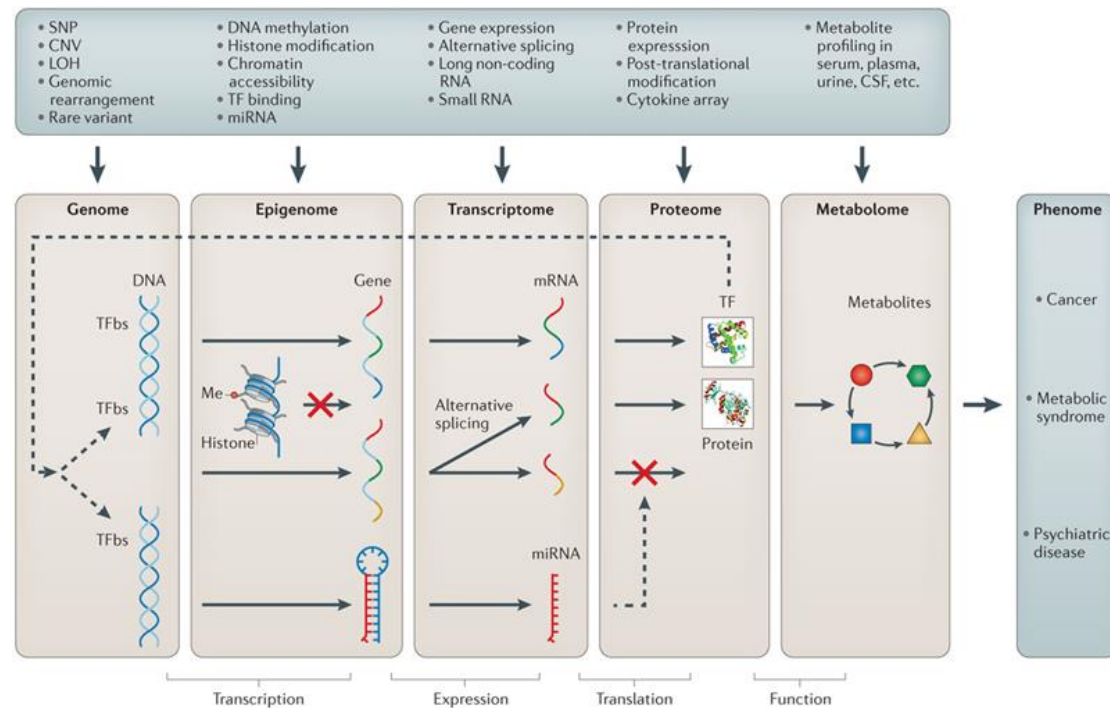
introduced by the central dogma of molecular biology was then updated from genes to RNA transcripts, these latter which were considered as fundamental units of the genome (Pennisi 2012).



**Figure 1.1-1:** An example of the human CD45 gene with complex AS patterns and its downstream consequences. The differential inclusion of exons 4, 5, and 6 is context-dependent and results in different protein isoforms, with different functional domains. In human naive T cells, the inclusion of exons 4, 5, and 6 is increased resulting in protein isoforms with glycosylated segments A, B and C. While in human memory T cells, the concentration of hnRNPLL is increased resulting in the inhibition of exons 4, 5, and 6 inclusion and production of the shortest CD45 isoform (CD45RO) lacking the glycosylated segments. ESS: Exonic splicing silencers. From (Zikherman and Weiss 2008).

Since the realization of the ENCODE project, it has become increasingly evident that the complexity of an organism genome does not only depend on its genome content, but also on how the production and the activity of genome products is regulated. Subsequently, it has now become clear that GE is a complex integrative process that is tightly regulated by multiple layers of regulation forming complex regulatory networks, contrary to the linear flow of information initially proposed by the central dogma of molecular biology. For instance, the expression of a given gene to produce a functional RNA or protein product implies regulations at *genomic*, *epigenomic*, and *transcriptomic* levels (Ritchie et al. 2015) (**Figure 1.1-2**).





**Figure 1.1-2:** A schematic overview of the GE process as an integrative process involving different layers of regulation. The arrows indicate the flow of genetic information from the genome to the proteome level and, consequently, to the level of cell phenotype (phenome). The red crosses indicate repression of a transcriptional or translational process. Examples of the molecular regulatory elements involved in each regulatory layer are given. SNP, single-nucleotide polymorphism; CNV, copy number variation; LOH, loss of heterozygosity; Me, methylation; TF, transcription factor; miRNA, microRNA; TFbs, transcription factor-binding sites. From *Nature Reviews Genetics* (Ritchie et al. 2015).

The *genome*, representing the first layer of regulation, harbours the constitutive sequence and structural information of a given organism, in addition to any kind of sequence variations; including sequence nucleotide variants, single nucleotide polymorphisms, or copy number variations, amplifications, insertions, and deletions; all of which influence the GE process (Bernstein, Meissner, and Lander 2007).

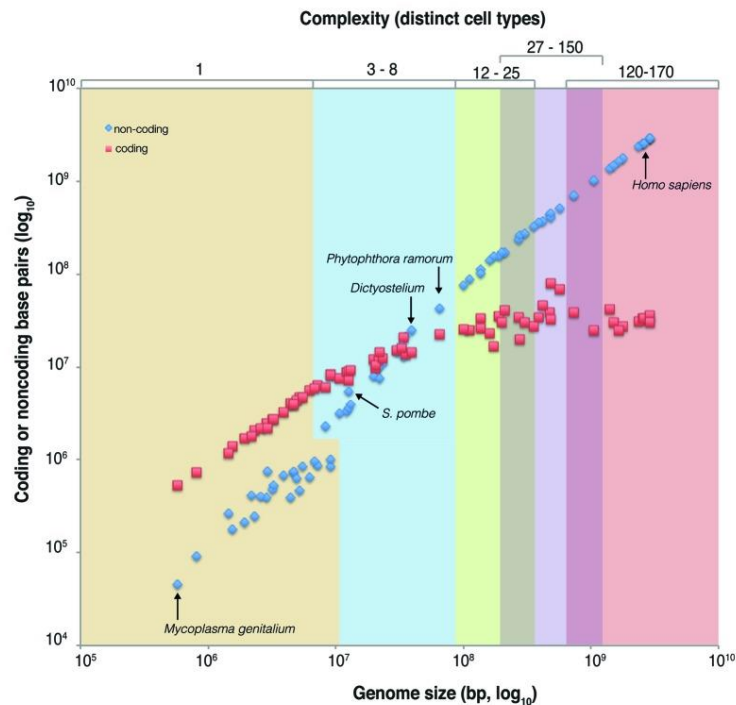
The second layer of regulation of the GE process is represented by the *epigenomic* content of the cell, which includes DNA methylation states and all possible histone Post-Translational Modifications (PTMs) along the chromatin (Rivera and Ren 2013). The *epigenome* regulates GE by organizing the nuclear architecture of chromosomes, either facilitating or repressing the accessibility of transcription factors (TFs) to DNA. The *epigenome* differs between tissues and cell types thus it constitutes a second layer of regulation that is pivotal to maintain cell type-specific GE patterns and ultimately cell-specific phenotypes.

The *transcriptome* which constitutes the whole set of molecular interactions involving cellular RNAs and their processing mechanisms, particularly by AS, constitutes the third layer of GE regulation. It has become recently clear that the number of functionally characterized regulatory RNAs is increasing (E. Park et al. 2018) and their dual ability to bind proteins or to mediate base-pairing with DNA or RNA molecules is being investigated (Statello et al. 2021).

Furthermore, the ENCODE project has estimated that, on average, each human gene has approximately 6.5 different RNA transcripts (ENCODE Project Consortium 2012). It is therefore not surprising that the regulation of RNA transcript expression plays an important role in many important biological processes (Kornblihtt et al. 2013). Vital biological functions such as development, differentiation, proliferation, and apoptosis have indeed been shown to be regulated by the expression of specific gene products (Gabut et al. 2011), and their misregulation has been associated with the development of several diseases, including cancer (Y. I. Li et al. 2016). Moreover, accumulating evidence has revealed that RNA isoforms are expressed in a tissue-specific manner and may alter several functional aspects, such as cellular localization, and stability of the RNA or protein molecules (Jiang et al. 2020). Notably, it was experimentally shown that the splicing of the pre-mRNA into different alternative RNA transcripts could result in the production of protein isoforms with remarkable differences in several functional aspects including cellular localization, stability, DNA binding properties, and could even behave as distinct proteins especially when considering their protein-protein interaction (PPI) networks (X. Yang et al. 2016).

AS, differential alternative promoter or transcription start site (TSS) usage, and alternative polyadenylation (polyA) site usage are the three main molecular mechanisms that are involved in the synthesis of different splice variants from the same gene (de Klerk and 't Hoen 2015). In highly complex organisms, like mammals, it has been estimated that at least 70% of genes have multiple alternative polyA sites, more than 50% have alternative TSSs, and up to 95% of genes undergo AS (Carninci et al. 2006). Hence, these three independent mechanisms by affecting the structure and the functional properties of RNA transcripts increase the repertoire and complexity of the genome of a given organism (Ni et al. 2013).

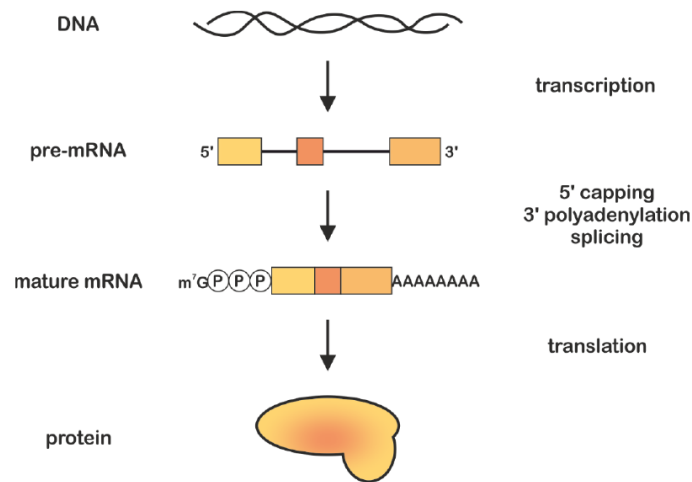
It is now well known that 98% of the human genome does not encode proteins (ENCODE Project Consortium 2012) (**Figure 1.1-3**). These non-protein coding regions of the human genome were initially considered as just “junk DNA” and were thought to undergo no selective pressure but needed only to accumulate mutations with no harm effects to the organism (Ohno 1972). However, The efforts of the ENCODE project provided evidence supporting that this junk DNA is indeed transcribed into RNA and it could, as RNA, participate in the control of the GE process by forming complex regulatory networks (ENCODE Project Consortium 2012). Notably, ncRNAs transcripts have been shown to be functionally important in different contexts where they function as key regulatory molecules of GE either at *epigenetic*, *transcriptional*, and *post-transcriptional* levels (Cech and Steitz 2014). Moreover, ncRNAs transcripts have also been found to be involved in physiological as well as in pathological conditions and were identified to act as tumor suppressors or oncogene in various cancers (Pavet et al. 2011). The discovery that most (98%) of our transcriptome is non-coding has become very exciting for researchers since understanding the different roles played by these ncRNAs would represent a great fundamental achievement in molecular biology.



**Figure 1.1-3:** In this scatterplot, 73 organisms with a previously defined relative biological complexity (expressed as the number of distinct cell types) are shown as pairs of data points, with total protein-coding sequence bases depicted in red and the total non-protein-coding bases depicted in blue which cumulatively give the total genome size (x-axis). Non-protein-coding sequence increases exponentially with the number of distinct cell types, while protein-coding sequence is asymptotic. From (G. Liu, Mattick, and Taft 2013).

## 1.2 Post-transcriptional control of gene expression

In eukaryotic cells, the decoding of a gene into functional RNA and protein products is generally achieved through the processes of transcription and translation, respectively (Society for Experimental Biology (Great Britain) 1958). The gene is first transcribed into a precursor mRNA (pre-mRNA), the latter which is translated into a protein. To be translated, the pre-mRNA molecule undergoes a preprocessing into mature mRNA molecule through three main stages: 5' capping, by addition of a 7-methylguanosine residue to the 5' terminal phosphate, polyadenylation, by addition of a poly-adenosine (poly-A) tail to the 3' end, and constitutive pre-mRNA splicing (**Figure 1.2-1**). The 5' capping of pre-mRNA is crucial for maintaining mRNA stability, nuclear export, and for translation initiation (Wilusz, Wormington, and Peltz 2001). The polyadenylation step of pre-mRNA constitutes another layer of regulation as it regulates the interaction with RBPs and microRNAs (miRNAs) (Tian and Manley 2017). The third step of constitutive splicing consists of a multi-step pathway aiming at the excision and removal of non-coding intervening sequences (introns) from the pre-mRNA and joining of exons together through the interaction with the spliceosomal complex (further discussed in the next sub-chapter).



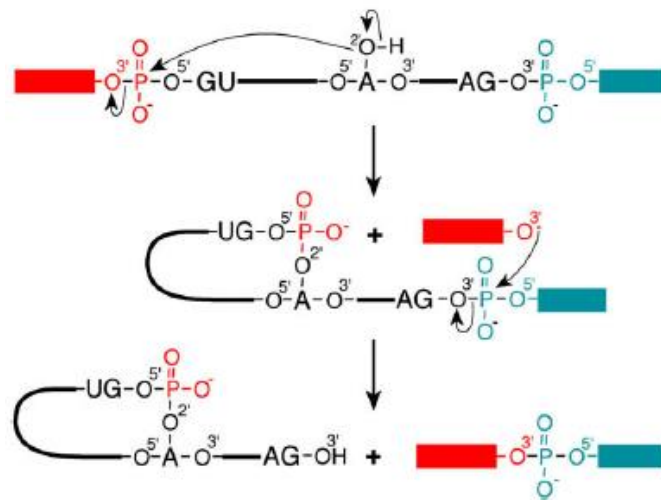
**Figure 1.2-1:** From DNA to protein formation as introduced by the central dogma of molecular biology. The DNA (double helix loop) is transcribed into pre-mRNA, consisting of a set of exons (boxes) intervened by non-coding introns (solid lines). The pre-mRNA is 5' capped, polyadenylated and constitutively spliced to give a mature mRNA molecule that is further translated into protein. m7G: 7-Methylguanosine, circled P: phosphate, A: adenosine. From (Neumann 2019).

### 1.2.1 Constitutive pre-mRNA splicing

Most eukaryotic unspliced pre-mRNA molecules contain one or more (sometimes dozens of) intervening non-coding sequences called introns of varying lengths that must be removed to ensure the proper placement of short sequences called exons (~20-300 nucleotides length) in the correct reading frame (Ast 2004). Since their discovery, a number of functions have been attributed to introns such as facilitating the apparition of novel genes by exon shuffling/duplication and production of diverse protein isoforms from the same gene through AS. However, it has recently become more clear that introns and their elimination by the spliceosomal machinery can influence different stages of RNA lifespan including initiation of transcription (Kwek et al. 2002), editing, polyadenylation (Proudfoot, Furger, and Dye 2002), export from nucleus (M.-J. Luo and Reed 1999), translation (Matsumoto, Wassarman, and Wolffe 1998) and mRNA degradation (Hir and Le Hir 2001). Taken together, these introns-induced effects can sum up and contribute to different expression profiles of intron-containing and intronless genes (Rose 2018).

During the process of pre-mRNA splicing, the splicing machinery recognises the exons with high precision, removes introns, and ligates the exons to form a mature RNA molecule. The efficiency and precision of introns removal from the pre-mRNA is ensured by trans-acting factors which includes the component of the spliceosomal complex; small nuclear ribonucleoprotein complexes (U1, U2, U4, U5, and U6 snRNPs) and other non-snRNP proteins; and by cis-acting sequence elements which includes 5' and 3' splice sites that mark exon-intron boundaries, a branching point, and a polypyrimidine tract that lie within the intron (Brow 2002). Additionally, other exonic and intronic splicing enhancer and suppressor signals constitute another control layer of pre-mRNA splicing through their interaction with the trans-acting factors (Sanford, Ellis, and Cáceres 2005).

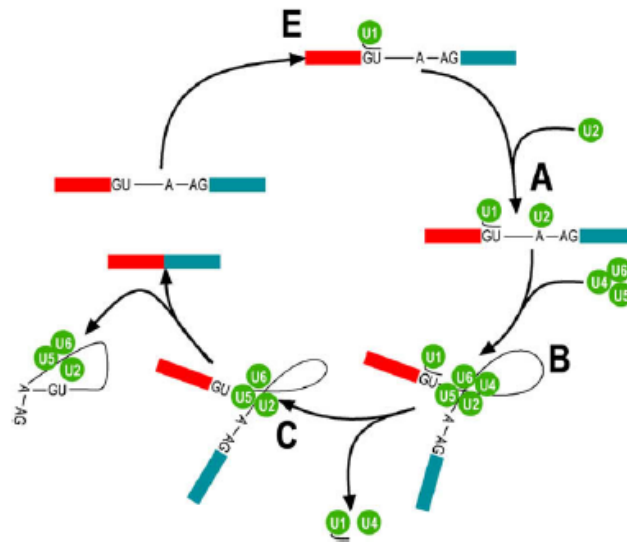
The intron-exon junction boundaries of the pre-mRNA are marked by the presence of specific splice sites that are recognisable by the spliceosomal machinery (Romfo et al. 2000). At the branch point, a two-step nucleophilic transesterification reaction takes place, where the 2' hydroxyl group of a specific adenosine becomes more nucleophilic and makes a nucleophilic substitution type 2 (N2) on the phosphodiester at the upstream 5' splice site, forming a free upstream 5' exon and a 3' exon-intron lariat (Sperling, Azubel, and Sperling 2008). Next, a second N2 reaction takes place where this time the free 3' hydroxyl anion of the upstream exon attacks the phosphodiester moiety at the 3' splice site to yield the spliced RNA and the spliced-out intron in a lariat form (Sperling, Azubel, and Sperling 2008) (**Figure 1.2-2**).



**Figure 1.2-2:** Constitutive splicing mechanisms of pre-mRNA molecules. First step: The 2' hydroxyl group of a specific adenosine at the branch site of a pre-mRNA makes a nucleophilic substitution type 2 (SN2) attack on the phosphodiester at the 5' splice site (marked in red), resulting in a free 5' exon (red) and a 3' exon-intron moiety. Second step: the free 3' hydroxyl anion of the 5' exon attacks the phosphodiester moiety at the 3' splice site (marked in blue), yielding the spliced RNA and the spliced-out intron. From (Sperling, Azubel, and Sperling 2008).

In addition, the pre-mRNA splicing process is mediated by a large poly-component complex called spliceosome. This complex is composed of the five small nuclear RNA (snRNA) molecules (U1, U2, U4, U5 and U6 snRNA) and as many as 250 proteins (J. Liu et al. 2020). Each of the five snRNA molecules interacts with a specific protein to form a small ribonuclear protein complex (snRNP). The coordinated binding of the five snRNPs within the splicing signals in the pre-mRNA occurs in a stepwise manner and results in the removal of introns and the ligation of the flanking exons (Ast 2004). The process is first initiated by the recognition and binding of the U1 snRNP at the 5' splice site forming the E complex, the recognition and binding of the U2 snRNP at the branch site in an ATP-dependent manner giving rise to A complex. Thereafter, a further step involves the addition of the U4/U6.U5 tri-snRNPs complex to form an inactive B complex. This B complex is later activated during the next step by the release of the U1 and U4 snRNPs to yield the splicing-active complex C. The active complex C catalyses a second reaction which results in the joining of the flanking exons and in the release and removal of the intron lariat as well as the U2, U5, and U6 snRNPs (Sperling,

Azubel, and Sperling 2008) (**Figure 1.2-3**). The whole process is accompanied by the interaction with other trans-acting SFs (discussed in detail in the next sub-chapters).



**Figure 1.2-3:** Schematic representation of the spliceosomal machinery assembly. U1 (green) binds to the 5' splice site of the pre-mRNA molecule, yielding complex E. Complex U2 snRNP interacts with the branch site in an ATP-dependent manner, yielding complex A. Complex B is formed by addition of the U4/U6.U5 tri-component. The U1 and U4 snRNPs are then released and the complex B is remodelled, yielding the splicing-active complex C. From (Sperling, Azubel, and Sperling 2008).

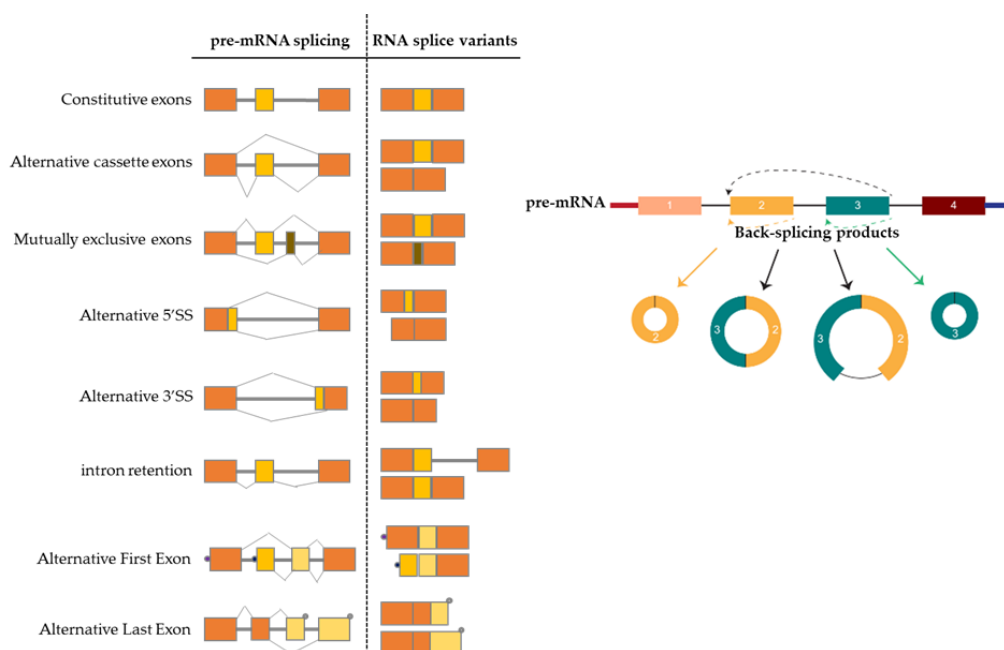
## 1.2.2 Alternative splicing of pre-mRNA

Most human genes are multi-exonic and could be considered as an island of exons placed in the middle of oceans of introns. An average human gene is in general 28,000 nucleotides long and consists of 8.8 exons of about 120 nucleotides on average that are separated by 7.8 introns of varying lengths (Consortium and International Human Genome Sequencing Consortium 2001). The relatively high number of exons per gene enables the spliceosomal machinery to splice-in different sets of exons from a single pre-mRNA generating different RNA transcripts with different exon composition and properties (Modrek and Lee 2002). Notably, the advent of transcriptomic sequencing technologies together with the development of bioinformatic tools revealed that RNA transcripts of nearly 95% of human multi-exonic genes are alternatively spliced into multiple RNA transcripts (Merkin et al. 2012). Hence, the AS process is a source of human proteome diversity and complexity (Woodley and Valcárcel 2002).

### 1.2.2.1 Modes and mechanisms of alternative splicing

The AS process in its linear form can be subdivided into five different major categories: (i), the usage of an alternative 5' or (ii) 3' splice site (A5'/A3'), resulting in the shortening of an exon; (iii), skipping of an entire cassette exon (exon skipping, ES); (iv), skipping of one of two consecutive exons and simultaneous inclusion of the other (mutually exclusive exon, MX); (v), and retention of an intron (intron retention, IR) (Sugnet et al 2004). In addition, advanced high-throughput sequencing techniques such as RNA-seq with the help of Cap

Analysis Gene Expression (CAGE) tags and full length RNA-seq projects have brought up two other transcript variations - not considered as AS events in *sensu stricto* - involving the differential usage of either alternative first (AF) exon; possibly resulting from TSSs differential usage; or alternative last (AL) exon; which could result in polyA sites differential usage (Trincado et al. 2018) (**Figure 1.2-4**). More recently, another mechanism of RNA transcripts production called back-splicing involving the circularization of one, two, or more exons generating circular RNA transcripts (circRNAs) was also discovered in both normal human tissues as well as in different cancer types (P. Zhang et al. 2020; Tarrero et al. 2018) (**Figure 1.2-4**). Despite their low expression levels in comparison to their linear counterparts, circRNAs represent the long living forms of RNA transcripts as they are extremely resistant to degradation by exonucleases due to the absence of free ends and their prognostic potential in different diseases is being recently investigated (Z. Zhao et al. 2017).

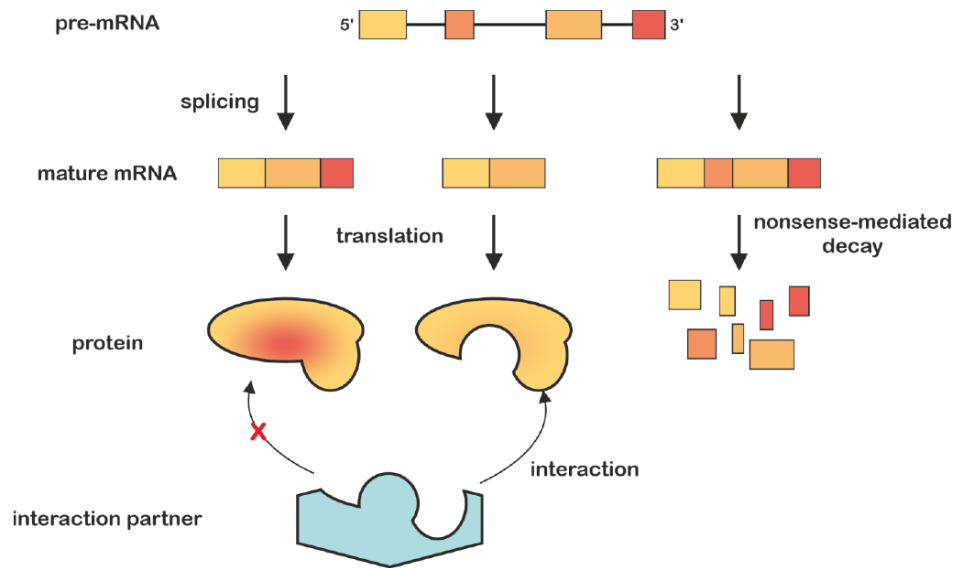


**Figure 1.2-4:** Types, modes, and resulting splice RNA products of the AS process. Shown on the left are constitutive splicing and the five main AS categories, in addition to AF and AL exon events and their resulting mRNA splice variants (right). constitutive and alternative exons are shown in dark and light colored boxes, respectively, and introns are represented by solid lines. In AF and AL splicing events, the dots before and after alternative exons represent different promoters and polyA sites, respectively. On the right side of the figure is shown the possible circular RNA transcripts (circRNAs) produced from a given pre-mRNA. The back-splice junctions involved in exon(s) circularization are highlighted by dashed arrows. The back-splicing could generate mono-exonic (exon 2 in yellow, exon 3 in green), or di-exonic circRNAs with or without intermediate intron retention.

### 1.2.2.2 Functional consequences of alternative splicing

AS produces different mRNA isoforms from the same gene locus and remarkably increases the proteome complexity and diversity in the cell (Yansheng Liu et al. 2017). Although genetically related and originate from the same gene locus, the different isoforms generated through AS may have different RNA stability, decay rate, and transcription efficiency (X. Yang et al. 2016). Furthermore, the different mRNA isoforms generated by AS of a single pre-mRNA molecule may differ by their exonic composition and may encode for

different protein isoforms. These protein isoforms have been shown on many occasions to act as distinct proteins, to exhibit different interaction profiles (M.-S. Kim et al. 2014), to have different localizations, and to be involved in different biological pathways, and in some cases to ensure completely opposite biological functions (X. Yang et al. 2016) (**Figure 1.2-5**).



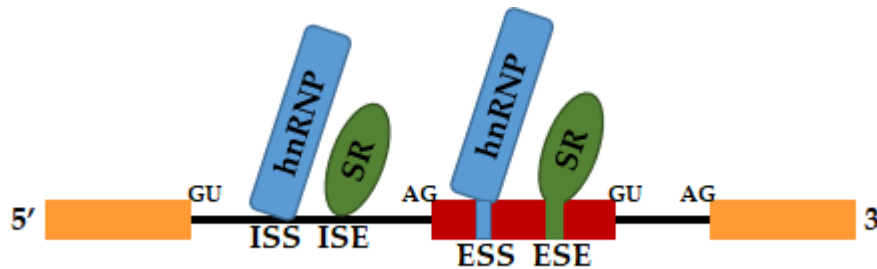
**Figure 1.2-5:** Possible functional consequences of AS. AS could result in different mRNA splice variants with different stabilities (NMD-sensitive versus NMD-insensitive). NMD, Nonsense mediated decay. AS could also affect the domains involved in interactions with specific partners of the encoded protein isoforms. From (Neumann 2019).

### 1.2.2.3 Regulation of alternative splicing

AS is an important molecular mechanism controlling GE and occurs in a tissue or context-dependent manner, thus representing a fine-tuning device of the cell fate and cellular decision-making process. It allows the fine-tuned production of specific RNA transcripts in a context-dependent manner (Saha et al. 2017). The precise selection of the set of exons to be spliced-in in a specific context by AS process is in turn dependent on a plethora of trans-acting proteins including SFs, as well as on cis-acting sequence elements (Douglas L. Black 2003). The SFs interact with the splicing machinery through their binding to the pre-mRNA at cis-acting sequence elements and can either promote or repress the usage of a particular splice site (Hamid and Makeyev 2017). The cis-acting elements can be subdivided into four different subtypes based on their position and type of regulation: exonic and intronic splicing enhancers (ESE, ISE), both promoting splice site usage, or exonic and intronic splicing silencers (ESS, ISS), both repressing splice site usage (**Figure 1.2-6**). When it comes to SFs, the most predominant trans-acting regulators of AS belong either to the serine/arginine-rich (SR) or heterogeneous nuclear ribonucleoprotein particle (hnRNP) protein families (Busch and Hertel 2012). Through their binding on cis-acting elements, they could either promote or repress the selection and usage of particular splice sites. Specifically, SR proteins contain serine-arginine-rich repeat domains and generally promote splice sites usage by recruitment of the spliceosomal components or other auxiliary factors (Erkelenz et al. 2013). The proteins from the hnRNP family are in contrast known to mainly repress splice site usage by involving



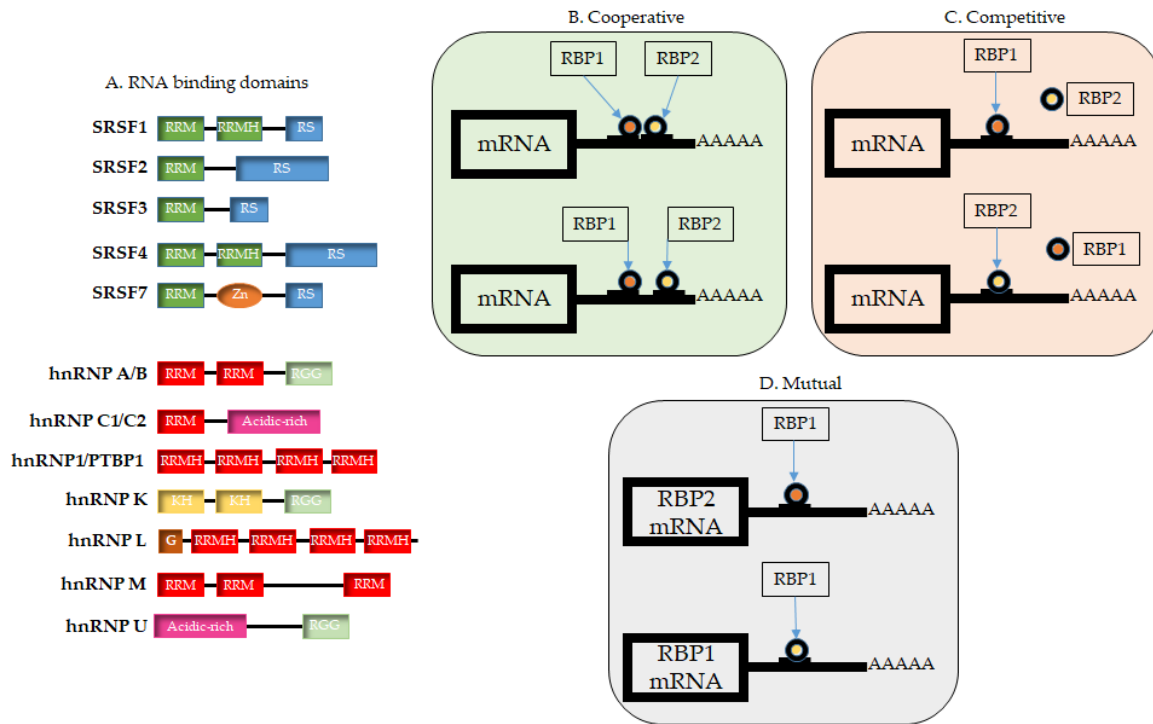
different mechanisms (Van Nostrand et al. 2020). These and other SR-like proteins most often regulate the same target RNA transcript, or either have overlapping binding sites on the same pre-mRNA, forming a complex regulatory network to yield the expression of their common target RNAs (Hamid and Makeyev 2017) (**Figure 1.2-6**).



**Figure 1.2-6:** Simplified representation of the cooperative control of AS by RBPs. In general, the splicing of the middle exon (red box) is modelled by binding of SFs. hnRNPs repress AS by binding to exonic or intronic splicing silencer (ISS, ESS) sites. In contrast, SR proteins promote AS through their binding of intronic or exonic splicing enhancer sites (ISE, ESE). hnRNPs and SR proteins interact with the components of the spliceosomal complex to modulate its binding on the pre-mRNA molecule.

### 1.2.2.3.1 Alternative splicing is tightly controlled by RNA-binding proteins networks

AS of the nascent pre-mRNA transcripts is a key determinant of the transcriptome, the protein levels, and the consequent phenotypes of the cell (Schwanhäusser et al. 2011). Each RNA transcript is tightly controlled by RBPs (Gerstberger, Hafner, and Tuschl 2014) which represent the post-transcriptional army of the cell that is encoded by at least 1,500 RBP-encoding genes (Castello et al. 2012). The RBPs influence mRNA metabolism or its life-cycle at different levels including splicing, as well as transport, localization, translation, and stability (Dreyfuss, Kim, and Kataoka 2002) and their mis-regulation has been associated with several diseases, including cancer (Qin et al. 2020). The binding of pre-mRNA molecules by RBPs is mediated by the presence of structurally well-defined RNA-binding domains (RBDs) contained within the RBPs, including the most frequent RNA Recognition Motifs (RRM), hnRNP K Homology (KH) domain, or Dead Box helicase Domain (Linder and Jankowsky 2011) (**Figure 1.2-7**).



**Figure 1.2-7:** Schematic representation of major classes of RNA SFs: Serine/arginine-rich (SR) and heterogeneous nuclear ribonucleoproteins (hnRNPs), and their regulatory interplays. A, Examples of RBPs with diverse RBDs involved in the pre-mRNA binding. RRM, RNA Recognition Motif; RRMH, RNA Recognition Homology Motif; RS, serine/arginine-rich repeat domains; Zn, Zn-binding domain; KH, K homology domain; RGG, arginine-glycine-glycine repeat-rich region; G, Glycine rich region; Acidic-rich, acidic amino acid residue-rich region. B, cooperative interplay between two RBPs having the same target with binding either proximal (top) or distal (bottom) binding sites. C, competitive interplay between two RBPs which compete for the binding of the same binding sites and results in opposite effects. D, Mutual interplay where an RBP regulates the expression of another RBP (top) or its cognate mRNA expression (bottom).

The fine-tuned control of AS by the RBPs relies on one hand on their relative cell- or tissue-specific expression (Van Nostrand et al. 2020) as well as on the molecular interactions between different RBPs (Quattrone and Dassi 2019). The aberrant expression as well as genetic variations of several RBPs-encoding genes have been associated with different cancer types and were observed in different cancer cell lines (B. Zhang et al. 2020). The number of characterized human RBPs has been recently increasing as confirmed by different methods such as mass spectrometry-based proteomics and RNA pulldown techniques (Brannan et al. 2016). This continuously increasing repertoire of RBPs is likely to underlie the complexity of the posttranscriptional regulation of GE process, and has motivated researchers to investigate the binding properties, RNA targets, and functional roles of these proteins under normal and disease states, including cancer (Bonnal, López-Oreja, and Valcárcel 2020).

The integration of different data generated by different techniques such as the crosslinking and immunoprecipitation followed by sequencing (CLIP-seq), systematic evolution of ligands by exponential enrichment (SELEX) and mass-spectrometry-based proteomics enabled the construction of RBP-RNA regulatory networks and helped elucidating their regulatory mechanisms. *In vivo* binding assays such as CLIP-

seq are widely used and provide candidate functional elements directly bound by individual RBPs (M. Zhang et al. 2020). While efforts in the past have been dedicated for the identification of RNA targets of each single RBP, recent studies instead have shown that RBPs form complex and hierarchical posttranscriptional regulatory networks by either direct physical interactions or indirectly through binding of same target RNA transcripts (Quattrone and Dassi 2019).

Different models were proposed to describe the regulatory interactions among RBPs. In the cooperative model, the regulatory actions of two or more RBPs on a common pre-mRNA target are modulated by either their direct physical interaction when they bind to proximal RBP-binding sites on the target transcript, or when they bind to distal sites that are brought in proximity by the secondary structure of the target RNA (Quattrone and Dassi 2019) (**Figure 1.2-7B**). Moreover, the two RBPs could exert their same regulatory effect on the common target independently of each other through binding to distal binding sites without being physically interacting. In contrast, in the competitive model, two or more RBPs compete for the binding of the same or overlapping RBP-binding sites and their binding induces opposing effects on the target transcript (**Figure 1.2-7C**). Thus, the binding of one RBP interferes with and inhibits the binding of the other RBP and this results in the creation of a balance between the concentration of transcripts bound by antagonistic RBPs, yielding a fine-tuned expression of the final RNA transcript (Quattrone and Dassi 2019). In the mutual model, one RBP could bind and regulate the expression of either its cognate mRNA or of another RBP, resulting in the mutual regulation of the expression of their target transcript (Dassi 2017) (**Figure 1.2-7D**).

Furthermore, SFs control AS by regulating the usage and selection of splice sites. Accumulating evidence, however, has shown that SFs could show antagonistic activities by either activating or repressing splice site selection, depending on their binding positions (X.-D. Fu and Ares 2014), thus adding another layer of complexity to the posttranscriptional control of GE process by AS. For instance, SFs of the SR and SR-like family enhance AS only when they are recruited to the alternative exon, and they repress AS when their binding is located on the opposite intronic side of the 5' splice site (S. Lin and Fu 2007). SFs members of the hnRNP family instead display analogous opposing effects on AS but in a reversed position-dependent manner (Huelga et al. 2012). Moreover, a recent study by Dominguez et al, using an RNA Bind-n-Seq (RBNS) assay, investigating the binding specificity of 78 human RBPs containing diverse RBDs, including RRM- and KH-domain containing-RBPs was performed (Dominguez et al. 2018). In this study, authors revealed that RBPs binding affinities are influenced in a sequence, structure and context dependent manner (Dominguez et al. 2018).

### **1.2.3 non-coding RNAs as regulators of gene expression at transcriptional and posttranscriptional levels**

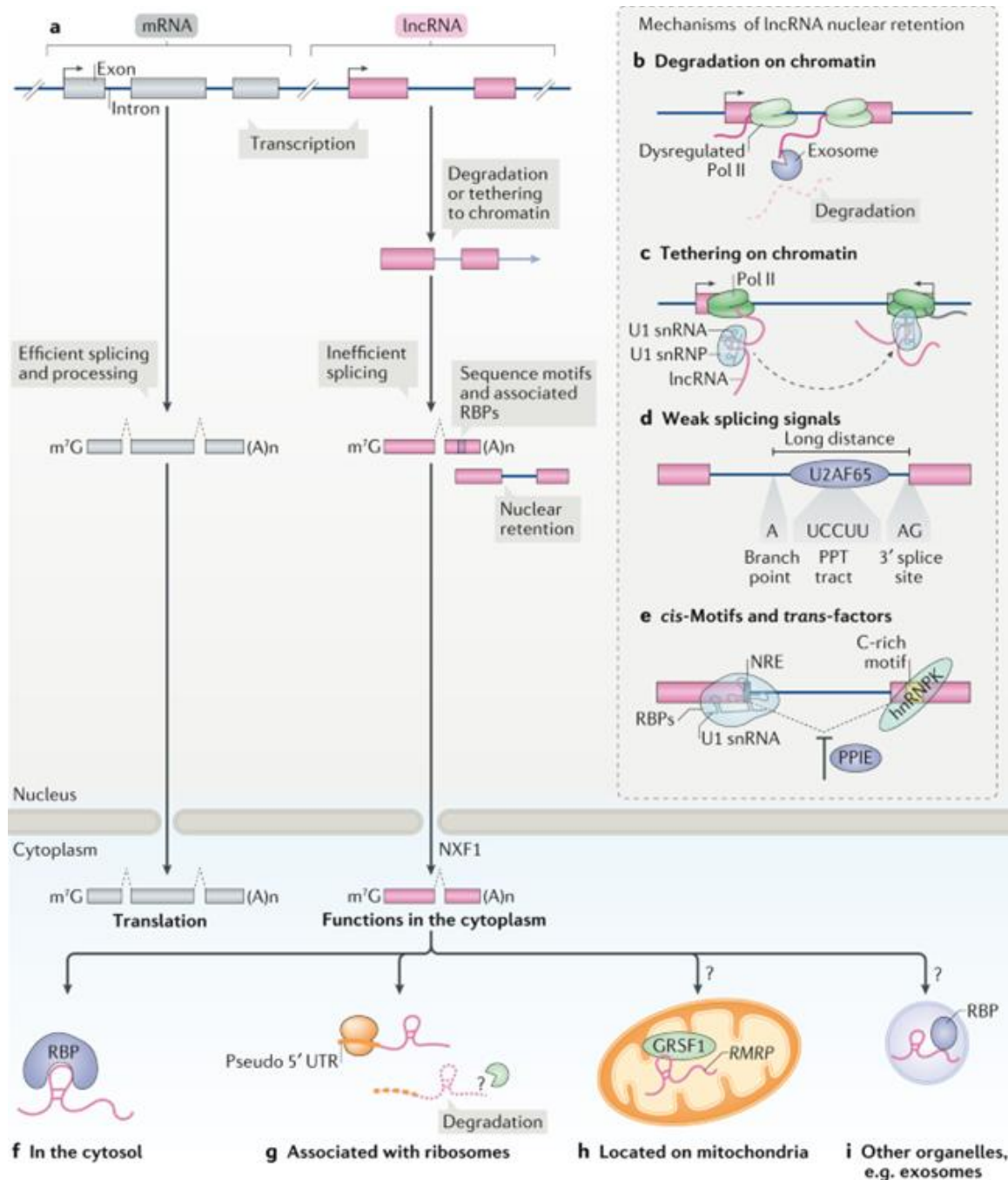
ncRNA transcripts constitute the vast portion of the human transcriptome; over 85,000 human transcripts not coding for proteins are annotated so far (Uszczyńska-Ratajczak et al. 2018); and similar to coding RNAs, are often polyadenylated, but lack evident open reading frames (ORFs) (Fatica and Bozzoni 2014). They are

globally classified based on their size into short and long ncRNAs. The short ncRNAs less than 200 bp like miRNAs or snRNAs are involved in different biological processes, including AS, transcription, and translation (Ameres and Zamore 2013; Morris and Mattick 2014). On the other hand, a growing number of functional ncRNAs longer than 200 bp (lncRNAs) have been characterized (Geisler and Collier 2013). LncRNAs could be classified based on their location into intergenic, exonic, intronic, in enhancer regions or in the region distal to protein-coding genes (Geisler and Collier 2013). Intragenic lncRNAs are defined also as sense or antisense lncRNAs based on the gene orientation with respect to the protein-coding gene (Ransohoff, Wei, and Khavari 2018). Most lncRNAs are transcribed by polymerase II (Pol II) and can be capped, polyadenylated, and some are processed similarly to mRNAs (**Figure 1.2-8a**). Recently, a number of studies have started to unravel distinct mechanisms of transcription, processing, export, and turnover of lncRNA transcripts, and such features are related to their functional activity (Statello et al. 2021). Compared to mRNAs, lncRNAs exhibit a differential cellular localization, with the vast majority being localized in the nucleus (Tian and Manley 2017). The comparison between lncRNA and mRNA genes has revealed that lncRNA genes are less evolutionary conserved, and are less abundantly expressed (Quinn et al. 2016). However, some studies have found that some lncRNA encoding genes are abundantly expressed in the brain (Serena Liu and Trapnell 2016), in specific cell types (Yao, Wang, and Chen 2019), or developmental processes (Sarropoulos et al. 2019).

### 1.2.3.1 lncRNAs transcription, processing and localization

LncRNAs exhibit distinctive features such as expression status, transcription mode, splicing efficiency, and cellular localization (Statello et al. 2021). For example, the low expression of certain lncRNAs has been linked to the presence of repressive chromatin markers at their promoters (Lagarde et al. 2017). Furthermore, as for protein-coding genes, the transcription stages of certain lncRNAs depend on the phosphorylation status of the Pol II at the C-terminal domain (Schlackow et al. 2017), but a high proportion of these lncRNAs are transcribed also by a dysregulated Pol II (Schlackow et al. 2017). The nuclear retention of most lncRNAs is related to their transcription mode and splicing efficiency (Melé et al. 2017). Some groups of lncRNAs appeared to be weakly spliced and the transcription termination at these genes occurs independently of polyadenylation signals, resulting in their accumulation on chromatin, followed by their degradation by RNA exosomes (Schlackow et al. 2017) (**Figure 1.2-8b**). The chromatin-tethering effects of lncRNAs was reported to be mediated by the presence of cis sequence motifs, embedded within the lncRNA sequences, that are recognized by certain nuclear factors (Shukla et al. 2018) (**Figure 1.2-8c**). For instance, a chromatin-tethering event of lncRNAs was observed when the Pol II-associated elongation factor SPT6 is not functional (Nojima et al. 2018). The active transcription mark H3k36me3 is redistributed from protein-coding genes to lncRNA genes when SPT6 is lost, thereby increasing their transcription (Vos et al. 2018; Nojima et al. 2018). The loss of SPT6 function impairs the recruitment of the transcription termination integrator complex to chromatin, leading to accumulation of lncRNAs on chromatin (Nojima et al. 2018). Furthermore, other aspects of lncRNAs processing is that they are spliced less efficiently compared to mRNAs (Guo et al. 2020). These molecules have weak internal splicing signals, and are characterized by the fact that the 3' splice site is located far away from the branch point

(Rosenberg et al. 2015), such feature is correlated with increased nuclear retention (**Figure 1.2-8c**). Other factors, such as the differential expression of certain SFs have been associated with the accumulation of lncRNAs within the nucleus (**Figure 1.2-8d**). For example, the higher expression of the splicing inhibitor peptidylprolyl isomerase E in mouse embryonic stem cells suppresses splicing of a subset of lncRNAs, leading to their nuclear accumulation (Guo et al. 2020) (**Figure 1.2-8e**). Alternative polyadenylation signals have also been associated with nuclear accumulation of lncRNAs (Lubelsky and Ulitsky 2018) (**Figure 1.2-8e**).



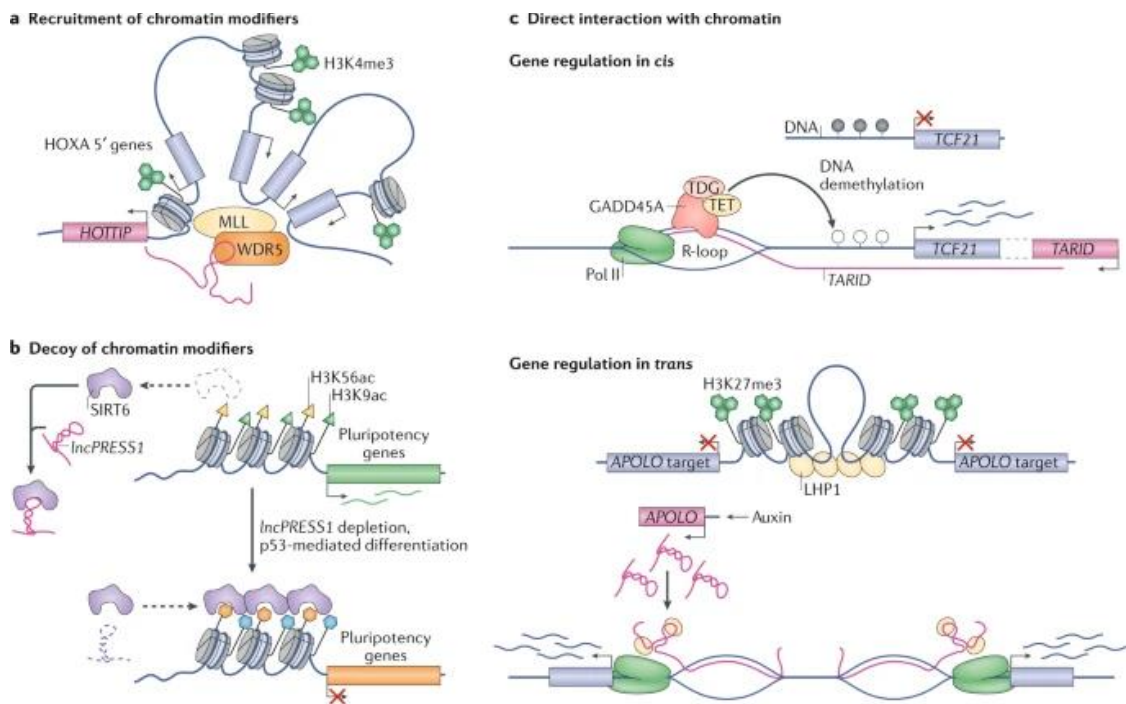
**Figure 1.2-8: Biogenesis, and processing of lncRNAs.** (a), biogenesis of lncRNAs. Some lncRNAs are transcribed by RNA polymerase II, weakly processed and are retained in the nucleus, while others undergo splicing and export to cytoplasm. Some lncRNAs having fewer exons are exported to the cytoplasm by the NXF1 pathway. (b), other lncRNAs are

*transcribed by deregulated Pol II, remain on chromatin, and get degraded by nuclear exosomes. (c), Numerous lncRNAs harbouring with U1 binding motif can recruit the U1 small nuclear ribonucleoprotein, which enables their association with Pol II at various loci. (d), In some lncRNAs, the sequence between the 3' splice site and the branch point is longer and causes inefficient splicing. (e), sequence motifs in cis containing nuclear retention element (NFRE) motifs and factors in trans like the differential expression of some SFs (hnRNPK) contribute to nuclear retention of lncRNAs. (f), some lncRNAs once in the cytoplasm are associated with many RBPs. (g), other lncRNAs associate with ribosomes through their pseudo 5' untranslated regions, while others (h) are sorted in the mitochondrial matrix. (i), some other lncRNAs associate with other organelles like exosomes probably by forming lncRNA-RBP complexes. m7G: 7-methyl guanosine 5' cap; (A)n, poly(A) 3' tail. From (Statello et al. 2021).*

On the other hand, a high proportion of lncRNAs are exported in the cytosol, presumably by the same export pathway of mRNAs (Zuckerman et al. 2020). Upon arrival to the cytoplasm, different lncRNAs are carefully assigned to different organelles or associate with different RBPs (**Figure 1.2-8f**), or with ribosomes under the influence of certain cis elements such as the long pseudo 5' untranslated regions (Zeng and Hamada 2018) (**Figure 1.2-8g**). Although 70% of cytoplasmic lncRNAs are found bound to ribosomes, whether they are engaged in translation or are degraded potentially by a translation-dependent mechanism is still under debate (Carlevaro-Fita et al. 2016). Other cytoplasmic lncRNAs are sorted into mitochondria where they interact with RBPs (Noh et al. 2016) (**Figure 1.2-8h**) while other lncRNAs accumulate within exosomes (**Figure 1.2-8i**), possibly by a mechanism likely involving the binding of specific sequence motifs by RBPs (Statello et al. 2018).

### 1.2.3.2 Gene expression regulation by lncRNAs at epigenetic level

GE process is regulated by lncRNAs at different levels (Gil and Ulitsky 2020). lncRNAs have the dual ability to bind proteins and make direct base-pairing with DNA and RNA molecules. The lncRNA-containing complexes can modulate chromatin functions, and transcription of both neighboring and distant genes and affect different aspects of RNA processing, including splicing, transport, stability, and translation (Statello et al. 2021). The application of the chromatin conformation capture (3C) techniques has revolutionized the complex lncRNA-mediated regulation of GE at the *epigenetic* level (Bonetti et al. 2020). lncRNAs could regulate chromatin architecture by directly binding to target DNA regions or indirectly by interacting with chromatin modifying enzymes, facilitating their recruitment to their target gene promoters, yielding inhibition or activation of the GE process (Mumbach et al. 2019) (**Figure 1.2-9a**). Furthermore, lncRNAs could also serve as decoys of chromatin modifying enzymes (Jain et al. 2016) (**Figure 1.2-9b**). Moreover, lncRNAs could directly interact with target DNA regions to form DNA-RNA R-loops recognized by chromatin modifying enzymes, which results in the activation or inhibition of target genes (Arab et al. 2019) (**Figure 1.2-9c**). Other lncRNAs control chromatin architecture in *trans* mode, such as the lncRNA auxin-regulated promoter loop (APOLO), which once is active by auxin, recognizes specific binding motifs on the promoters of target genes, subsequently binds and generates an R-loop that acts as a decoy for the Polycomb factor like heterochromatin protein 1 (LHP1), allowing the expression of target genes (Ariel et al. 2020).



**Figure 1.2-9:** GE regulation by lncRNAs at the epigenetic level. (a), lncRNAs can physically interact with chromatin modifiers and recruit them to target gene promoters, resulting in activation or inhibition of transcription. The example illustrates HOTTIP regulation of HOXA 5' genes transcription. HOTTIP, HOXA transcript at the distal tip; MLL, Mixed-Lineage Leukaemia; WDR5, WD repeat-containing protein 5; H3K4me3, Histone H3 Lysine 4 trimethylation. (b), lncRNAs can act as decoys of chromatin modifiers sequestering them from target gene promoters. SIRT6, histone deacetylase sirtuin 6; lncPRESS1, p53-regulated and embryonic stem cell-specific lncRNA; H3K56ac, Histone 3 Lysine 56 Acetylation; H3K9ac, Histone 3 Lysine 9 acetylation. (c), lncRNAs can regulate GE in cis or in trans by directly interacting with DNA forming DNA-lncRNA complexes recognized by chromatin modifying enzymes or by transcription factors. TCF21, Transcription Factor 21; TARID, the lncRNA TCF21 antisense RNA inducing demethylation; TDG, thymine-DNA glycosylase; TET1, ten-eleven translocation 1; APOLO, auxin-regulated promoter loop; LHP1, heterochromatin protein 1. From (Statello et al. 2021).

### 1.2.3.3 Gene expression regulation by lncRNAs at the transcriptional level

Several examples of RNA-DNA complexes involving lncRNAs have been reported under both physiological and pathological conditions. lncRNAs interactions with DNA could result in either activation (Mondal et al. 2015) or inhibition (O'Leary et al. 2015) of the GE process. A classic example of this mechanism is the lncRNA Xist, which is responsible for chromosome X inactivation process in cells of female mammals (Wutz 2011). During embryonic development, the lncRNA Xist causes the silencing of a subset of genes located on one of the two X chromosomes involving a complex interplay with different proteins (Colognori et al. 2019). Another example of many is the lncRNA ANRIL, which mediates the recruitment of PRC1 and PRC2 to the promoters of its neighbouring genes, *CDKN2A* and *CDKN2B*, thereby controlling their expression and regulating cell senescence (Yap et al. 2010). The same lncRNA could control GE in trans-acting fashion through alu sequences where ANRIL drives the recruitment of PRC1 and PRC2 to the promoters of distant target genes (Holdt et al. 2013). Yin Zhang and colleagues, by analysing TCGA RNA-seq and ChIP-Seq data and Xenograft tumors, they confirmed previous evidence from Miano and colleagues (Miano et al. 2018), that

lncRNA *DSCAM-AS1* expression is regulated by two super-enhancers. These enhancers are driven by FoxA1 binding and *DSCAM-AS1* promotes cancer progression by interacting with YBX1, an essential complex for the expression of FoxA1 and ER $\alpha$  (Y. Zhang et al. 2020). A number of studies have reported the implications of lncRNAs in controlling transcription are summarized in (table 1.2-1).

**Table 1.2-1:** Examples of lncRNAs, their functions and mechanisms in controlling GE at the transcriptional level.

LncRNA	Interacting partners	Functional mode	Related pathologies
ANRIL (Yap et al. 2010)	PRC1, PRC2, YY1	recruits polycomb repressive complex to the promoters of <i>CDKN2A</i> and <i>CDKN2B</i> to inhibit their expression in <i>cis</i> and to regulate distant genes in <i>trans</i>	multiple roles in diseases including cancers (Y. Kong, Hsieh, and Alonso 2018)
<i>LINC-PINT</i> (Marín-Béjar et al. 2017)	PRC2	Suppresses gene-expression signature of cancer cell invasion	generally downregulated in multiple cancers; reduces cell migration and proliferation when overexpressed in lung cancer cells.
<i>CCAT1-L</i> (Xiang et al. 2014; Cai et al. 2020)	CTCF, hnRNPk	induces chromatin looping by changing enhancer availability for MYC and verus PVT1	induces cancer progression, metastasis and resistance in multiple cancers (Z. Liu, Chen, and Hann 2019)
<i>DSCAM-AS1</i> (Y. Zhang et al. 2020)	YBX1	Physically interacts with YBX1 and regulates its recruitment at FOXA1 and ER $\alpha$ promoters.	overexpressed in lung, breast and prostate cancers. Induces cell growth in xenograft tumors (Y. Zhang et al. 2020)



<i>SWINGN</i> (Grossi et al. 2020)	SWI/SNF	activate the proximal gene GAS6 and distant genes in trans through chromatin looping	SWINGN depletion reduces cancer cell growth in lung cancer xenografts (Grossi et al. 2020).
------------------------------------	---------	--	---

#### 1.2.3.4 Gene expression regulation by lncRNAs at the post-transcriptional level

In addition to their roles in controlling transcription and chromatin organization, lncRNAs also regulate the GE process post-transcriptionally. One of the mechanisms by which lncRNAs regulate GE at the post-transcriptional level is by acting as competitive endogenous RNA or miRNAs sponges. A number of lncRNAs were reported to bear miRNA-complementary sites and their competitive behaviour with miRNAs for binding mRNAs has been reported (Salmena et al. 2011). The lncRNA *DSCAM-AS1* was reported to accelerate the progression of hepatocellular carcinoma (HCC) via sponging miR-338-3p (D. Ji et al. 2019). The knock-down of *DSCAM-AS1* inhibited proliferation, migration and invasion in HCC cells and those effects were at least partially reversed by inhibition of miR-338-3p, suggesting they have common targets (D. Ji et al. 2019). Similarly, the lncRNA *MALAT1* regulates *ZEB1* expression by sponging miR-143-3p and promotes progression of HCC (Lisha Chen et al. 2017). Another example of many is the lncRNA *LINC01426* which by sponging miR-345-3p induces an upregulation of *VAMP8* gene and thereby facilitates glioblastoma progression (Cao, Tang, and Su 2020). Other studies have shown that some lncRNAs could act on important biological processes such as the Epithelial-to-Mesenchymal Transition (EMT) process via sponging target miRNAs (Wan et al. 2020). A number of example lncRNAs with miRNAs sponging effects and their implications in diseases are reported in (Table 1.2-2).

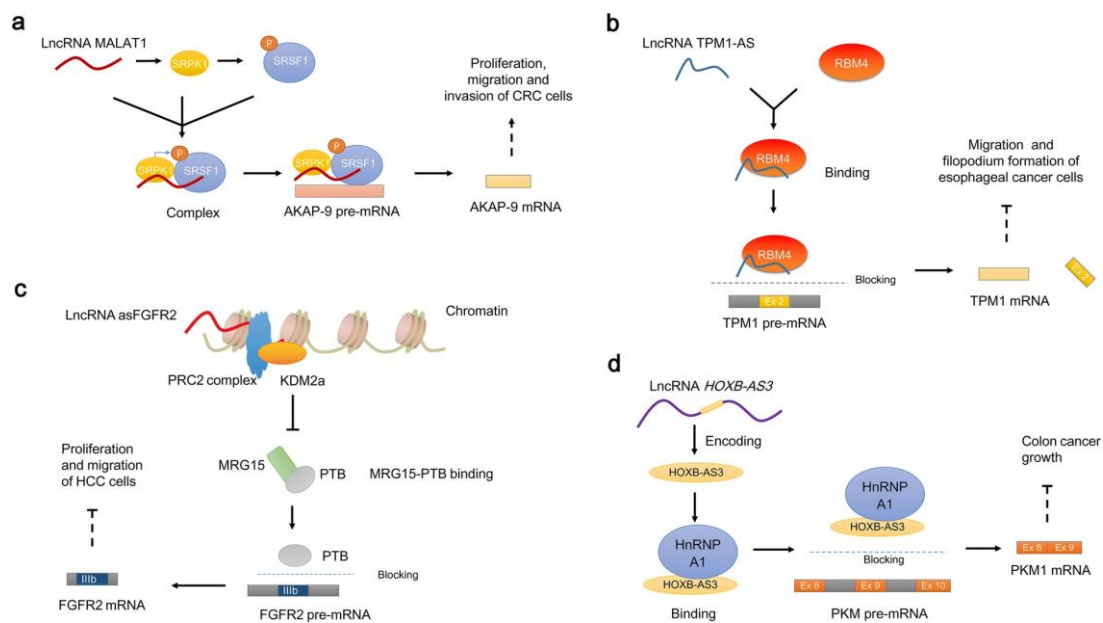
**Table 1.2-2:** example lncRNAs mediating post transcriptional control of GE process via miRNAs sponging mechanism.

LncRNA	Target miRNA	Physiopathological process	References
<i>DSCAM-AS1</i>	miR-338-3p	Induces cell proliferation, migration and invasion in HCC cells by upregulating Cyclin-D expression	(D. Ji et al. 2019)
<i>DSCAM-AS1</i>	miR-101-3p	promotes the progression of osteosarcoma through the upregulation of <i>USP47</i> expression	(Shanyong Zhang et al. 2020)

DSCAM-AS1	miR-384	promotes cell proliferation and colorectal cancer progression by inducing <i>AKT3</i> expression	(B. Li, Sun, and Zhang 2020)
MACC1-AS1	miR-384, miR-145-3p	Induces cell proliferation and breast tumor progression by enhancing PTN and c-MYC mRNAs expression	(X. Zhang et al. 2019)
lncRNA-PAGBC	miR-133b, miR-511	promotes tumor growth and metastasis in gallbladder cancer by inducing the AKT/mTOR signaling pathway	(X.-S. Wu et al. 2017)
TRPM2-AS	miR-612	promotes cell proliferation and gastric cancer progression and radioresistance by inducing the expression of IGF2BP1, c-MYC and FOXM1	(Xiao et al. 2020)
UCA1	miR-184	accelerates the proliferation and cisplatin resistance of oral squamous cell carcinoma through an upregulation of SF1-mediated AS	(Fang et al. 2017)
CCAT1	miR-490	promotes gastric cancer cell migration via increasing hnRNPA1-mediated AS	(B. Zhou et al. 2016)
circRNA 1000,146	miR-361-3P, miR-615-5p	promotes the proliferation and invasion of non-small lung cancer cells by inducing SF3B3-mediated AS	(Lijian Chen et al. 2019)

Moreover, in the last few years, a new role of lncRNAs in the post-transcriptional control of GE has emerged (Statello et al. 2021). Particularly in cancer, certain lncRNAs were found to interact with many RBPs including SFs, which results in the control of the RNA AS process (He, Luo, and Mo 2019) (**Figure 1.2-10**). A well-established lncRNA associated with RNA splicing is MALAT1. Tripathi and coworkers, by using *in-silico*

analysis and RNA immunoprecipitation assays combined with knock-down experiments, identified that SRSF1 can interact with MALAT1 via its RRM domain (Tripathi et al. 2010) (**Figure 1.2-10a**). Such interaction is required for the proper localization of SRSF1 as well as for other SFs to nuclear speckles, and MALAT1 depletion induced changes in the AS pattern of a subset of transcripts (Romero-Barrios et al. 2018). Furthermore, MALAT1 was reported to disrupt the SFPQ-PTBP2 (*polypyrimidine tract-binding protein 2/proline- and glutamine-rich SF*) complex by hijacking SFPQ, which results in the release of PTBP2, thereby allowing cancer progression and metastasis (Q. Ji et al. 2014). The lncRNA LINC01133 which is linked to different cancers has been proposed to inhibit EMT and metastasis in colorectal cancer through its interaction with SRSF6 (J. Kong et al. 2016). Using RNA pulldown and luciferase assays, Zhang et al., showed that the lncRNA MACC1-AS1 harbour binding sites for multiple tumor suppressor miRNAs, including miR-384 and miR-145-3p which repress pleiotrophin (*PTN*) and *c-MYC* expression (X. Zhang et al. 2019). The binding of miRNAs alters breast tumor growth and progression through an increase in the expression of *PTN* and *c-MYC* mRNAs. In addition, MACC1-AS1 competitively interacts with PTBP1, via its conserved pyrimidine-rich motif, and this binding further enhanced the miRNAs sponging effects of this lncRNA, and decreased the availability of PTBP1 to target mRNAs (X. Zhang et al. 2019). A number of other examples of lncRNA directly or indirectly regulate AS in different contexts and cell types are reported elsewhere (Yunze Liu et al. 2021).



**Figure 1.2-10:** lncRNAs regulate AS through the post transcriptional control of SFs and resulting downstream effects. (a), The lncRNA MALAT1 promotes the phosphorylation of SRSF1 by inducing the expression and activity of SRPK1, thereby promoting the SRSF1-mediated AS of AKAP-9 pre-mRNA and enhancing the expression of AKAP-9 isoform which exacerbates CRC. (b), lncRNAs could also act as decoys for SFs. The lncRNA TPM1-AS can directly bind RBM4 and inhibits its binding to TPM1 pre-mRNA, regulating its AS, thereby inhibiting cancer progression. (c), lncRNA regulation of AS by a chromatin remodeling mechanism. The lncRNA asFGFR2 promotes PRC2 and KDM2 recruitment to chromatin, which disrupts MRG15-PTB complex formation, thereby altering the location of PTB and promoting exon IIIb inclusion in the final FGFR2 mRNA which inhibits HCC proliferation and migration. (d), lncRNAs could have weak coding potential and the resulting peptides could be functional. The lncRNA HOXB-AS3 encodes a peptide HOXB-AS3

---

*which targets and inhibits the binding of hnRNPA1 to PKM pre-mRNA, thereby reducing PKM2 isoform expression and consequently repressing the growth of colon cancer. From (Yunze Liu et al. 2021).*

#### **1.2.4 Alternative splicing process as a regulator of development, differentiation, and tissue identity definition**

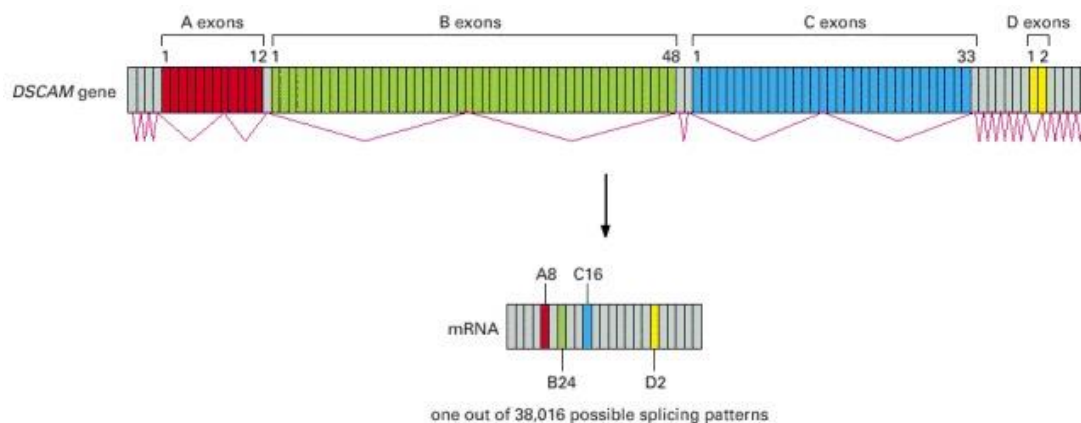
The expression level differences between alternative RNA transcripts could in some cases be nearly undetectable but enough to change the functional properties or localization of the gene products (Andreassi, Crerar, and Riccio 2018). A fine-tuned control of the RNA transcripts balance is, therefore, needed through important processes such as development, differentiation, and it is not surprising that several diseases are caused by AS deregulation (Gallego-Paez et al. 2017). It has been shown that approximately up to 30% of AS events occur in a tissue-specific manner (Q. Xu, Modrek, and Lee 2002). Several studies have shown that AS events are differentially regulated in different tissues, suggesting that AS is a major contributing factor to phenotypic heterogeneity in mammals, and over 22,000 tissue-specific alternative RNA transcripts were identified (E. T. Wang et al. 2008). More importantly, the vast majority of genes that undergo AS during development or cell differentiation are not modulated in terms of their overall expression (up and downregulation), suggesting the important role of AS dynamics as a driver of tissue identity and development (Baralle and Giudice 2017).

For example, AS plays an important role in stem cell renewal and differentiation (Han et al. 2013). In this study, authors showed that the muscle blind-like RBPs, MBNL1 and MBNL2, differentially regulate cassette exon AS events in embryonic stem cells. The inhibition of these proteins in differentiated cells induces embryonic stem cell-like patterns of AS, while in contrast, the overexpression of these proteins in embryonic stem cells induces patterns of AS similar to those of differentiated cells (Han et al. 2013). The transcription factor 7-like 2 (TCF7L2), which has been strongly associated with type 2 diabetes, is another excellent example of tissue-specific AS. In the work of Prokunina et al., the expression of different TCF7L2 mRNA splice isoforms was examined in eight different human tissues observing a tissue-specific pattern of AS (Prokunina-Olsson et al. 2009). Specifically, significant differences were observed in the expression pattern of *TCF7L2* exon 7-8 in the analyzed tissues with respect to that in the pancreas, in which the gene has a pivotal role in controlling apoptosis and where the two exons were associated with single polymorphisms (rs7903146 and rs12255372) that represent a risk factor for type 2 diabetes (Prokunina-Olsson et al. 2009).

In the human brain, AS is more abundant and extensively used as compared to other organs and explains the intrinsic complexity and protein diversity of the brain (Ule and Darnell 2006). Moreover, the AS process in the nervous system is under a tight regulation in relation with a unique RBP expression pattern of specific neuronal populations, suggesting it may regulate cell type- and synapse-specific functions (Traunmüller et al. 2016). For example, during neurogenesis, the patterns of AS vary considerably, particularly are under the control of polypyrimidine tract binding (PTBP) 1 and 2, and serine/arginine repetitive matrix protein 4 (SRRM4) (Vuong, Black, and Zheng 2016). PTBP1 is highly expressed in progenitor and stem neuronal cells

in which it represses neuronal genes. Upon differentiation, PTBP1 is downregulated allowing the induction of *PTBP2* and *PBX1* which in turns allow the differentiation and activation of neuronal genes (Vuong, Black, and Zheng 2016). SRRM4 in turns controls the transcriptional repressor of genes REST which is required for neurogenesis. Through AS of REST, SRRM4 promotes the production of REST4 isoform that has a reduced repressive activity (Norris and Calarco 2012). Furthermore, neuronal and axonal migration is under the control of another subset of AS regulators including NOVA1, NOVA2, RBFOX1, RBFOX2, and RBFOX3 (Leggere et al. 2016).

As a classical example, the gene with the greatest known number of alternatively spliced mRNAs is a *Drosophila* (fruit fly) gene called The *D. melanogaster* cell adhesion molecule (*Dscam*), which encodes an axon guidance receptor (Schmucker et al. 2000). This gene can express 38,016 possible mRNA isoforms by virtue of AS (Celotto and Graveley 2001), three times as many as different proteins as there are genes in the *Drosophila* genome. The *Dscam* gene contains a high number (95) of alternative exons that are organized into four different clusters formed of 12, 48, 33, and 2 mutually exclusive exons each (Schmucker et al. 2000) (**Figure 1.2-11**). These *Dscam* isoforms were found to be expressed in a stochastic and combinatorial manner and exhibit isoform-specific homophilic binding, and are required for neuronal wiring and self-avoidance processes (Hattori et al. 2009).

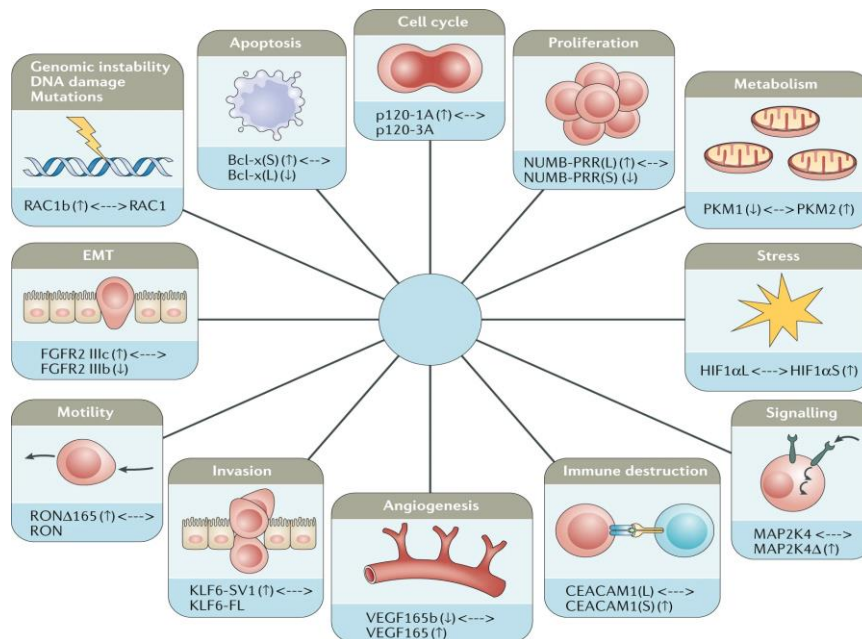


**Figure 1.2-11:** alternative splicing of RNA transcripts of the *Drosophila Dscam* gene. The gene contains 4 distinct clusters A, B, C and D, formed of 12, 48, 33, and 2 mutually exclusive exons each. One of the 38,016 possible mRNAs is shown. From (D. L. Black 2000).

#### 1.2.4.1 Dysregulation of alternative splicing in cancer

Defects in AS are frequently observed in human tumors and could result either from point mutations of the splicing sites of cancer-related genes or from changes in the expression levels of RBPs including SFs (S. Li et al. 2019). A set of SFs have been identified to act as oncoproteins or tumour suppressors, and contributes to the disease progression by modulating the expression of specific RNA isoforms involved in different cancer hallmark processes (Bonnal, López-Oreja, and Valcárcel 2020) (**Figure 1.2-12**). As an evolutionary process, in cancer tissues, cells are positively selected based on the expression of RNA isoforms that promote cell

proliferation, migration, resistance to endocrine therapy, or enable escape from cell death processes (Biamonti et al. 2014).



**Figure 1.2-12:** A schematic depiction of different cancer hallmarks (Hanahan and Weinberg 2011) along with examples of the AS alterations involved. Up and down arrows near to each RNA isoform indicate those which contribute most or least to the corresponding process, respectively. EMT, Epithelial-to-Mesenchymal Transition. From (Bonnal, López-Oreja, and Valcárcel 2020).

The AS process and its regulation mechanisms have recently become recognized as highly relevant for the understanding of every cancer hallmark, to the point that splicing alterations have been considered as another cancer hallmark (Rahman, Krainer, and Abdel-Wahab 2020; Desterro, Bak-Gordon, and Carmo-Fonseca 2020) (Figure 1.2-14). The global analysis of AS alterations in over 8,000 tumors across 32 cancer subtypes has revealed the presence of thousands of cancer-specific RNA splice variants that were completely absent in the non-malignant tissues, which likely to generate cancer-specific markers and neoantigens (Frankiw, Baltimore, and Li 2019). These and other observations continue to provide evidence for the AS process as a potential therapeutic target (Pardi et al. 2018).

The extensive analysis of AS process from The Cancer Genome Atlas (TCGA) data across 32 cancer types indicates that AS changes that confer the cancer cell with selective advantages may be caused by mutations in AS-regulatory sequences (Kelemen et al. 2013) or in *trans*-acting factors (Brooks et al. 2014). As a consequence, those mutations impair the proper recognition of splice sites, thereby affecting the AS patterns of multiple genes, which may include those with oncogenic or tumor suppressor functions (E. Kim et al. 2015). On the other hand, changes in the relative concentrations of SFs was reported to trigger oncogenic processes. For example, many SR-like and hnRNP SFs are overexpressed in multiple tumors and induce splicing changes promoting tumor cell proliferation (Golan-Gerstl et al. 2011). Conversely, the downregulation of other tumor suppressor SFs was also observed (Zong et al. 2014).

RBPs are frequently de-regulated in tumors and in particular the expression patterns of SFs were found to characterize different tumor types. The analysis of differentially expressed (DE) RBPs in 11 tumor types as compared to their normal counterparts revealed the vast majority (1,143 out of 1,348, 84.8%) of RBPs encoding genes are DE in at least one tumor type. The list includes 168 RBPs known as putative SFs and their expression profiles distinguish between different tumor types (Sebestyén et al. 2016). Similarly, a recent study conducted by analyzing the co-expression patterns of SFs in TCGA RNA-seq data, revealed that SFs are co-expressed in multiple tumor types forming two distinct subgroups, enhancer-SFs and suppressor-SFs. The observed co-expression profiles of SFs act by enhancing tumor aggressiveness, progression and promote metastasis formation (Koedoot et al. 2019).

Deregulated AS was linked to several processes, including cell proliferation, apoptosis, angiogenesis, and resistance to therapy (Sebestyén, Zawisza, and Eyras 2015). An excellent example of such alterations is the key apoptotic regulatory protein *Bcl-x* encoding gene which is alternatively spliced into two different isoforms generating proteins with completely opposite functions (Warren, Wong-Brown, and Bowden 2019). Specifically, the short isoform Bcl-xS promotes apoptosis, whereas the long isoform Bcl-xL inhibits apoptosis in cancer cells and its overexpression is associated with increased risk of metastasis in breast cancer (BC) (Mercatante et al. 2001). Another example is, the cell surface molecule *CD44* which harbours nine alternative exons between its constitutive exons. Differential inclusion of these variable exons generate over 20 splice variants (Brown et al. 2011). The inclusion of one of the variable exons generates CD44 variable isoform (CD44v), while skipping produces the standard isoform (CD44s). CD44v generally exhibits an epithelial expression, while the CD44s is mainly mesenchymal. Furthermore, the isoform switch from CD44v to CD44s is important for EMT and BC metastasis (H. Zhang et al. 2019). The tumor suppressor gene, Breast cancer type 1 (*BRCA1*), essential for DNA damage repairing process and for maintaining genomic stability, is aberrantly spliced in BC (Nielsen, van Overeem Hansen, and Sørensen 2016). AS of this gene produces three isoforms depending on the regulation of exon 11: BRCA1 full-length isoform (including the exon 11), BRCA1- $\Delta$ 11 (skipping of exon 11), and BRCA1- $\Delta$ 11q (partial skipping of exon 11). BC patients expressing isoform harbouring a mutation on exon 11 have a worse overall survival (OS) as compared to those with no mutation on exon 11. In addition, isoform-specific association with clinical data indicate BRCA1- $\Delta$ 11q is positively correlated to tumorigenesis and drug resistance (Nielsen, van Overeem Hansen, and Sørensen 2016).

Moreover, the oncogene *HER2*, whose overexpression delineates the HER2-positive BC subtype, is also aberrantly spliced in BC. The gene has a splice variant,  $\Delta$ 16HER2, lacking exon 20 which encodes a small extracellular domain. When it is co-expressed with the full length, wild type HER2 isoform,  $\Delta$ 16HER2 has been linked to resistance to the HER2-targeting monoclonal antibody, trastuzumab, in metastatic BC (Siyuan Zhang et al. 2011), affecting its biology and treatment response (Weigelt and Reis-Filho 2013). It is also reported that the  $\Delta$ 16HER2 variant has significant impacts on HER2-driven BC stemness (Castiglioni et al. 2006), on tumorigenesis (Turpin et al. 2016), and drug resistance (Jackson et al. 2013) as compared to the full length wild type isoform.

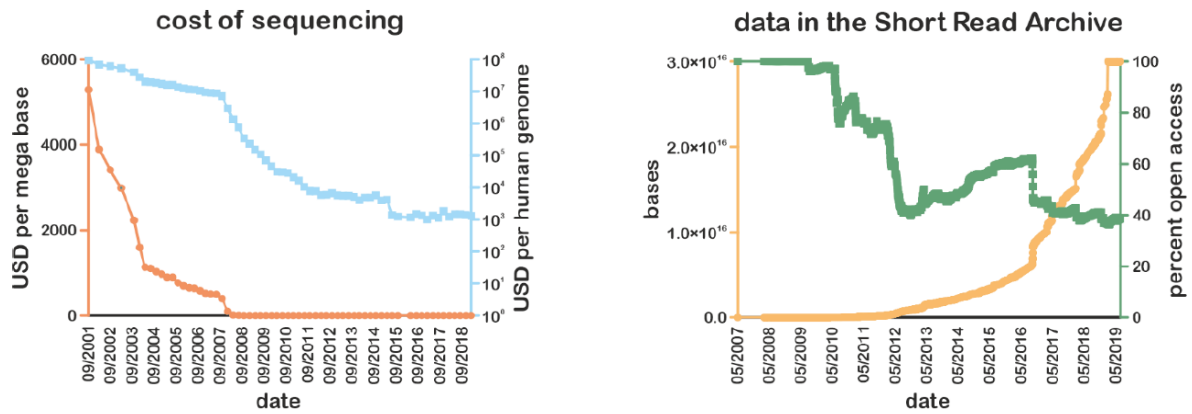
The ER $\alpha$  gene, which delineates the ER $\alpha$ -positive BC tumors, is also aberrantly spliced in BC. The gene produces multiple isoforms through AS in a tissue and disease-specific manner (Taylor, Martin-Hirsch, and Martin 2010). The full length ER $\alpha$ 66 isoform harbours two activation domains AF1 and AF2. The short isoform ER $\alpha$ 36 encodes a 29-amino acid protein lacking AF1 and AF2 domains. Another splice isoform ER $\alpha$ 46 only contains AF1 domain (Chantalat et al. 2016) and share the same 174-595 sequence with the full length ER $\alpha$ 66 isoform (Inoue and Fry 2015). The splicing isoform ER $\alpha$ 46 which is frequently expressed in BC tissues antagonises the function of the full length ER $\alpha$ 66 isoform in mammary cancer cells, and contributes to cancer development and drug resistance (Klinge et al. 2010).



### 1.3 High throughput techniques as tool for alternative splicing characterization on a genome wide scale

The conventional quantification approach for the characterization of AS is reverse transcription coupled to a target-specific and splicing-sensitive polymerase chain reaction (PCR). This was later followed by expressed sequence tags (ESTs) which enabled the characterization of widespread AS changes among different organisms (Modrek and Lee 2002). To characterize AS changes on a genome-wide scale, microarrays were developed and successfully used to examine AS across different tissues and species (Castle et al. 2008). However, these techniques were challenged by the nature of their low throughput, by a low signal to noise ratio, or by their limited ability to characterize only known AS events. To cope with these limitations, high throughput RNA-seq as well as development of AS quantification tools allowed the investigation of AS on a genome-wide scale, profoundly impacting our understanding of AS and its implication in biomedical research (Mardis 2011; Levy and Myers 2016).

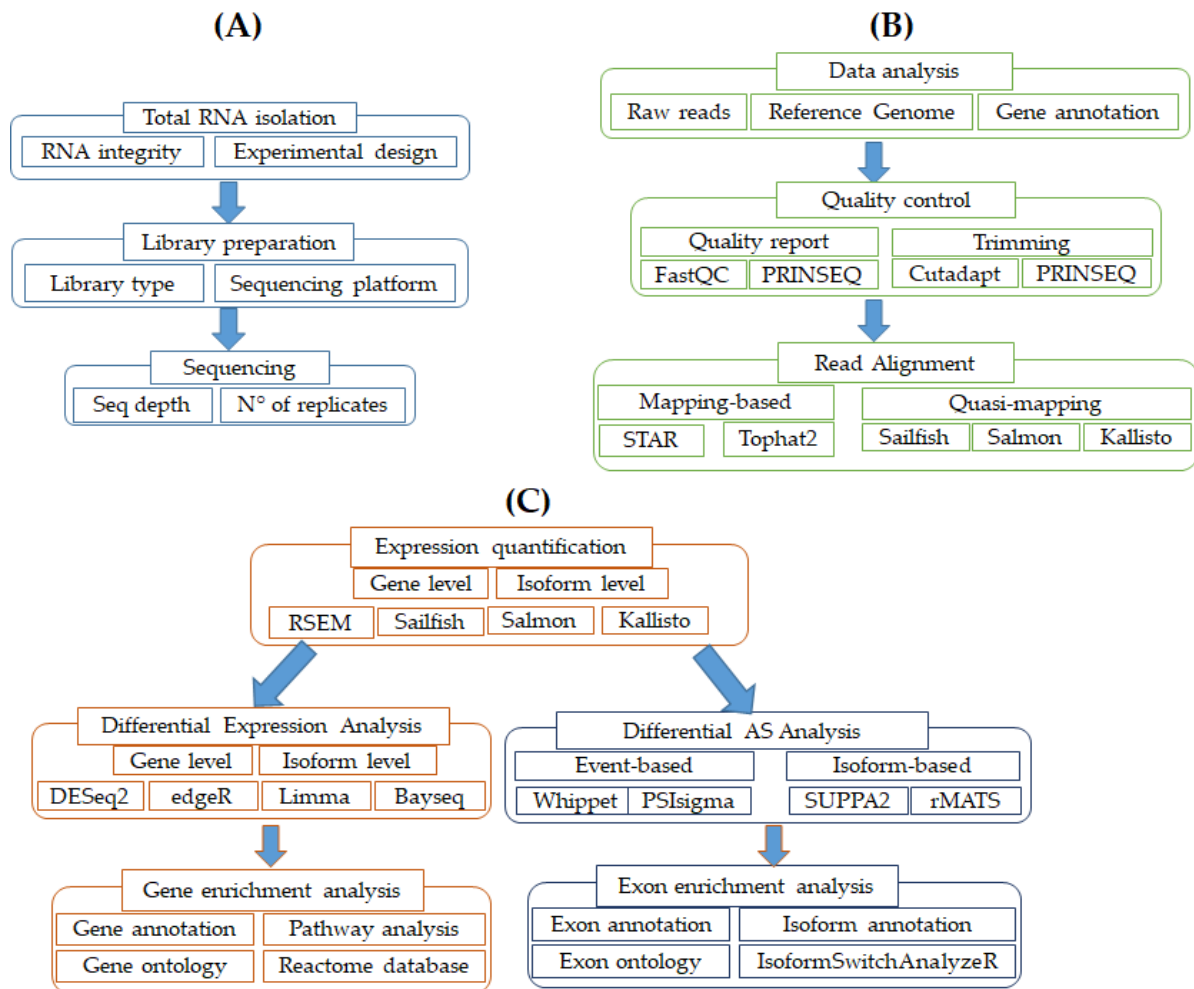
The RNA-seq technique has revolutionized the investigation of GE regulation on a genome-wide scale (Levy and Myers 2016). It allows the quantification as well as the identification of the sequences of virtually all the transcripts at once in a given sample. The parallel development of bioinformatics tools make it possible to process the RNA-seq outputs (in fastq format) and perform different analyses, including a differential gene/transcript expression analysis as well as differential AS analysis among data from different samples (Trapnell et al. 2012; Dobin et al. 2013; Shen et al. 2014; Bray et al. 2016; Trincado et al. 2018; Sterne-Weiler et al. 2018; K.-T. Lin and Krainer 2019). The diffusion of Next Generation Sequencing (NGS) techniques has been greatly accelerated by a significant drop in the sequencing cost that has occurred during the last decade. From 2001 to 2017, the cost of sequencing the human genome has dropped from almost 100 million dollars to just over 1,000 dollars in 2017 due to technical improvements (**Figure 1.3-1-left**). Subsequently, many research groups were allowed to perform RNA-Seq for their research and to date, a wide variety of RNA-Seq data is publically available. For example, RNA-Seq data of over 1.5 million samples are publically accessible in the Short Read Archive (SRA) for the National Center of Biotechnology Information (NCBI) and this number is steadily increasing (**Figure 1.3-1-right**). Finally, the description of the splicing code which strives to predict tissue-dependent AS outcomes based on hundreds of RNA features became possible (Barash et al. 2010).



**Figure 1.3-1:** Enhancement of next generation sequencing accelerates data production over the years. Left: the cost of sequencing per megabase (orange) and per human genome (blue) in US\$ between 2001 and 2019. Right: the line plot reports the amount of data stored in the Short Read Archive in bases (orange) and the proportion of data publicly accessible (green) from 2017 to 2019. From (Neumann 2019).

### 1.3.1 Overview of a typical RNA-seq data analysis pipeline

A typical RNA-seq analysis workflow comparing the genes/transcripts expression levels between two experimental conditions consists of four main steps starting from millions of nucleotide strings (defined as sequencing reads) generated by the sequencing instrument: i) read quality control; ii) reads mapping against a reference genome and expression quantification; iii) differential expression analysis; and iv) functional enrichment analysis to identify the regulated biological pathways and processes (**Figure 1.3-2**). The RNA-seq reads generated by Illumina Sequencers (the most diffuse sequencing instrument) are usually between 75 to 150 nucleotides in length. In the alignment step, these reads are aligned against a reference genome or transcriptome in order to identify the set of genomic regions they originally stem from. A set of bioinformatics tools called short-read aligners such as BWA (H. Li and Durbin 2009), bowtie2 (Langmead and Salzberg 2012), Tophat2 (D. Kim et al. 2013), and STAR (Dobin et al. 2013) have been developed to perform this analysis. This analysis step is one of the most computationally challenging in an RNA-seq analysis workflow as the short reads consist of sequences from multiple sets of exons. To achieve a proper and faster read alignment, the aligner generally splits them into subsequences during mapping. However, the correct position where the reads should be split into small parts is not known beforehand and thus several fragmentations are needed to yield optimal alignments results which makes this step computationally expensive. To cope with these challenges, quasi-mapping aligners such Sailfish (Patro, Mount, and Kingsford 2014), Salmon (Patro et al. 2017) and Kallisto (Bray et al. 2016) have been introduced to reduce the data structure complexity and improve the alignment time frame. While they improve the alignment time, these set of quasi-mapping aligners suffer from a potential downside pitfall in analysing and quantifying lowly expressed genes and small RNAs which may contain biological variations (D. C. Wu et al. 2018).



**Figure 1.3-2:** A schematic representation of the different steps performed in an RNA-seq experiment. A, a summary of RNA-Seq libraries preparation steps during which RNA samples are extracted and prepared for the sequencing instrument. B, steps of the computational steps where RNA-seq short reads are controlled for quality and aligned to a reference genome. C, Expression of genes and isoforms is quantified and experimental conditions are compared for the identification of DE and spliced genes.

### 1.3.2 Bioinformatic approaches to exploring and quantifying alternative splicing

Bioinformatics methods to explore AS are classified into three main categories: (i), AS detection methods using AS sequence conservation among species (Fiszbein et al. 2019), (ii), AS detection using microarray data (Lapuk et al. 2010), and (iii), and AS detection using RNA-seq data (Alamancos et al. 2015). With the availability of RNA-Seq, multiple transcriptomes have been sequenced and it has become feasible to explore AS on a genomic scale. Thereafter, a large number of alternative RNA transcripts have been identified along with the extraction of specific features of alternatively spliced exons using bioinformatics (Tranchevent et al. 2017). Specifically, bioinformatics tools have enabled easy exploration of AS dysregulation leading to diseases including cancer (Eswaran et al. 2013).

RNA-seq represents an unprecedented opportunity to quantitatively study AS in a systematic way (H. Feng, Qin, and Zhang 2013). In recent years, many researchers have been using RNA-seq and were able to

characterize AS alterations in different contexts (Kanitz et al. 2015). The application of bioinformatics AS detection tools and the steps to perform quantitative study of AS are now well established, including different visualization approaches (Garrido-Martín et al. 2018). Generally, two different approaches have been proposed to explore quantitative study of AS including (i) event-based and (ii) isoform-based approaches (C. Zhang et al. 2016). In the event-based approach, the short RNA-seq reads aligned against specific exons or splice junctions are counted, and appropriate statistical models are applied to quantify AS events and detect differentially regulated ones between distinct biological conditions (Shen et al. 2014; Sterne-Weiler et al. 2018; K.-T. Lin and Krainer 2019). The AS is quantified in this event-based approach using the widely used Percent Spliced-In (PSI) ratio which represents the percentage of usage of a specific gene's transcript exon (Schafer et al. 2015). The PSI value is calculated by counting the reads that support a specific exon and/or splice junction. The second approach of AS quantification tools, isoform-based, seeks to estimate the relative abundances and proportions of each transcript using RNA-seq short reads data (Alamancos et al. 2015; Trincado et al. 2018; Vitting-Seerup and Sandelin 2017b, 2019; Froussios et al. 2019). The relative abundance of different transcripts is possible by first aligning the short reads to a specific genome or transcriptome and then estimating the relative abundances by applying an Expectation-Maximization algorithm (Merino and Fernández 2020). The drawback of this approach is that inferring the abundance of full transcripts from short reads is not trivial as the results are sensitive to the chosen transcript annotation. Moreover, attributing changes of relative abundances to differential splicing regulation of specific exons or splices sites is not straightforward (Kanitz et al. 2015).

### 1.3.2.1 Human RNA transcripts annotation databases

The advent of RNA-seq technologies have recently switched the paradigm of genetic analysis from a genome to a transcriptome-based perspective. However, the precise functions of most individual splice isoforms (notably splicing events) are yet to be elucidated. Moreover, Gene Ontology databases provide annotation of gene products according to their biological processes, molecular functions, or cellular localizations, but no isoform-specific annotations are provided. Until recently, efforts have been investigated aiming at solving this issue where several algorithms and databases have been developed to provide isoform-specific annotations. For example, ISOGO (ISOform + GO function imputation) is a recent algorithm that has been developed to predict the functions of coding isoforms based on their protein domains and their correlation of expression among 11,373 cancer patients (Ferrer-Bonsoms et al. 2020). PEGASAS (Pathway Enrichment-Guided Activity Study of Alternative Splicing) is an algorithm that helps perform a correlation-based analysis between alternatively spliced exons and oncogenetic pathways (Phillips et al. 2020). The tool at a first step assesses the activity of signaling pathways by incorporating GE levels from RNA-seq experiments together with different functional gene sets and gives an activity score to each enriched pathway. Then, the tool identifies in a second step the pathway activity-correlated alternatively spliced events identified based on a correlation score between the alternatively spliced events and the enriched pathways or gene sets (Phillips et al. 2020). DIGGER is a tool that was developed for exploring the potential functional role of AS changes at the levels of protein-protein

interactions (Louadi et al. 2021). Using this approach, the alternatively spliced exons are judged for their downstream effects on a system biology level, where all the possible protein isoform specific interactions involving each alternatively spliced exon could be predicted, thus enabling the identification of the protein-protein interactions networks affected by each AS event (Louadi et al. 2021). Other tools based on the relative isoforms abundances such as the IsoformSwitchAnalyzeR R bioconductor package (Vitting-Seerup and Sandelin 2017b, 2019) were developed at the aim of helping understand the downstream functional consequences of the individual changes of the relative abundances of isoforms when comparing their expression between different conditions (e.g. cancer versus normal tissues). This tool helps users to integrate isoforms abundances together with their annotations for their coding potential (Y.-J. Kang et al. 2017), protein domains (Finn et al. 2014), and potential splicing changes leading to their expression.

### **1.3.2.2 RNA-binding proteins databases: RBPs-RNAs interaction networks and binding motifs**

The analysis of AS changes landscape not only implies the identification of mis-regulated splicing events, but also the determination of the RBPs possibly involved. To facilitate the identification of RBP-RNA pairs, several databases holding information on RBPs and their putative RNA targets have been developed. For example, the RNA-Binding Protein Database (RBPDB) is a database holding a collection of experimental observations of RNA binding sites information on 272 RBPs confirmed either *in vivo* or *in vitro* (Cook et al. 2011). As the number of characterized RBPs is continuously increasing, the number of RBPs annotation databases also increases. The Catalog of Inferred Sequence Binding Proteins of RNA (cisBP-RNA) is an annotation-rich database holding information on a high number of RBPs and their targets (Ray et al. 2013). These databases have made it possible to identify RBP-RNA pairs and enabled the determination of post-transcriptional regulatory mechanisms of GE.

## 1.4 Breast cancer and Estrogen Receptors

### 1.4.1 ER-alpha signaling pathway in mammary gland morphogenesis and development

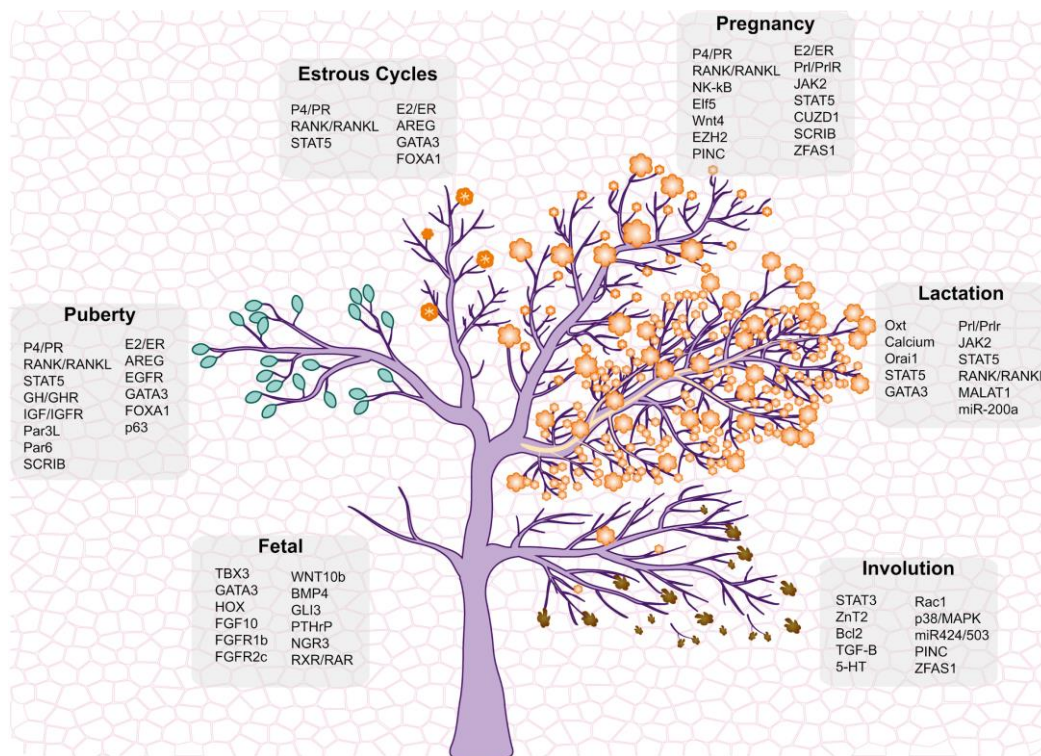
The estrogen receptor signaling pathway comprises estrogens and their two types of receptors called Estrogen Receptor  $\alpha$  (ER $\alpha$ ) and Estrogen Receptor  $\beta$  (ER $\beta$ ), members of the steroid/thyroid hormone nuclear receptor superfamily of ligand-regulated TFs (Moudgil 2013). Estrogens are steroid compounds playing a central role in the physiological control of reproductive tissues and functions, participating both in the development of the secondary sex characteristics and in the regulation of the menstrual cycle process in females. Estrogens are lipophilic molecules that control the GE process by passively diffusing through cell membranes and binding either one of their two receptors ER $\alpha$  and ER $\beta$  (Moudgil 2013). Different types of estrogens naturally exist in females and affect the mammary gland development at different stages, including estrone (E1), 17-estradiol (E2), and estriol (E3), with E2 is the most potent form of mammalian estrogenic steroids.

The mammary gland is one of the organs whose development and physiology are strictly dependent on estrogenic hormones and researchers have found that a woman's risk to develop BC is directly linked to her reproductive history and to her lifetime hormonal exposure (Briskin and O'Malley 2010; Rosen 2012). Anatomically, the mammary epithelium is composed of two different compartments, luminal and basal. The luminal compartment is formed by a layer of polarised cells around the duct lumen. The epithelial compartment is instead formed by myoepithelial and progenitor cells. The mammary epithelium is encapsulated into a basement membrane and surrounded by a complex stroma. The complex stroma tissue is composed of myoepithelial cells, extracellular matrix components within the basement membrane, fibroblasts, adipocytes, blood vessels, nerves, and various immune cells (Inman et al. 2015). There exists increasing evidence that interactions between the mammary epithelium and the surrounding stroma are crucial for normal mammary gland development (Dzięgielewska and Gajewska 2019). The ER $\alpha$  is expressed by both epithelial and stromal cells, and exhibits its activity involving both types of cells (Mallepell et al. 2006). The development and maturation of the mammary epithelium can be distinguished into two different phases including hormone-independent phase up to puberty, and hormone-dependent phase thereafter (Briskin and O'Malley 2010).

Several morphological and molecular changes occur during the mammary gland developmental process (**Figure 1.4-1**). At the fetal stage, the first stage of mammary gland development occurs during embryogenesis simultaneously as the specialization and maturation of the ectoderm and mesoderm occur. During this stage, the structural organization of the mammary gland is shaped by the ectoderm, whereas the mesenchymal signaling networks are responsible for guiding the ectodermal modifications and expression during mammary line positioning, placode assembly, and mammary bud formation and elongation. This stage is achieved through implication of molecular signaling pathways including members of the Fibroblast Growth Factor (FGF) and the Wingless-related integration site (WNT) protein families, which control the expression of TFs from the Homeobox gene family (HOX), GATA binding protein 3 (GATA3), and the T-box family (TBX), which are expressed either in the endoderm or mesoderm (L. S. Carroll and Capecchi 2015). The branching morphogenesis and expansion is regulated at this stage by Tbx2–3, Wnt genes, Parathyroid related hormone

(PTHrP), Msh homeobox 2 (MSX2), and Nuclear factor kappa B (NF- $\kappa$ B). On the other hand, the maternal hormones are the initial stimuli to the mammary gland for ductal development. At the puberty stage, the levels of ovarian hormones mainly estrogens (E2) and progesterones (P4) are increased. This increase in hormones promotes other developmental modifications, where it induces mammary gland growth and elongation of the epithelial ducts (Slepicka, Somasundara, and Dos Santos 2020).

Studies from ER $\alpha$  knockout mice models clearly demonstrated a strong relationship between ER $\alpha$  signaling and mammary gland development. The mammary phenotype of female mice with a disrupted ER $\alpha$  gene is normal before puberty and the developmental process at this stage is independent of ER $\alpha$ . However, the ducts failed to differentiate thereafter into terminal end-buds (TEBs), suggesting that ER $\alpha$  is required for ductal elongation during puberty and complete mammary gland development in the mature mouse (Bocchinfuso and Korach 1997). Furthermore, ER $\alpha$  is also important for secondary branching of the ducts during pregnancy and for the proliferation and maintenance of differentiated alveolar cells (Y. Feng et al. 2007). In contrast, Progesterone Receptor (PR) deficient mice models showed that the role of PR is to mediate full lobuloalveolar development of the mammary gland (Lydon et al. 1995). The exposure to estrogens upregulates the expression of PR in BC cell lines, and more recently, a genomic interaction between PR and ER $\alpha$  was reported and associated with a good clinical outcome in BC patients (Mohammed et al. 2015).



**Figure 1.4-1:** The blooming of mammary gland development. Schematic illustration of mammary gland developmental stages, showing fetal, puberty, estrous cycles, pregnancy, lactation and involution (from left to right). In puberty, green buds represent TEBs. Mammary alveoli are shown as orange flowers in estrous cycles, pregnancy and lactation. In lactation, the milk is represented as yellow sap flowing from the alveoli (flowers) to the ducts (branches). During involution, the regression of the mammary tissue is depicted with falling dead flowers and branches into the background, which portrays the fat pad. The basal compartment and luminal compartment are delineated with darker and lighter

---

colors in the tree, respectively. The main molecular regulators of each developmental stage are highlighted in the grey squares. From (Slepicka, Somasundara, and Dos Santos 2020).

## 1.4.2 ER-alpha signaling pathway in breast cancer development and progression

Breast cancer (BC) is one of the most significant and common diseases affecting women's health worldwide, and has been recognized as one of the top leading causes of cancer-related deaths in women (Ferlay et al. 2015). BC is a heterogeneous disease that affects different cellular types of the breast and results in variable phenotypes. The use of immunohistochemistry as well as GE profiling techniques resulted in the identification of different intrinsic BC subtypes that associate with variations in phenotypes, treatment-response, and disease-specific outcomes (Sørlie et al. 2003).

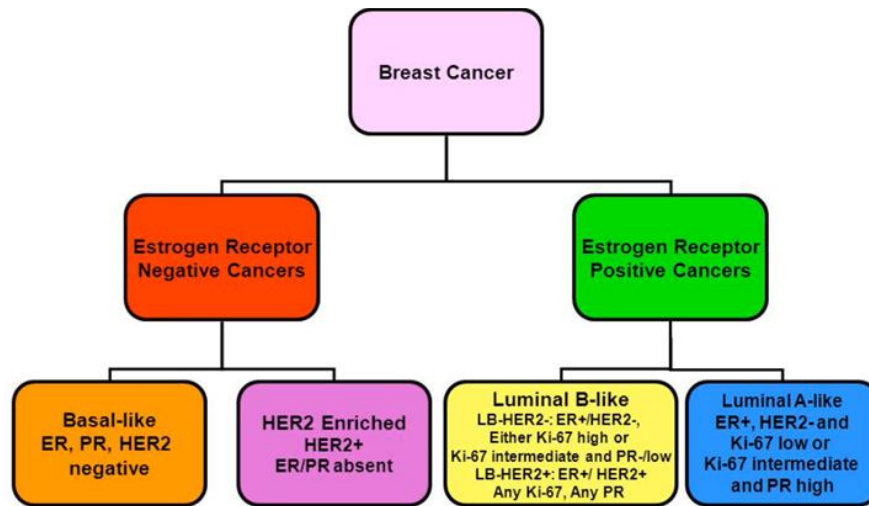
### 1.4.2.1 Histological classification of breast cancers

BC occurs in any cellular type of mammary gland and is known to exhibit a wide scope of morphological features and different immunohistological profiles, resulting in unique histological subtypes that have specific clinical course and outcome. The clinical profiling of BC cases results into different subtypes of BCs based on the tumor stage, histological grade, anatomical origin and immunohistochemical staining (Makki 2015). Consequently, BC can be widely classified into *in-situ* and *invasive* (infiltrating) carcinoma, with the latter being the most common form of BC at diagnosis. Based on its anatomical origin, the *in-situ* BC carcinoma is further subdivided into lobular (LCIS) and ductal (DCIS) subtypes, with the latter being the most frequently occurring form (over 20% of BC cases diagnosed as DCIS in USA). The progression of BC involves a series of clinical and pathological events, ranging from the appearance of a local atypical breast epithelial hyperproliferation, followed by a subsequent evolution into *in-situ* or *invasive* carcinoma and finally into metastasis formation (Bombonati and Sgroi 2011).

### 1.4.2.2 Molecular classification of breast cancer subtypes

Globally, BC is classified into significantly different subtypes based on the expression status of three receptors, including ER $\alpha$ , PR, and human growth factor receptor-2 (HER2) (Rouzier et al. 2005). Among these subtypes, the luminal ER $\alpha$ -positive subgroup representing the  $\frac{2}{3}$  of BC cases, HER2-amplified representing 20% of cases, and Basal-like triple negative (TNBC) subgroups (Reis-Filho and Pusztai 2011) (**Figure 1.4-2**). These molecular subgroups also differ based on other variable clinical features, including tumor morphology and grade classification, tumor size, and presence of lymph nodes metastases (Rouzier et al. 2005). While the receptor status classification remains the top most significant, the stratification of BC patients based on the expression profiles of ncRNAs transcripts have been recently described as a good classification approach of BC patients into different subtypes (Miano et al. 2016).





**Figure 1.4-2:** Classification of BC into the main major 4 subtypes based on molecular characteristics and GE profiling.

The classification of BC cases into different intrinsic subtypes based on GE profiles has started to significantly influence BC patients' diagnosis since the beginnings of the 2000's. The microarray-based GE profiling of BC cases enabled the definition of gene signatures such as the 50-gene set called the Prediction Analysis of Microarrays (PAM50) (Parker et al. 2009). The PAM50 gene signature allows the identification of different intrinsic BC subtypes and patients overall survival by measuring the expression of a 50 classifier genes and of 5 control genes by qRT-PCR assays which have been validated on formalin-fixed paraffin-embedded (FFPE) BC tissues (Kittaneh, Montero, and Glück 2013). The classification of BC cases based on this microarrays-based GE analysis and unbiased hierarchical clustering have enabled the identification of the main BC intrinsic subtypes (Sørlie et al. 2003) (**Figure 1.4-2**):

- (i) *The luminal BC subtypes*, including luminal A and B subgroups, is the major and most predominant subtype representing more than  $\frac{2}{3}$  of all BC cases. These subtypes are characterized by ER $\alpha$  positivity, low tumor grades, good prognosis, and most favorable clinical outcomes. As compared to other subtypes, the luminal cancers are generally characterized by the highest expression of ER $\alpha$  as well as other transcription factors such as Forkhead-box protein A1 (FoxA1) and GATA binding protein 3 (GATA3), and luminal cytokeratin. Specifically, the luminal A subgroup is characterized by more favorable prognosis and is less aggressive than the luminal B subgroup which shows a variable expression of the oncogene HER2 and lower expression of the PR (Mohammed et al. 2015).
- (ii) *HER2-overexpressing (HER2+) BC subtype*, which represents the second most frequent subtype of BC cases, is characterized by a high expression of HER2/ERBB2 marker. As compared to the other subgroups, the HER2+ BC subgroup is characterized by a high proliferative potential, higher tumor grades, and most often worse prognosis. Nonetheless, inside this group is present a high proportion (up to 50%) of cases that are also ER $\alpha$ +, showing, however, a minimal clinical response to treatments.
- (iii) *The Triple Negative Breast Cancers (TNBC) subtype*, also called *basal-like subtype*, is characterized by the complete negativity for PR ER $\alpha$ , and HER2 expression. This subgroup, representing up to 20% of all BC

caes, is characterized by the worst disease outcome, high histological grade, high proliferative index, and shortest overall survival.

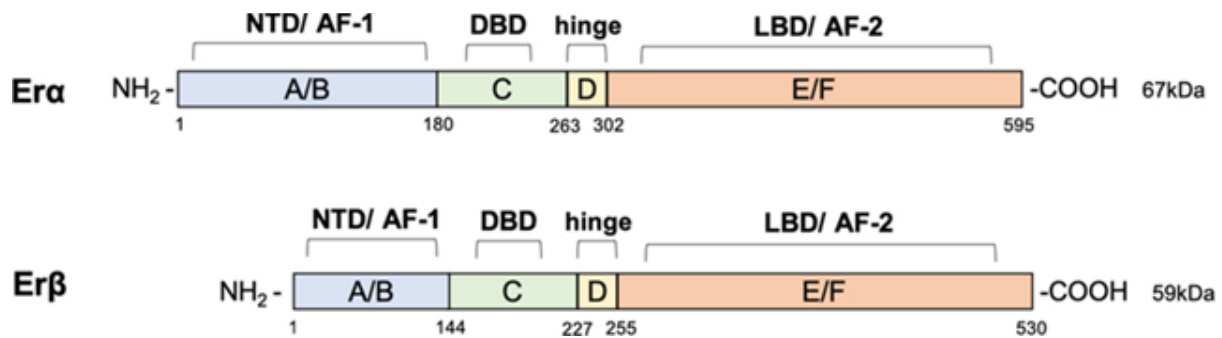
(iv) *The claudin-low breast cancer subtype*, which is defined by specific GE signatures, has recently emerged as an independent BC subtype that is characterized by low expression of cell-adhesion genes, high expression of epithelial-to-mesenchymal transition gene signatures, and stem-like features. This BC subtype is reported to be mostly triple-negative, and it is associated with poor prognosis. This subtype is characterized by a highly variable incidence ranging from 1.5 to 14% of BC cases (Fougnier et al. 2020).

(v) *The normal-like breast cancer subtype*, which is similar to the basal-like subgroup, is characterized by the high expression of genes characteristic of the basal epithelial cells and adipose tissues, and the low expression of genes characteristic of the luminal epithelial cells (Perou et al. 2000).

### 1.4.3 Estrogen receptors and their transcriptional activities

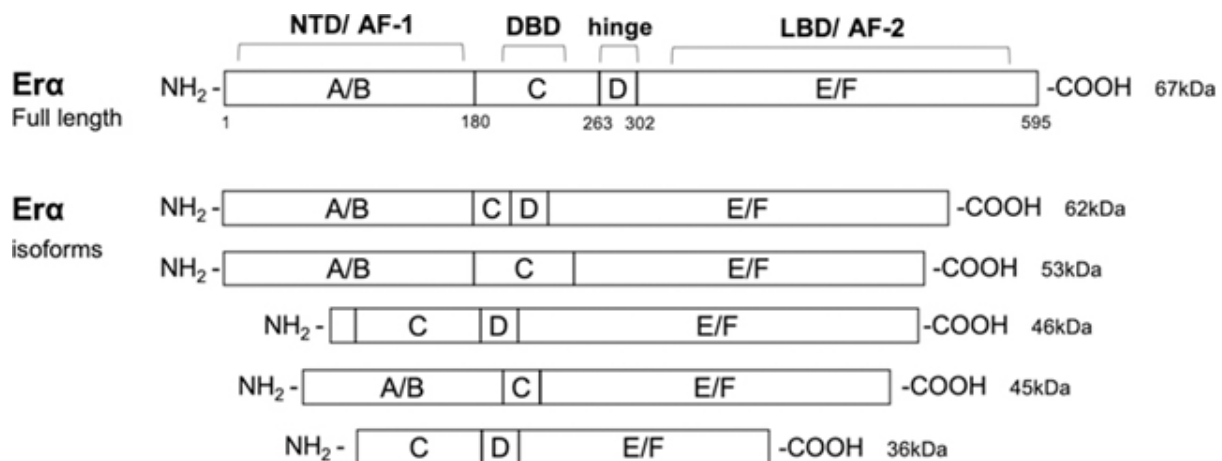
#### 1.4.3.1 Estrogen receptors structural properties

In 1958, the ER $\alpha$  was discovered by Elwood Jensen as the first receptor ever encountered for any hormone, by showing that reproductive female tissues were able to uptake estrogens from the circulation by binding to proteins (Jensen et al. 1968). It was later demonstrated that estrogen-bound receptors are able to migrate to the nucleus, where they could activate gene transcription (Jensen et al. 1968). Several years later, a second estrogen receptor, ER $\beta$ , highly similar to the firstly identified receptor (DNA-binding domain 95%, ligand-binding domain 55%), was described (Kuiper et al. 1996). More recently, a third type of estrogen receptors, membrane estrogen receptor, (also called G protein-coupled estrogen receptor, GPER1) was identified (Filardo and Thomas 2012) and its role in mediating fast estrogen responses and its implication in physiological and pathological processes has been extensively studied in human and mouse models (Molina et al. 2017). Estrogen receptors are structurally composed of various functional domains. ER $\alpha$  and ER $\beta$  are 595 (67kDa) and 530 (59kDa) amino acids length, respectively, and mainly differ by the N-terminal domain which is shorter in ER $\beta$  than in ER $\alpha$  (Schwabe and Teichmann 2004) (**Figure 1.4-3**). The N-terminal domain contains the AF-1 (A/B) domain which is involved in gene transcription transactivation, and the binding of target sequences which is mediated by a zinc finger domain (Kumar et al. 2011). The C-domain corresponds to the DNA-binding domain (DBD), which contributes to receptor dimerization and binding to specific sequences in the chromatin (Scheidereit et al. 1986). The D-domain is a hinge connecting C and E domains, and is involved in the binding of resting estrogen receptors to chaperons, co-chaperons and immunophilin proteins (Echeverria and Picard 2010). The C-terminal domain corresponds to the ligand binding domain (LBD, AF-2) and contains the estrogen binding area, along with binding sites for activators and co-repressors (Wärnmark et al. 2003).



**Figure 1.4-3:** Structural and functional organization of estrogen receptors. Functional domains of the full-length isoforms are labelled A to F. Both have 6 functional domains: NTD: A/B (AF-1 domain), DNA-Binding Domain (DBD, C), the hinge (D domain), Ligand Binding Domain (LBD, E/F, AF-2 domain). From (Fuentes and Silveyra 2019).

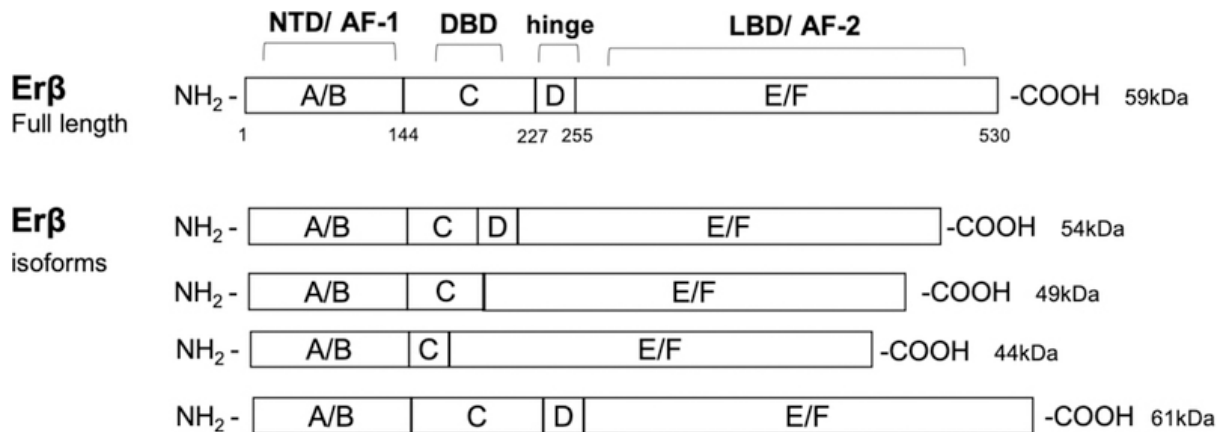
In humans, ER $\alpha$  is encoded by the *ESR1* gene located on chromosome 6, locus 6q25.1 (Gosden, Middleton, and Rout 1986). In addition to the full length isoform ER $\alpha$ 66, several shorter isoforms (62kDa, 53kDa, 46kDa, 45kDa, and 36kDa) have been described and result either from differential alternative promoter usage or from AS of internal exons (Ferro et al. 2003) (**Figure 1.4-4**). These short isoforms lack the AF-1 domain and therefore they cannot activate gene transcription. However, they have the ability to form heterodimers with the full length isoform ER $\alpha$ 66, where they could activate or inhibit its transcriptional activity (Penot et al. 2005). Specifically, the shortest isoform ER $\alpha$ 36, which lacks both AF-1 and AF-2 transcriptional activation domains, was reported to exert membrane-initiated signaling events upon binding of ligands (Y. Gu et al. 2014). On the other hand, the isoform ER $\alpha$ 46 was identified in the year 2000 as a new isoform of the human ER $\alpha$  lacking the N-terminal 173 amino acids present in the full length ER $\alpha$ 66 isoform (Flouriot et al. 2000). This isoform is encoded by a distinct transcript that lacks the the first coding exon (exon 1A) of the full length isoform, and has been identified as antagonizing the proliferative effects of ER $\alpha$ 66 isoform in MCF-7 cells BC cells by inhibiting the ER $\alpha$ 66 AF-1 activity, and its overexpression in proliferating MCF-7 cells provokes a cell cycle arrest in G0/G1 phase (Penot et al. 2005).



**Figure 1.4-4:** Depiction of estrogen receptor alpha (ER $\alpha$ ) isoforms. The domain organization of the full-length 595 amino acid ER $\alpha$  (67kDa), and truncated shorter isoforms (62kDa, 53kDa, 46kDa, 45kDa, and 36kDa) resulting from alternative splicing and/or alternate translation start sites are illustrated. Protein domains are labeled as A to F with numbering denoting the amino acid sequence number based on the full-length protein (595 aa). ER $\alpha$  domains: N-terminal (NTD,

*A/B domains, AF-1), DNA binding domain (DBD, C domain), hinge (D) domain, and C-terminal region containing the ligand binding domain (LBD, E/F domain, AF-2). From (Fuentes and Silveyra 2019).*

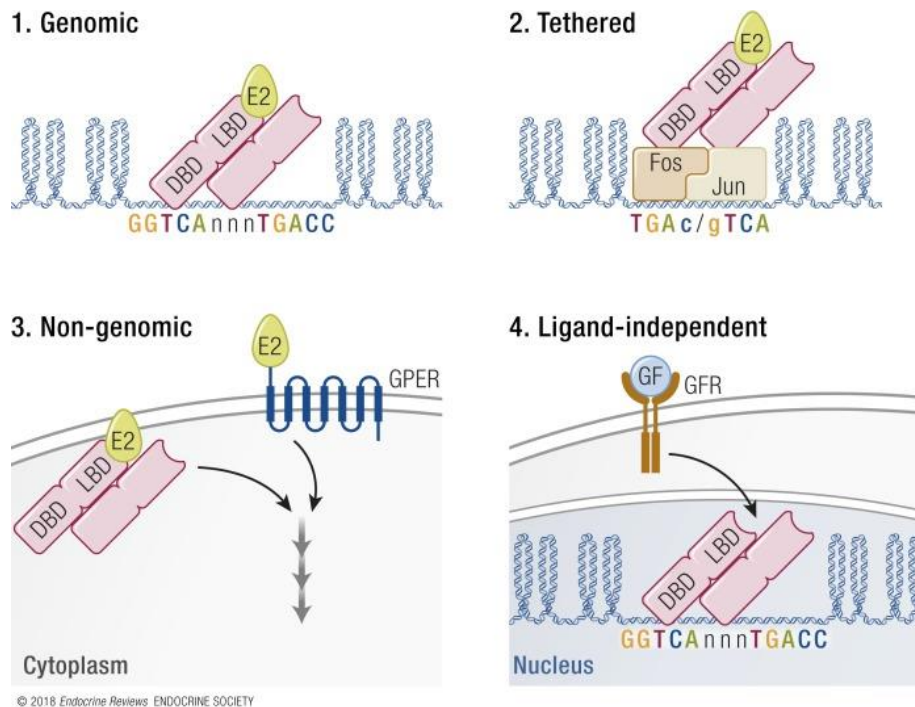
On the other hand, ER $\beta$  is encoded by the *ESR2* gene located on chromosome 14 (14q23-24), and is present in five known isoforms (59kDa, 54kDa, 49kDa, 44kDa, and 61kDa) (Enmark et al. 1997) (**Figure 1.4-5**). The full length and short ER $\beta$  isoforms differ mainly by the C-terminal LBD domain. The short ER $\beta$  isoforms that have no transcriptional activity were described to form dimers with ER $\alpha$  and therefore suppress its transcriptional activity (Vrtačnik et al. 2014).



**Figure 1.4-5:** Estrogen receptor beta (ER $\beta$ ) isoforms. Structural domain organization of the full length (530 amino acids, 59 kDa), short (54 kDa, 49 kDa, 44 kDa) and elongated (61 kDa) ER $\beta$  isoform. The reported isoforms result from AS or alternative translation start site usage of the full length isoform. ER $\beta$  domains: N-terminal (NTD, A/B domains, AF-1), DNA binding domain (DBD, C domain), hinge (D) domain, and C-terminal region containing the ligand binding domain (LBD, E/F domain, AF-2). From (Fuentes and Silveyra 2019).

### 1.4.3.2 Estrogen receptor alpha classical and non-classical genomic activities

Since the first review of estrogen receptors structure many years ago, the understanding of biological and molecular mechanisms underlying E2/ER $\alpha$ -mediated responses has greatly progressed. Moreover, the continuous development of cell culture and animal models has enabled closer study of both molecular and cellular details of E2/ER $\alpha$  signaling. Thus, the mechanisms of E2-ER $\alpha$  transcriptional activities have recently been revised. The next chapter will be organized in different sections describing in details the core understanding of ER $\alpha$  functions including the basic genomic mechanisms, expanded to include tethered, non-genomic, and ligand-independent mechanisms of ER $\alpha$  actions (**Figure 1.4-6**).

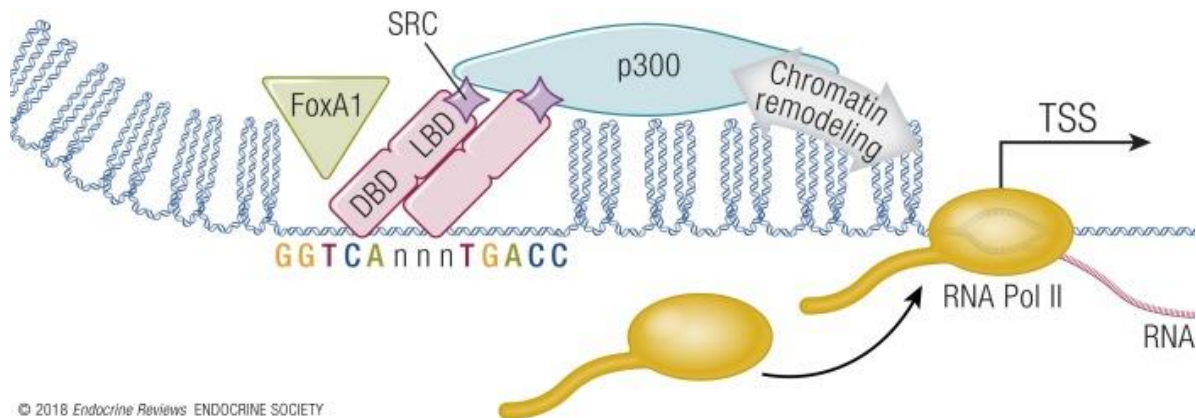


**Figure 1.4-6:** Variations in the basic genomic mechanisms of E2/ER $\alpha$ -response recently revised, resulting in four different mechanisms. (1), Genomic mechanisms involving the interaction of ER $\alpha$  with ERE DNA motifs. (2), the tethering mechanism of ER $\alpha$  involves the indirect interaction of ER $\alpha$  with DNA where ER $\alpha$  is tethered to the target DNA via other transcription factors like Fos/Jun dimer, in this example, which binds to its AP-1 DNA motif. (3), Non genomic signaling involves the extracellular E2 which initiates a rapid signaling cascade in the cytoplasm thus the response does not involve interaction with genomic features. These responses are mediated by other factors like membrane-associated estrogen receptor, or G protein-coupled Receptor (GPER1). (4), Ligand-independent signaling involves transduction of extracellular growth factor receptor pathways which initiates signaling cascades, such as the MAPK pathway. The GFR activation signal is perceived by ER $\alpha$  (PTMs) activating its transcriptional modulation of target genes, despite lacking E2. From (Hewitt and Korach 2018).

#### 1.4.3.2.1 Estrogen receptor alpha genomic and epigenomic activity

Physiological actions of estrogens and their receptors are principally mediated through the transcriptional regulation of target genes. The classical mechanism of ER $\alpha$  actions involves estrogens binding to receptors in the nucleus, after which, the receptor dimerizes and binds to specific DNA consensus motif called *Estrogen Responsive Elements* (ERE) that is composed of a minimal core sequence of 13bp palindromic inverted motifs “GGTCANNNTGACC”, located on target genes (Hewitt et al. 2012). Moreover, ER $\alpha$  has also the ability to bind a wide range of varying ERE motifs and many of estrogen-responsive genes contain non-consensus ERE sequences in their promoters and principally in enhancers (Coons et al. 2017) (**Figure 1.4-7**). The application of advanced techniques such as Chromatin Immunoprecipitation on chip (ChIP-chip) and ChIP-Seq experiments have enabled the global genome-wide profiling of ER $\alpha$ -chromatin interactions and characterized the genomic distribution of ER $\alpha$  binding sites in BC cell lines and primary tissues (Gilfillan, Fiorito, and Hurtado 2012; Cheung and Kraus 2010). Specifically, The Carroll laboratory performed the first ER $\alpha$  genome-wide binding profiling using E2-stimulated MCF-7 BC cell lines (J. S. Carroll and Brown 2006). In this study,

the authors identified 3,665 ER $\alpha$  binding sites and 3,629 RNA Pol II binding sites (J. S. Carroll and Brown 2006). Interestingly, while the vast majority of RNA Pol II bindings were located at target gene promoters, only 4% of ER $\alpha$  binding sites located in these regions, while the remaining 96% of binding sites were located distal from promoters, suggesting that enhancers are the main interaction sites of ER $\alpha$  with chromatin (J. S. Carroll and Brown 2006).



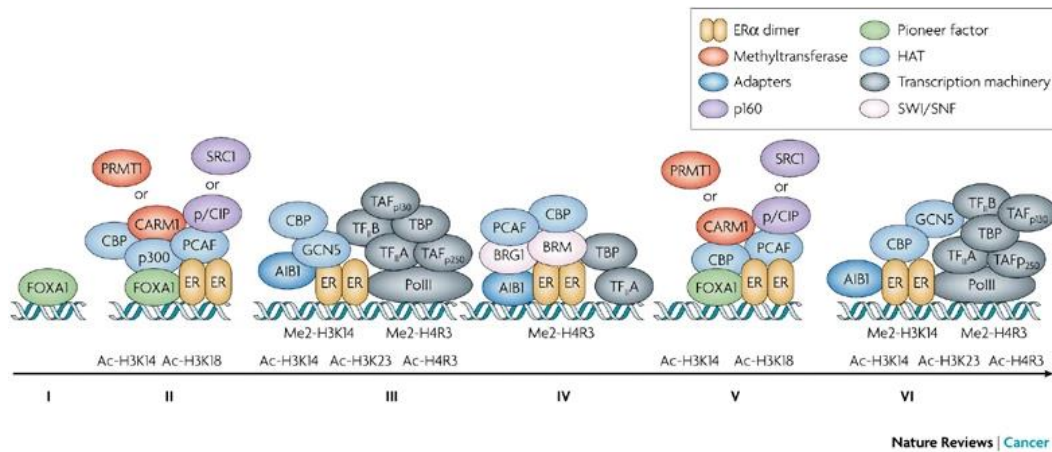
**Figure 1.4-7:** Basic and classical mechanisms of ER $\alpha$ -mediated regulation of transcription. ER $\alpha$  access to target gene is, in part, controlled by chromatin states. This is controlled by other pioneer factors such as FoxA1 which provides more open chromatin, facilitating access of ER $\alpha$  to target ERE DNA motifs (mainly at enhancers). Exposure to E2 recruits SRC complex and mediates the interaction with p300 which then induces chromatin remodeling facilitating access of RNA Pol II to ER $\alpha$  target genes. From (Hewitt and Korach 2018).

#### 1.4.3.2.2 Estrogen receptor alpha tethering (indirect genomic) mechanism

In addition to the ERE motifs that mediate the basic mechanisms of ER $\alpha$ -mediated response, the ER $\alpha$  transcriptional activity can be enhanced by pioneer factors by a tethering mechanism. The most known factor is FoxA1 that provides chromatin accessibility by binding and opening chromatin, which leads to enhanced ER $\alpha$ -ERE interactions at appropriate sites in the cell (Hurtado et al. 2011) (**Figure 1.4-8**). Furthermore, ER $\alpha$  binding to the chromatin correlates with the expression of FoxA1 in BC and this later was identified as a key factor influencing the differential interaction of ER $\alpha$  with chromatin (Hurtado et al. 2011). Notably, siRNA-mediated silencing of FoxA1 in tamoxifen-resistant BC cell lines also significantly inhibited cell growth and reduced the binding signal of ER $\alpha$  to chromatin. The distribution of ER $\alpha$  binding to chromatin was independent of ligand but dependent on FoxA1 expression (Hurtado et al. 2011). More specifically, Hurtado and colleagues, by analysing ChIP-seq in MCF-7 cell lines, identified 79,651 FoxA1 peaks in cells maintained in estrogen enriched medium, in addition to 14,059 ER $\alpha$  binding sites, of which 52% overlapped with FoxA1 peaks. In hormone-starved conditions, the silencing of FoxA1 reduced the number of E2-induced ER $\alpha$  binding sites and globally affected the E2-mediated control of transcriptome where the expression of more than 95% of E2-regulated genes were altered (Hurtado et al. 2011). Similarly, the binding motifs of other pioneer transcription regulators were enriched within ER $\alpha$  binding sites. The enrichment of Activator Protein 2 gamma (AP2 $\gamma$ ) binding motif in ER $\alpha$  binding sites was observed in a study where ER $\alpha$  binding correlated with time-course analysis of the E2-mediated gene expression of both coding and noncoding genes (Cicatiello et al. 2010). Other

groups by performing a ChIP-Seq against AP2 $\gamma$  in E2-treated MCF-7 cells identified 25,567 AP2 $\gamma$  peaks (Tan et al. 2011). Among the 14,468 E2-induced ER $\alpha$  binding sites identified, 50% overlapped with AP2 $\gamma$  peak signal and colocalized with FoxA1 binding. Importantly, silencing AP2 $\gamma$  in this same experimental set up reduced the expression of 60% of estrogen-induced genes and decreased both ER $\alpha$  and FoxA1 bindings at target sites. Similarly, the silencing of FoxA1 also decreased AP2 $\gamma$  binding, suggestive of a cooperative binding of these pioneer factors to mediate ER $\alpha$  transcriptional activity (Tan et al. 2011). In addition, other ER $\alpha$ -cooperating factors were identified, such as GATA3, a frequently mutated transcription in breast tumors, was identified as a pivotal factor enabling chromatin accessibility at enhancers involved in ER $\alpha$ -mediated transcription (Theodorou et al. 2013). Authors performed an extensive analysis of GATA3 ChIP-Seq data using unstimulated or estrogen-treated MCF-7 cells, and explored the effects of GATA3 silencing on ER $\alpha$  genomic occupancy. GATA3 silencing resulted in a significant redistribution of ER $\alpha$ -cofactors such as p300 occupancy and histone markers such as H3K27ac, and H3k4me1 modifications in absence of ligands (Theodorou et al. 2013). In addition, the identified set of ER $\alpha$  binding sites observed upon GATA3 silencing were characterized by an increase of ER $\alpha$  binding upon exposure to E2 and were accompanied by changes in gene expression. Importantly, when GATA3 was silenced, chromatin loops at the *TFF* locus involving ER $\alpha$ -bound enhancers occurred independently of ER $\alpha$  binding, suggestive of a licensing role mediated by GATA3 for E2/ER $\alpha$ -mediated interactions with enhancers (Theodorou et al. 2013).

In addition, the binding of estrogens induces a conformational change of the receptors, and this allows a number of coactivator proteins to be recruited (Yi, Wang, and O'Malley 2019). To date, a variety of coactivators have been identified. The most well-documented coactivator complexes include the steroid receptor coactivator proteins (SRCs), CREB binding protein (CBP), p300, and coactivator associated arginine methyltransferase 1 (CARM1) (Yi, Wang, and O'Malley 2019). The recruitment of coactivator complexes impacts the chromatin states of target genes thereby inducing chromatin decondensation, thus facilitating the access of the transcriptional machinery and subsequently target gene transcription (X. Liu et al. 2008). For example, using ChIP experiments, the Brown and the Gannon laboratories have shown *in vivo* the dynamic on/off cycling of ER $\alpha$  binding and its coactivator multiprotein complex p160 at the promoter of the cathepsin (*CTDS*) and Trefoil Factor 1 (*TFF1*) promoters (Métivier et al. 2003) (**Figure 1.4-8**). This mechanism is abrogated by applying ER $\alpha$  antagonist drugs, such as tamoxifen, which induces a conformational change of the protein that blocks access to coactivator complexes and in parallel fosters co-repressors binding, which results in the inhibition of ER $\alpha$ -mediated transcription (Celik, Lund, and Schiøtt 2007). Furthermore, other factors such as the PTMs of ER $\alpha$  protein, particularly phosphorylation at different residues, were described to influence the recruitment of specific coactivators and are also involved in the activation process of ER $\alpha$  in absence of ligands (Le Romancer et al. 2011).



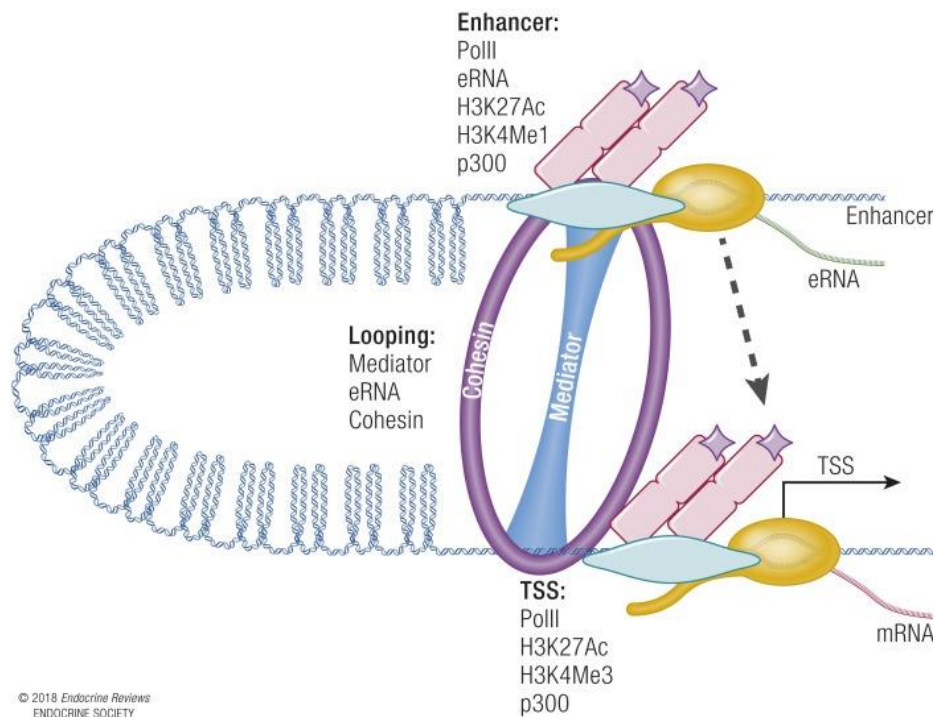
**Figure 1.4-8:** ER $\alpha$  cycles on and off the chromatin in cycles of approximately 45 minutes. First, the pioneer factors bind to the heterochromatin (I) to determine where in the genome ER $\alpha$  can bind. (II), After oestrogen stimulation, ER $\alpha$  associates with the TFF1 promoter and subsequently histone acetyl transferases (HATs) are recruited and acetylation of the local histones occurs. Concurrently, binding of histone methyltransferases (HMTs) and p160 co-activators also increases at the TFF1 promoter. (III), RNA Polymerase II (Pol II) and the transcription machinery lag behind ER $\alpha$  association and correlate with increases in histone methylation. (IV), As the levels of ER $\alpha$  protein on the chromatin decreases, the remaining ER $\alpha$  is associated with components of the SWI/SNF complex, low levels of HATs and AIB1 as well as minimal transcriptional machinery. (V), As the ER $\alpha$  levels begin to increase in subsequent cycles there is a concomitant increase in HMTs, transcription factors, HATs (not p300), adaptor proteins and p160 coactivators. As ER $\alpha$  reaches its maximum load on the chromatin, the HMTs and p160 proteins decrease to leave just the HATs, adaptor proteins and transcription machinery (VI). From (K. A. Green and Carroll 2007).

### 1.4.3.3 Estrogen receptor alpha-bound enhancers and differential chromatin interactions

In the last years, strong lines of evidence started to appreciate that most (96%) of ER $\alpha$  binding sites on chromatin are distal from TSSs and are located on enhancer (often >100kb away from target transcripts), and these are enhanced and expanded by E2 stimulation, which results in the formation of super-enhancers (Hewitt et al. 2020). This is particularly important since the active chromatin-enriched enhancers and super-enhancers are intricate regulatory DNA elements that drive cell differentiation processes and often control the expression of cell-specific transcription factors, which are important for the shaping of every organ and tissue (Blanco et al. 2020). Moreover, chromatin features at enhancer and super-enhancer regions provide the permissive landscape required for the differential access of gene expression regulators and signaling molecules to drive cell-specific gene expression programmes, in time and space, and their misregulation is a key contributor to carcinogenesis (Blanco et al. 2020). For instance, it was reported that enhancers are decorated with specific chromatin marks earlier to the onset of the cell decision-making process, which suggests that chromatin states at enhancers are crucial for the subsequent differentiation process and enables the lineage-specific gene expression programmes to take place (H. Xu et al. 2020). The interactions between distal ER $\alpha$ -chromatin binding sites and target TSSs occur via a chromatin looping mechanism, which involves the synthesis of small enhancer RNAs from the distal sites, followed by the assembly and binding of CTCF and cohesin complexes, further facilitating enhancer-promoter interactions (J. Wang et al. 2016) (**Figure 1.4-9**). The enhancer



activation process is initiated by binding of p300 which exhibits its acetyltransferase activity leading to increased transcriptional activity at the promoter of target genes. In contrast, the mechanisms of ER $\alpha$ -mediated transcriptional repression are not well understood. However, it was reported that the recruitment of the polycomb complexes to ER-bound enhancers could inhibit the interaction with p300/CBP complexes, preventing transcriptional progression (Piunti and Shilatifard 2016).



**Figure 1.4-9:** Model of chromatin looping to facilitate interaction between enhancers and promoters/TSSs. RNA PolIII, enhancer RNA (eRNA) transcription, acetylation of histone H3 lysine 27 (H3K27Ac), monomethylation of histone H3 lysine 4 (H3K4Me1), and p300 are found at enhancers. TSSs have PolIII, H3K27Ac, trimethylation of histone H3 lysine 4 (H3K4Me3), and p300. Cohesin/mediator form a looping structure that facilitates interaction between the enhancer and TSSs (dashed arrow). From (Hewitt and Korach 2018).

ER $\alpha$  binding to enhancers was also reported *in vivo* in mice hormone-responsive uterine tissue (Hewitt et al. 2020). These ER $\alpha$ -bound super-enhancers were found to be formed prior to the exposure to estrogens at the onset of puberty, and associated with genes encoding for critical developmental factors, including retinoic acid receptor alpha (RARA), and homeobox D (HOXD) (Hewitt et al. 2020). Moreover, authors demonstrated by the chromatin conformation capture (Hi-C) techniques that the genes which associated with the ER $\alpha$ -bound enhancers are located at the anchors of chromatin loops and are regulated by estrogens. Importantly, the expression of the enhancer-associated genes acquired a strong dependency on ER $\alpha$  binding after reproductive maturity (Hewitt et al. 2020). These observations suggested that enhancers assembly and their binding by ER $\alpha$  during the pubertal process are pivotal factors for estrogen-mediated regulation of key genes mediating uterine functions, such as transforming growth factor beta (TGFB) and LIF interleukin-6 family cytokine signalling pathways (Hewitt et al. 2020). Furthermore, Hi-C interaction maps enabled a detailed description of the multilevel genome organization and the 3D chromatin interactions in different contexts.

In the context of BC, Clark and colleagues extensively characterized the effects of ER $\alpha$ -enhancers binding on the 3D chromatin topological interactions in both endocrine-sensitive and endocrine-resistant BC cellular models (Achinger-Kawecka et al. 2020). Authors identified a number of differential chromatin interactions between parental wild type MCF-7 cells, tamoxifen-resistant (TAMR), and fulvestrant-resistant (FASR) cellular systems (Achinger-Kawecka et al. 2020). Specifically, using diffHiC at 20kb resolution, authors identified 981 differential chromatin interactions between wild type MCF-7 cells and TAMR cells, and 2596 differential chromatin interactions between MCF-7 cells and FASR model. More precisely, authors observed that chromatin interactions are more often lost with the development of fulvestrant resistance as compared to MCF-7 cells, (62% are MCF-7-specific), while differential chromatin interactions were often gained with the development of tamoxifen resistance (42% TAMR-specific). Importantly, the vast majority of differential chromatin interactions in TAMR model were completely absent in FASR model, potentially consistent with the different modes of action between tamoxifen and fulvestrant and the different pathways underlying the development of endocrine resistance (Achinger-Kawecka et al. 2020). Furthermore, authors explored whether the identified differential (lost/gained) chromatin interactions include direct enhancer-promoter interactions, by integrating chromatin interactions data with chromatin state information, based on five ChIP-Seq markers, including H3K27ac, H3K4me1, H2AZac, and H2K4me3 chromatin states. Interestingly, regardless of the endocrine resistant model (TAMR or FASR), the differential chromatin interactions were enriched for active enhancers and promoters, as well as CTCF sites. However, gained chromatin interactions showed more enrichment for active enhancer marks (H3K4me1, H3K27ac) as compared to lost interactions. Similarly, there was an increased enrichment of active promoter marks (H2K4me1) in TAMR and FASR as compared to parental MCF-7 cells. Moreover, authors further explored whether differential chromatin interactions present in TAMR and FASR associate with changes in gene expression of genes located at anchors of these differential chromatin interactions as compared to their expression levels in parental, wild type, MCF-7 cells. Interestingly, genes located on anchors of differential chromatin interactions were related to important pathways known for endocrine resistance development, such as estrogen response pathway, and cancer, such as EMT and angiogenesis. More precisely, differential chromatin interactions identified in FASR model overlapped with promoter regions of 2069 genes, and loss of these interactions associated with significant decreases in the expression of 219 genes, and gained chromatin interactions in this model associated with significant increase in the expression of 170 genes. Similarly, in the TAMR cellular model, 500 genes associated with differential chromatin interactions, among which 50 genes associated with loss in chromatin interactions were downregulated, while gained chromatin interactions associated with increased expression of 21 genes (Achinger-Kawecka et al. 2020).

From a topology perspective, in both TAMR and FASR cellular systems, gained and lost chromatin interactions associated with differentially expressed genes, with most of downregulated genes located at the anchor of lost interactions, while upregulated genes located at the ectopic/gained chromatin interactions in both systems (Achinger-Kawecka et al. 2020). Despite the fact that differential chromatin interactions either in TAMR or FASR were related to similar gene ontology terms including transcription, cell-cell adhesion, and

G2/M phase transition, some differential chromatin interactions were specifically enriched in each cellular system. For instance, in TAMR cellular model, some of the differential chromatin interactions were exclusively related to ERBB2 signalling pathway, response to estradiol and Wnt signalling pathways, while FASR-specific chromatin interactions were mainly related to apoptosis, MAPK cascade, cell division and migration (Achinger-Kawecka et al. 2020). Finally, authors observed a significant differential enrichment of binding motifs for a large number of transcription factors at anchors of the differential chromatin interactions between the investigated cellular models. In particular, ESR1, MYC, CTCF, NR2F1, FOXA1, and PR were among the most enriched binding sites at anchors of differential chromatin interactions in parental MCF-7 cells. In the TAMR model, chromatin interactions lost principally occurred at regions of ERE, SOX2, FOXA1, and HOX cluster binding motifs, and gained interactions in TAMR were present at ZNF143, OCT4, FOXA1, and RUNX2 binding sites, in accordance with previously reported results showing that tamoxifen resistance results in aberrant ER $\alpha$  signaling through the genome (M. Fan et al. 2006). In fulvestrant resistance, lost chromatin interactions were mainly present at SOX6, NRF2, and ATF3 binding motifs, while gained interactions were enriched for OTX2, and SMAD4 binding sites. In both endocrine-resistance models, differential chromatin interactions occurred commonly at cMYC binding motifs, as previously reported (A. R. Green et al. 2016).

#### 1.4.3.4 Estrogen receptor alpha indirect non-genomic mechanisms

In addition to genomic actions of E2/ER $\alpha$  signaling pathway, non-genomic actions are common to steroid hormones. The non-genomic actions of E2/ER $\alpha$  signaling pathway have been known for a long time and are associated with the rapid activation of various protein-kinase cascades which leads to indirect changes in gene expression due to phosphorylation of transcription factors (Y. Li et al. 2010). The binding of estrogens to membrane-bound estrogen receptor GPER1 and certain ER $\alpha$  and ER $\beta$  variants (Barton et al. 2018) can mobilize intracellular calcium, stimulate the activity of adenylate cyclase and cyclic adenosine monophosphate (cAMP) production, activation of the mitogen-activated protein kinase (MAPK) signaling pathway, activation of phosphoinositol 3-kinase (PI3K) signaling pathway, as well as the activation of membrane receptor tyrosine kinases (RTKs) (Ajj et al. 2013). Several examples of transcription factors regulated by these mechanisms include Elk1, CREB1 (Cavalcanti et al. 2015), CCAAT-enhancer-binding protein beta (C/EBP $\beta$ ), the NF- $\kappa$ B complex (Fox, Andrade, and Shupnik 2009), and members of the signal transducer and activator of transcription (STAT) family (Furth 2014). Thus, by its non-genomic actions, ER $\alpha$  and ER $\beta$ , indirectly regulate the expression of indirect target genes.

#### 1.4.3.5 Estrogen receptor alpha ligand-independent activation

Apart from being the main mediator of estrogens actions in various tissues and cancers, ER $\alpha$  displays an estrogen-independent activity in its unliganded status (Bennesch and Picard 2015). The first evidence of hormone-independent activation of ER $\alpha$  was reported in 1991 where Power and colleagues demonstrated that the receptor can be activated by neurotransmitters like dopamine; via cAMP/PKA activation (Ince, Montano, and Katzenellenbogen 1994); in absence of ligand (Power et al. 1991). Similarly, other studies demonstrated the activation of ER $\alpha$  by epidermal growth factor (EGF) *in vivo* in ovariectomized mice (Curtis et al. 1996),

and by insulin-like growth factor (IGF) (Kato et al. 1995). These was further supported by the analysis of ER $\alpha$  ChIP-Seq from uterine tissue of ovariectomized mice which showed over 5,000 regions of constitutive ER $\alpha$ -DNA interactions (Hewitt et al. 2012). However, In spite of the solid evidence demonstrating the transcriptional activation of ER $\alpha$  in the absence of ligands, it has proven difficult to characterize the mechanisms underlying this process. Studies on transfected cells reported relevance for selected kinases such as MAPK (D. Chen et al. 2000), PKA (Al-Dhaheeri and Rowan 2007), and ras/ERK (Klotz et al. 2002) for ER $\alpha$  activation. Furthermore, studies using mutant ER $\alpha$  showed that in each cell system, specific ER $\alpha$  phosphorylation sites are required for the unliganded ER $\alpha$  activity (Sheeler, Singleton, and Khan 2003), suggesting that the activation of unliganded ER $\alpha$  may fulfill cell-specific functions. Although PTMs of ER $\alpha$  were reported relevant for the transcriptional activation of unliganded ER $\alpha$ , it was however reported, in some cases, that the phosphorylation of ER $\alpha$  was insufficient by itself for the transcriptional activation of unliganded ER $\alpha$ , even if allowed the recruitment of coregulator complexes and splicing factors (Masuhiro et al. 2005).

One of the mechanisms of the clinical resistance to tamoxifen is the upregulated signaling of growth factor pathways such as EGF, IGF1, both trigger an alteration of the ER $\alpha$  and co-regulators PTMs code (Murphy, Seekallu, and Watson 2011). One of the most studied PTMs of ER $\alpha$  is the phosphorylation at serine 118 (Ser118), which was a common target for many signaling pathways (Lannigan 2003). Both E2 and growth factor signaling pathways such as EGF, IGF1 and prolactin pathways stimulate the phosphorylation of Ser118 residue (González et al. 2009). The mitogen activated protein kinase (MAPK), which is activated by multiple growth factor receptor pathways, can phosphorylate Ser118 in a ligand independent manner both *in vitro* (Kato et al. 1995) and *in vivo* (Joel et al. 1998). The phosphorylation of ER $\alpha$  at Ser118 residue was reported to be involved in protein turnover via a proteasome-mediated mechanism (Grisouard et al. 2007), and is important for protein dimerization (Sheeler, Singleton, and Khan 2003), and for ER $\alpha$  interaction with coactivator complexes in absence of ligand (Dutertre and Smith 2003). Other important ER $\alpha$  PTMs were reported. The phosphorylation of serine 167 (Ser167) is important for enhanced BC cell proliferation (Yamnik et al. 2009), and is stimulated by pp90rsk (Joel et al. 1998), and by AKT (Campbell et al. 2001), whose increased phosphorylation associated with poor clinical outcomes in patients treated with tamoxifen (Kirkegaard et al. 2005). ER $\alpha$  Serine 305 (Ser305) phosphorylation by a PKA1-mediated mechanism was also reported to be important for ER $\alpha$  interactions with DNA and coactivators (Zwart et al. 2007), and experimental data suggested that it could have a role in tamoxifen resistance in BC cells (Holm et al. 2009). Other functional individual and collective ER $\alpha$  PTMs, their functional roles and implications in resistance to tamoxifen in BC, and candidate signaling pathways and kinases involved are reviewed elsewhere (Murphy, Seekallu, and Watson 2011).

More recently, the transcriptional activity of unliganded ER $\alpha$  became relevant in the context of BC, especially when investigating clinical resistance to endocrine therapies. In 2014, Caizzi and colleagues performed the first genome-wide analysis of unliganded ER $\alpha$  chromatin interactions in MCF-7 cells. In this work, authors performed a ChIP-Seq against ER $\alpha$  in hormone-deprived MCF-7 cells, transfected with control or ER $\alpha$ -

specific siRNA, to determine the genome-wide chromatin binding of unliganded-ER $\alpha$ . The analysis of this ChIP-Seq data led to the definition of unliganded ER $\alpha$  cistrome in MCF-7 cells. Specifically, 4,232 unliganded-ER $\alpha$  binding sites were identified (Caizzi et al. 2014). The identified binding sites were almost all contained within the ER $\alpha$  cistrome identified in MCF-7 cultured in full medium (FM) or after E2-treatment and showed a similar genomic distribution as compared to the other conditions, with a prevalence of intergenic location. Furthermore, siRNA-mediated silencing of ER $\alpha$  strongly reduced ChIP-Seq signal at the unliganded-ER $\alpha$  binding sites, confirming that the 4,232 ER $\alpha$  binding sites in absence of hormone are specific. Moreover, gene ontology analysis showed that these binding events associate to genes enriched in development, cell differentiation, and morphogenesis, while E2-specific and FM-specific ER $\alpha$  binding sites were associated to genes enriched in different terms such as metabolism, lipid metabolism, and biosynthesis terms, suggesting that unliganded-ER $\alpha$  binding sites may have a different role than that of E2-induced binding (Caizzi et al. 2014). In addition, transcription factor binding sites analysis on top significant peaks revealed a significant enrichment for different factors such as FOXA1, GATA3, NR2F2 and AP2 $\gamma$ . The identified unliganded-ER $\alpha$  binding sites were confirmed in independent datasets and in different BC cell lines, and further validated in human breast tumor samples. Authors confirmed the transcriptional effects of the unliganded-ER $\alpha$  binding sites by performing a paired-end RNA-seq experiment consisting of silencing ER $\alpha$  in absence of ligand. This revealed changes in the expression of hundreds of protein-coding and non-coding genes, mainly related to cell growth and survival and to the maintenance of the epithelial phenotype, further supporting the functional role of unliganded ER $\alpha$  activity.

#### **1.4.3.5.1 Estrogen receptor alpha activity in the maintenance of luminal gene expression**

ER $\alpha$  has been accepted as a biomarker of differentiated forms of BC (Parl et al. 1984) and is expressed in non-aggressive BCs while its expression is lost in more invasive and aggressive forms of the disease (Lari and Kuerer 2011). In vitro experiments demonstrated the opposing effects of ER $\alpha$  on the EMT process (Guttilla, Adams, and White 2012), and in BC cells cultured in absence of hormones, the depletion of ER $\alpha$  induces a response similar to EMT, by activating the expression of mesenchymal genes and growth-inhibiting pathways (Caizzi et al. 2014; Al Saleh, Al Mulla, and Luqmani 2011). In other cellular systems, the re-expression of ER $\alpha$  was able to induce the re-appearance of epithelial gene expression (Fortunati et al. 2010).

The E2-ER $\alpha$  signaling pathway opposes to the onset of EMT process by inhibiting the expression of EMT-mediating TFs and upregulating epithelial markers such as E-cadherins (Guttilla, Adams, and White 2012). The ligand-independent transcriptional activity of ER $\alpha$ , together with other TFs such as Forkhead box protein A1 (FoxA1) and Activating enhancer binding Protein 2  $\gamma$  (AP-2 $\gamma$ ) is important for maintaining the luminal phenotype and blocking the EMT of BC cells (Caizzi et al. 2014). The stable silencing of ER $\alpha$  expression by application of short-hairpin (shRNA) lentiviral particles in ER $\alpha$ -positive MCF-7 BC cells, strongly induced cellular phenotypic changes such as enhanced proliferation, migration and invasion, accompanied by significant changes in gene and protein expression of several markers typical of EMT (Al Saleh, Al Mulla, and

Luqmani 2011; Bouris et al. 2015) including EGFR and HER2 receptor tyrosine kinases and various extracellular matrix effectors.

#### **1.4.3.5.2 ER $\alpha$ activity in post-transcriptional regulation of gene expression**

Hormone signalling pathways such as the E2/ER $\alpha$  signalling pathway have been widely studied for their effects on transcription, but their roles in the post-transcriptional control of GE have recently been suggested, but not rigorously explored. However, a number of studies reported the possible implications of the E2/ER $\alpha$  axis in the process of AS. For instance, a clear link between steroid-regulated transcription and AS has been established (Auboeuf et al. 2002). Three years later, a collection of at least 25 proteins have been described to have a transcription coregulator and AS activity (Auboeuf et al. 2005). Among the identified proteins involved in steroid-regulated AS are the U2AF65-related proteins, CAPER $\alpha$  and CAPER $\beta$ , ASC-1, ASC-2, and CoAA, several of which are known to interact with ER $\alpha$  (Dowhan et al. 2005). Furthermore, the E2/ER $\alpha$  pathway has been shown to control the production rate of cyclin D1 mRNA by co-transcriptionally controlling the splicing level of its promoter-proximal intron (Bittencourt et al. 2008). In addition, a one study has reported the effects of the growth factor receptor-induced ER $\alpha$  phosphorylation on splicing, where MAPK-induced phosphorylation of ER $\alpha$  at Ser118 residue mediates its interaction with the U2 snRNP-related splicing factor SF3a and promotes exon skipping of target genes (Masuhiro et al. 2005). Moreover, BC subtypes are carefully characterized for prognosis and therapy purposes and represent the basis of patients stratification due to their different biological behaviors and response to treatments, which are underlined by their specific and distinct gene expression profiles (Sørliie et al. 2003). Interestingly, it was recently shown that not only gene expression profiles but also the expression pattern of transcript isoforms differs among ER $\alpha$ -positive and triple negative (TNBC) subtypes, and is capable of distinguishing these tumor types with higher fidelity than the standard gene expression profiles (Stricker et al. 2017). The different transcript isoforms expression patterns were a consequence of differential promoter usage, AS, and alternative 3'UTR polyadenylation sites usage which were differentially regulated in the investigated BC subtypes (Stricker et al. 2017). More importantly, a significant number of RNA processing factors were differentially expressed between tumor subtypes and are regulated by ER $\alpha$  (Stricker et al. 2017). These and other studies started to point evidence for the implication of ER $\alpha$ , and ER $\beta$  notably, in the post-transcriptional control of GE, either directly, by forming complexes with splicing factors and with certain components of the spliceosomal machinery, or indirectly, where ER $\alpha$ , notably ER $\beta$ , can control the post-transcriptional events by promoting or inhibiting the expression of the key players of this process, including miRNAs, RBPs, and RBP-associated lncRNAs.

##### **1.4.3.5.2.1 Estrogen Receptor Alpha is directly involved in the post-transcriptional control of gene expression**

The nuclear interactome proteins of the two nuclear receptors ER $\alpha$  and ER $\beta$  in BC cells nuclei were previously identified by Tandem Affinity Purification (TAP) techniques (Tarallo et al. 2011). The two categories of ER $\alpha$ - and ER $\beta$ -interacting proteins identified showed their ability to associate *in vivo* with either hormone-activated ER $\alpha$  and ER $\beta$  receptors (Tarallo et al. 2011). The computational analysis of these two datasets by Nassa and

colleagues in 2011 further supported the possible direct implication of ER $\alpha$  and ER $\beta$  in the post-transcriptional control of GE (Nassa et al. 2011). Notably, in this study, the computational analysis of ER $\alpha$ -interactome revealed three different clusters of interacting proteins, among which a cluster that was specifically formed by components of the splicing machinery such as U1, and U2 snRNPs, and other related factors such as the U2 Small Nuclear RNA Auxiliary Factor 2 (U2AF2), FUS RNA Binding Protein (FUS), and the RNA Binding Motif Protein X-Linked (RBMX), which are known for their dual role in controlling GE at both transcriptional and post-transcriptional levels, in addition to Elongation Factor Tu GTP Binding Domain Containing 2 (EFTUD2), and Small Nuclear Ribonucleoprotein D1 Polypeptide (SNRPD1), two important factors for spliceosome assembly (Little and Jurica 2008), and Y box binding protein 1 (YBX1), which is known to regulate alternative splice sites selection through its interaction with SR proteins and on the other hand it promotes cell proliferation of BC cells through activation of cell cycle-related genes (Raffetseder et al. 2003). In addition, two other distinct clusters of ER $\alpha$ -interacting proteins were identified, of which a cluster that comprised chromatin modifying enzymes such as HDAC1, HDAC2, and a third cluster included the components of the translational initiation complex and 3' polyA site interacting proteins such as PABPC1, while others proteins are involved in the 5' cap recognition such as EIF4A1 and EIF4G1 (S. Fan et al. 2010). These studies shed the light of the direct implication of ER $\alpha$  and ER $\beta$  in the AS process. Two years later, a study by Bhat-Nakshatri and colleagues reported the functional role of the interplay between E2/ER $\alpha$  axis and the AKT signaling pathway, which phosphorylates ER $\alpha$ , and evidenced their impacts on AS in BC (Bhat-Nakshatri et al. 2013). In study, authors identified, by using microarray analysis, a total of 463 AS events targeting 154 genes were significantly affected by E2 treatment, 18 of which were experimentally validated by qRT-PCR (Bhat-Nakshatri et al. 2013). Interestingly, among the 154 genes that underwent AS changes under E2 treatment, 89 genes (60%) contained intergenic binding sites for ER $\alpha$ , and 80% of which did not show any changes in their transcription rate, suggestive of a licencing role of ER $\alpha$  for the local recruitment of other factors such as histone modifying enzymes that favor AS regulation. Furthermore, the E2-induced AS changes in MCF-7 cells were influenced by AKT-induced phosphorylation of ER $\alpha$  and were reported to play a role in the anti-estrogen response to tamoxifen and fulvestrant in BC. Among the E2-induced alternatively spliced genes are the Fas cell surface receptor death (FAS/CD95), the fibroblast growth factor receptor 2 (FGFR2), and axin 1 (AXIN1). Specifically, E2 through AS regulation induced the mRAN abundance of FGFR2 C1 isoform while decreasing that of C3 isoform. These E2-induced AS changes of FAS and FGFR2 genes in MCF-7 BC cells correlated with resistance to FAS-activation induced apoptosis and to response to keratinocyte growth factor (KGF), respectively. The E2/ER $\alpha$ -induced overproduction of FGFR2 C1 isoform associated with resistance to tamoxifen in MCF-7 cells, while the altered KGF response correlated with an ER $\alpha$ -dependent isoform switching events, primarily associated with resistance to fulvestrant (Bhat-Nakshatri et al. 2013). These and other results (Bhat-Nakshatri et al. 2008) which show that the E2/ER $\alpha$  axis and its crosstalk with other signaling pathways such as the AKT pathway controls the expression of many splicing factors further support E2-mediated AS in BC.

In addition, on the basis of the study by (Tarallo et al. 2011), a study by (Dago et al. 2015) further investigated in depth the differences in the early transcriptional events and RNA splicing patterns induced by E2 in wild type MCF-7 cells expressing ER $\alpha$  alone (ER $\alpha$ +/ER $\beta$ -) or cells expression ER $\alpha$  and ER $\beta$  together (ER $\alpha$ +/ER $\beta$ +), generating by stably transfecting cells with ER $\beta$  fused to a TAP-tag at C- (CtER $\beta$ ) or N-terminus (NtER $\beta$ ) of the protein. The analysis of the RNA-Seq datasets of the three E2-stimulated cell lines revealed exon skipping (ES) as the most abundant splicing events in the post-transcriptional control induced by E2 (Dago et al. 2015). Furthermore, the comparison of ER $\alpha$ +/ER $\beta$ + to ER $\alpha$ +/ER $\beta$ - BC cells, revealed significant differences in the set of expressed isoforms between cell lines, and ER $\beta$  when expressed strongly impacted E2-induced splicing patterns, further potentiating E2-induced splicing in ER $\beta$ + BC cells by 2-fold more as compared to ER $\beta$ - cells (Dago et al. 2015). In particular, 1,264 (involving 1,016 genes), 1,402 (1,117 genes), and 975 ES events (involving 816 genes) were revealed to be induced by E2 in CtER $\beta$ , NtER $\beta$ , and wild type MCF-7 cells, respectively, of which 575 ES events were common to all cell lines analyzed, while 115 ES events showed opposite regulation in ER $\beta$ + as compared to ER $\beta$ - cell lines. Furthermore, the differences in E2-induced splicing patterns in ER $\beta$ + as compared to ER $\beta$ - cells revealed a fraction of 56 ER $\beta$ -genes both transcriptionally and post-transcriptionally, including transcriptional regulators, RNA metabolism, and genes encoding for actin cytoskeleton remodeling and protein transporters. Moreover, 35 genes whose isoforms composition significantly changed by E2 in an ER $\beta$ -dependent fashion included genes involved in apoptosis, such as *BAD*, lipid metabolism, such as *ACADM*, *PLSCR1*, *SLC27A2*, and *STARD4*, nutrient transport (*SLC25A19*, *SLC35C2*), transmembrane receptor signaling (*IFNGR2*, *LDLRAD4*), Notch signaling (*PSEN2*, *POGLUT1*, *SGKI*, *SLC35C2*), as well as some non-coding RNAs (*MCM3AP-AS1*, *SNHG17*), suggestive of a potential role of ER $\beta$  in inducing significant switches in E2-mediated splicing patterns, potentially affecting the expression or function of ER target genes. Furthermore, ER $\beta$  expression promoted significant differences in promoter usage either by attenuating those differentially used upon E2-stimulation in wild type MCF-7 cells or by inducing new switches in promoter usage. In particular, 61 ER $\beta$  promoter-switching genes were involved in important pathways controlled by E2 in BC cells, such as transcription, DNA metabolism and repair, pre-mRNA maturation and splicing, cell adhesion and polarity, and other important cellular functions. Interestingly, the integration of ChIP-Seq data for both ER $\alpha$  and ER $\beta$  together with the estrogen receptors-dependent splicing revealed a significant correlation between the binding and ER-regulated AS patterns, revealing three distinct groups of spliced genes, involved in important cellular functions such transcription, histone modifying enzymes, and apoptosis showing a binding peak for either one or for both receptors (Dago et al. 2015).

Interestingly, at the gene level, the expression of ER $\beta$  strongly impacted the E2-dependent GE profile. In particular, the regulation of (230 out of 895) genes, representing 25% of E2-regulated genes, was lost in both cell lines expressing ER $\beta$ , while a large number of new genes, not regulated in wild type MCF-7 cells, became significantly regulated in both CtER $\beta$ + (2,396 out of 2,899) and NtER $\beta$ + (2,463 out of 3043) cell lines, with the most affected genes by the expression of ER $\beta$  are the E2-inhibited genes than E2-activated (40% of E2-induced inhibition was lost versus 14% of E2-activated genes). These groups of genes whose E2-mediated



regulation was lost by expression of ER $\beta$  were related to DNA replication, recombination and repair, cell cycle and cell morphology (Dago et al. 2015).

In human primary endometrial stromal cells, the exposure to E2 increases the expression of pyruvate kinase M2 (PKM2) isoform by upregulating the c-Myc-hnRNP axis activity, which results in metabolism reprogramming and further fostering cell proliferation (Salama et al. 2014). In these cells, authors revealed that E2 also controls the PTMs of pyruvate PKM2, inducing its oxidation, phosphorylation, and nuclear translocation, where it functions as a coactivator of many transcription factors (W. Luo and Semenza 2012). The E2-induced PKM2 isoform in addition to reprogramming glucose metabolism is physically interacting with ER $\alpha$  and functions as its own coactivator. The application of small-molecule PKM2 activators inhibited the transcriptional activity of ER $\alpha$  and reduced the E2-induced proliferation of the investigated cell lines (Salama et al. 2014). Similarly, E2 was shown to impact the splicing pattern of corticotropin releasing hormone receptors (CRH-R) 1 and 2 in ER $\alpha$ -positive BCs (Lal et al. 2013). In particular, E2 induced an increase in the expression of CRH-R2 mRNA and promoted skipping of exon 12 of the type 1 receptor (CRH-R1) by inhibiting the SRSF6-SRp55 axis, which resulted in cell invasion inhibition in MCF-7 cells. Interestingly, ER $\alpha$ -positive tumors were characterized by increased abundance of CRH-R2 and CRH-R1 lacking exon 12 (CRH-R1- $\Delta$ 12) which was concordantly associated with a decreased abundance of SRp55 as compared to ER $\alpha$ -negative tumors, suggestive of a potential role of E2 in the onset of the disease by altering the splicing patterns of CRH receptors and disrupting its mediated signaling mechanisms in BC (Lal et al. 2013).

#### **1.4.3.5.2.2 Estrogen Receptor alpha acts on post-transcriptional events by controlling the expression of its key players**

ER $\alpha$  is known to control proliferation and survival of the luminal BC subtypes. The extensive analysis of genome-wide ER $\alpha$  binding sites has enabled a better understanding of the molecular mechanisms and determinants underlying this phenotype. In E2-stimulated MCF-7 and ZR-75 BC cell lines, ChIP-Seq data of ER $\alpha$  in addition to RNA-Seq data analysis revealed that ER $\alpha$  control GE process at both transcriptional and post-transcriptional levels (Cicatiello et al. 2010). The E2/ER $\alpha$  axis controls a gene regulatory network that maintains the luminal phenotype of this cell lines. This included several transcription factors such as AP-2 $\gamma$ , E2F1 and 2, ELF3, GTF2IRD1, MYB, SMAD3, RAR $\alpha$ , and RXR $\alpha$ , in addition to several miRNAs such as miR-107, miR-424, miR-570, miR-618, and miR-760 which were identified as integral components of this gene regulatory network, enabling a dynamic post-transcriptional control of the concentration and activity of expressed mRNAs levels and translation efficiency of E2-responsive genes (Cicatiello et al. 2010). ER $\alpha$  was also reported to suppress the aggressiveness of BC cells by inhibiting the expression of genes preferentially expressed in basal-like and TNBC tumors through a mRNA-mediated mechanism (Sanawar et al. 2019). ER $\alpha$ , in absence of ligands, controls the expression of *FAM171A1* transcript and protein that is preferentially expressed in basal-like breast tumors, via stimulating the expression of miR590-5p, which in turns, targets *FAM171A1* (Sanawar et al. 2019).

Moreover, the expression of several RBPs encoding genes, key players of the AS process, is regulated by the E2-ER $\alpha$  axis. Notably, the expression of the RBP NOVA1 was shown to be regulated by E2 stimulation which in turn triggers the specific expression of ER $\beta$  isoforms by regulating the AS pattern of the gene in the aging female rat brain (Shults et al. 2018, 2015). Recently, Koedoot and colleagues performed a gene co-expression analysis of RBPs encoding genes using RNA sequencing data from different primary and metastatic tumor types including BC from the TCGA and BASIS databases (Koedoot et al. 2019). Interestingly, authors identified two main clusters 1 and 2 formed by 61 and 24 splicing factors, respectively, named Enhancers-SFs and Suppressors-SFs which exhibited different expression behaviour in BC samples depending on the clinical features including PAM50, AIMS subtype, pleomorphism score, tumor grading, and hormone receptor status (Koedoot et al. 2019). In particular, cluster 1 splicing factors which were overexpressed in less aggressive tumors were overexpressed in ER $^{+}$  tumors while the splicing factors within cluster 2 which correlated with more aggressive tumors and with poor relapse- and metastasis-free survival were higher expressed in ER $^{-}$  tumors. More importantly, further stratification of ER $^{+}$  tumors into luminal A and B subtypes showed that cluster 2 splicing factors were higher expressed in the more aggressive luminal B subtype as compared to the less aggressive luminal A subtype, as they were overexpressed in basal-like and HE2-amplified compared to normal-like subtype. However, gene pairwise correlation analysis did not reveal a significant correlation between ESR1 and either of the two clusters (Koedoot et al. 2019). Another example of ER $\alpha^{+}$  signaling implication in the post transcriptional control of GE process is shown by the work of Gökmen-Polar and colleagues. In this study, authors demonstrated that *ESRP1* expression correlates with worse ER $\alpha^{+}$  BC patient overall survival and endocrine treatment outcome, showing also the gene overexpression in endocrine-resistance BC cell models. Interestingly, no significant results were obtained for ER $\alpha^{-}$  tumors supporting a relation between *ESRP1* expression and ER $\alpha^{+}$  signaling (Gökmen-Polar et al. 2019).

Furthermore, ER $\alpha$  ligand-independent transcriptional activity was also reported to be important for the expression of luminal-specific non-coding transcripts including miRNAs (Cicatiello et al. 2010) and lncRNAs (Miano et al. 2016; Caizzi et al. 2014). Miano and colleagues successfully identified a list of 133 unliganded-ER $\alpha$ -controlled lncRNAs. The extensive characterization of the identified lncRNAs using public data from BC cell lines or tumor tissues showed these lncRNAs to be only marginally overlapping E2-induced transcripts, and their expression was exclusive to luminal BC subtypes, and was able to perfectly classify BC subtypes (Miano et al. 2016) consistent with results obtained by (Niknafs, Han, Ma, Speers, et al. 2016). Specifically, DSCAM antisense RNA 1 (*DSCAM-AS1*) lncRNA showed as the most expressed lncRNA in luminal cells, whose expression is regulated by unliganded ER $\alpha$  and inversely correlated with EMT markers. When depleted in luminal BC cells, *DSCAM-AS1* was able to recapitulate some of the effects of silencing ER $\alpha$  such as inducing a growth arrest and inducing expression of EMT markers (Miano et al. 2016).

*DSCAM-AS1* is a lncRNA transcribed in antisense from intron 3 of the *DSCAM* gene on chromosome 21. The lncRNA has four different transcripts of length less than 2 Kb annotated in Gencode, resulting from AS of the gene and differing mainly by the presence or absence of a central exon. The analysis of the expression of

DSCAM-AS1 isoforms in MCF-7 BC cells using isoform-specific primers revealed an enrichment of the isoform containing the central exon in the nucleus, while the other three isoforms were mainly cytoplasmic (Miano et al. 2016).

*DSCAM-AS1* is a lncRNA that exhibits a highly cancer-specific expression pattern, mostly in BC (Niknafs, Han, Ma, Speers, et al. 2016; Miano et al. 2016, 2018), lung carcinoma (Liao and Xie 2019), and prostate (Y. Zhang et al. 2020), where it functions as an oncogenic lncRNA. In vitro experiments in different cellular systems have elucidated the possible molecular mechanisms underlying the specific expression of *DSCAM-AS1*. For example, The association between ER $\alpha$  and *DSCAM-AS1* expression was supported by ChIP-Seq and ChIP-qPCR from different luminal BC cell lines showing the binding of unliganded ER $\alpha$  at the promoter of the gene, which is further increased upon E2 stimulation (Miano et al. 2016; Niknafs, Han, Ma, Speers, et al. 2016). Furthermore, despite the increased binding of ER $\alpha$  upon E2-treatment at the promoter of *DSCAM-AS1*, its expression is largely unaffected. The analysis of chromatin states such as the super-enhancer specific histone marks, H3K27ac, identified a set of super-enhancers occupied by unliganded-ER $\alpha$ , among which a super-enhancer is mapped in proximity of *DSCAM-AS1* gene body and is occupied by unliganded-ER $\alpha$  in addition to other TFs including p300, GATA3, FoxM1 and CTCF (Miano et al. 2018). Similarly, Zhang and colleagues recently confirmed the previously identified super-enhancers in proximity to *DSCAM-AS1* gene and additionally identified that the expression of the lncRNA is regulated by FoxA1 in BC and lung cancer cell lines (Y. Zhang et al. 2020). Taken together, these data supported the lineage-specific expression of *DSCAM-AS1* in different cancer types.

Furthermore, DSCAM-AS1 is involved in the control of gene expression at the post transcriptional level through the binding of RBPs which are the top regulators of AS. The first interaction between DSCAM-AS1 and RBPs was reported by Niknafs and colleagues, demonstrating that DSCAM-AS1 is physically interacting with heterogeneous nuclear ribonucleoprotein L (hnRNPL) (Niknafs, Han, Ma, Zhang, et al. 2016). hnRNPL is a well-known splicing factor belonging to the heterogeneous nuclear ribonucleoprotein family and is involved in the regulation of AS by binding to C/A-rich motifs particularly at introns and the 3' Untranslated region (3'UTR) of the gene (Rossbach et al. 2014; Hung et al. 2007; Fei et al. 2017). Moreover, hnRNPL activity in the AS process was related mainly to the maintenance of mRNA stability by the regulation of nonsense mediated decay (NMD) pathway through the binding at gene 3'UTR region (Kishor, Ge, and Robert Hogg 2019). Such activity was particularly relevant in cancer contexts where the mRNA stability of genes with oncogenic functions such as Bcl2 (Kishor, Ge, and Robert Hogg 2019), SRSF3 (Jia et al. 2016), or tumor suppressor potential such as p53 (Gaudreau et al. 2016) was demonstrated to be regulated by hnRNPL. Although evidence existed for DSCAM-AS1 and hnRNPL physical interaction, its functional role particularly in cancer was not elucidated.

## 1.5 Aims of the Thesis:

ER $\alpha$  activity is essential to sustain epithelial cell growth and differentiation in mammary cells and to maintain the epithelial phenotype in breast tumors, and prompts the cells to respond to estrogenic hormones.

Recent studies from our laboratory and other research groups have shown that ER $\alpha$  controls transcription of a number of protein coding and noncoding genes, both when activated and in the unliganded form. Among these, a remarkable group was that of RBPs, suggesting that ER $\alpha$  controls post-transcriptional events as well. In the present work, we have taken into consideration as paradigms, (i) the lncRNA DSCAM-AS1; which is highly expressed in luminal BC in an ER $\alpha$  dependent manner, and was shown to interact with the important regulator of alternative splicing, hnRNPL; (ii) the epithelial-specific, ER $\alpha$ -regulated splicing factors ESRP1 and ESRP2, in order to evaluate a minimal network of post transcriptional control by ER $\alpha$ .

RNA-sequencing datasets were obtained through the downregulation of ER $\alpha$ , ESRP1/2, DSCAM-AS1, and HNRNPL were analysed for resulting downstream effects at both gene and isoform (e.g. splicing) levels to evidence alternative splicing and 3'UTR choice. The four datasets were then compared in order to evidence common post transcriptional events. In particular, the following questions were investigated:

- 1) How ER $\alpha$  controls AS in luminal BC? What are the mechanisms underlying this regulation?
- 2) What SFs or lncRNAs are controlled by ER $\alpha$  and how in turn control the process of AS in luminal BC?
- 3) What is the functional role of DSCAM-AS1, as one of the top luminal-specific lncRNA transcripts whose expression is strongly dependent on ER $\alpha$ ? and how it impacts post-transcriptional events, taking into consideration its physical interaction with hnRNPL in luminal BC?
- 4) Does the identified post transcriptional events have a prognostic value in the context of BC.

---

## chapter 2. Results

### Chapter organization

This results section is organized into four different chapters:

The first chapter reports the results obtained from the analysis of the hormone-independent activity of ER $\alpha$  at both transcriptional and post transcriptional levels in the ER $\alpha$ + MCF-7 luminal BC cell lines.

The second chapter reports the results related to the functional role of the core alternative splicing (AS) regulators, ESRP1 and ESRP2 proteins, in the epithelial cells. In this chapter, the role of unliganded ER $\alpha$  activity in regulating the expression of these two genes is evidenced and the effects of the combined silencing of *ESRP1* and *ESRP2* in MCF-7 cells at both transcriptional and post transcriptional levels are reported and are compared with those changes observed upon silencing ER $\alpha$ . The results are discussed in the context of gene expression as well as AS regulation.

In the third chapter are described the results obtained from the analysis of the effect of silencing the ER $\alpha$ -regulated lncRNA, *DSCAM-ASI*. In this chapter, the expression of *DSCAM-ASI* is analysed in 31 public microarray datasets in addition to two independent in-house cohorts of patients. The association of *DSCAM-ASI* gene expression with distinct patient's clinical and molecular features are discussed. The second axis of this chapter describes the effect of silencing *DSCAM-ASI* on gene expression as well as on cell proliferation and phenotypic changes in MCF-7 and other BC cell lines. The last axis of this chapter reports the effects of silencing *DSCAM-ASI* on AS and isoform expression in MCF-7 cells. In this axis, the physical interaction of *DSCAM-AS1* with the splicing factor hnRNPL showed in MCF-7 cells is evidenced and suggests that the disruption of this interaction, by silencing *DSCAM-ASI*, induces significant changes in RNA transcripts by a 3'UTR shortening and exon skipping mechanisms.

The fourth chapter reports the effects of silencing the splicing factor hnRNPL at both gene and isoform levels through the analysis of a paired-end RNA-seq dataset consisting of siRNA-mediated silencing of *HNRNPL* in MCF-7 BC cells. Therefore, a differential gene/isoform expression as well as AS analyses are performed to decipher the core biological pathways affected by hnRNPL depletion. Then, this chapter reports transcriptional and post transcriptional changes occurring in both *HNRNPL* and *DSCAM-ASI* silencing experiments. The final axis of this chapter discusses the prognostic value of hnRNPL in ER $\alpha$ + and ER $\alpha$ - BC patients. The association between hnRNPL gene expression and patient's clinical features are discussed.

Finally, at the end of each chapter, the main findings are presented and discussed in the context of gene expression and AS levels.

## 2.1 The functional investigation of the hormone-independent Estrogen Receptor alpha signaling pathway on transcriptional and post-transcriptional levels

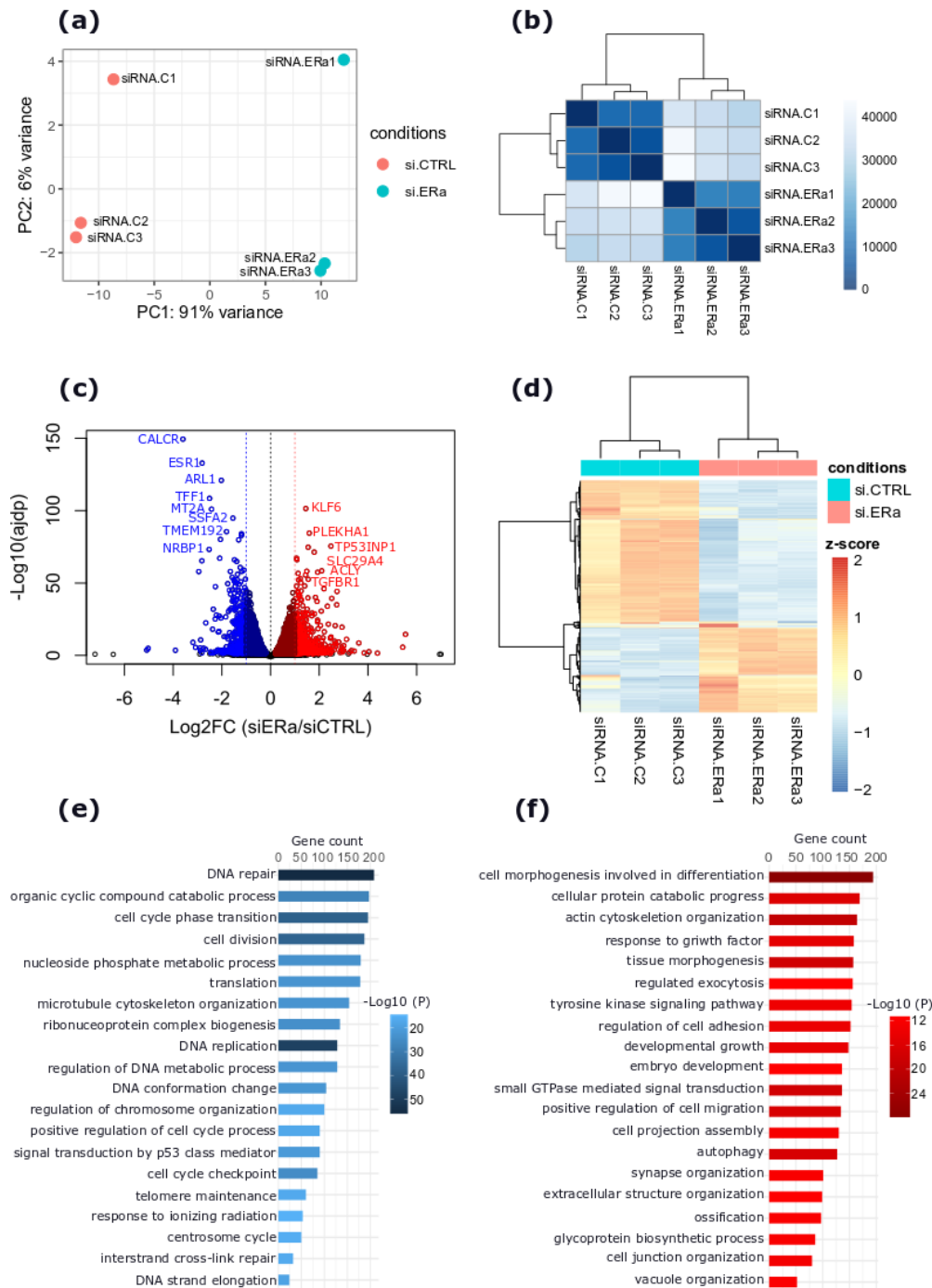
Our groups have previously reported the functional significance of the hormone independent activity of ER $\alpha$  in different BC cell lines and primary BC tissues by performing an integrative analysis of high throughput data such as CHIP-Seq and epigenomics data (Caizzi et al. 2014; Miano et al. 2016, 2018). In this chapter, to characterise the functional significance of the hormone-independent activity of ER $\alpha$  on both transcriptional and post-transcriptional levels in BC cells, we developed an integrative computational pipeline that performs a differential expression analysis at both gene and isoform levels, allowing the characterization of AS changes at a genome-wide scale and its downstream consequences at the protein isoform level. Therefore, the results of this chapter will be organized into three main different sections:

In **section 2.1.1** the results obtained from the hormone-independent activity of ER $\alpha$  on transcription are reported. To identify genes whose expression is regulated by the hormone-independent ER $\alpha$  activity, an RNA-seq experiment consisting of silencing the ER $\alpha$  gene in MCF-7 cells was performed (Miano et al. 2018). Next, a differential expression analysis was performed at both gene and isoform levels. genes whose expression responded to ER $\alpha$  depletion were screened to identify RNA-binding protein (RBP) genes that are potentially involved in post-transcriptional events. Similarly, transcription factors (TFs) genes which are potentially involved in the epithelial-to-mesenchymal transition (EMT) process were also identified.

In **section 2.1.2** the results obtained from hormone-independent activity of ER $\alpha$  on the post transcriptional level are reported. The effects of unliganded-ER $\alpha$  activity depletion on isoform usage as well as on AS process are presented. To this end, a differential isoform usage (dIU) analysis was performed and genes with switching isoforms upon ER $\alpha$  silencing were identified. In addition, to identify alternatively spliced exons, a differential AS analysis considering local alternative splicing events (ASEs) was performed using rMATS (Shen et al. 2014, 2012). Significant ASEs were identified and annotated using IsoformSwitchAnalyzeR (Vitting-Seerup and Sandelin 2019) to identify their potential effects at protein isoform level. Next, to identify putative RBPs regulators of the significant ASEs upon ER $\alpha$  silencing, an RBP binding motif enrichment analysis was performed for event, region, and direction of regulation of the ASE using MoSEA (Sebestyén et al. 2016) and fimo (Bailey et al. 2015). Furthermore, the expression of significant ASEs in primary BC tissues was analysed and correlated with *ESR1* gene expression in these samples. To identify the potential functional role(s) of the ER $\alpha$ -regulated ASEs, a molecular pathway-guided enrichment analysis was performed reporting the correlation between ASE expression level (Percent Splice-in Index, PSI) in primary BC samples, and enriched hallmarks using PEGASAS algorithm (Phillips et al. 2020).

### 2.1.1 The hormone independent activity of ER $\alpha$ is crucial for cell proliferation and for maintaining the epithelial phenotype

To explore the functional role of the unliganded ER $\alpha$  activity on gene expression process in BC, analysis of gene and isoform differential expression (dGE, dIE, respectively) was carried out on a triplicate paired-end RNA-seq experiment consisting of hormone-starved MCF-7 BC cells treated with control or with *ESR1*-targeting siRNA (siCTRL vs siER $\alpha$ ). The quality of the RNA-Seq replicates are reported in the PCA and heat map plots in (**Figure 2.1-1a,b**), respectively. As a result, silencing ER $\alpha$  in MCF-7 cells in absence of hormone defined a set of differentially expressed (DE) genes, including protein coding and non-coding genes. Notably, a total of 6611 genes were DE ( $|\log_2FC| > 0.2$  and  $adjp < 0.05$ ), including 3741 downregulated and 3140 upregulated genes. The  $\log_2FC$  and statistical significance of these DE genes are reported in (**Figure 2.1-1c** and **Supplementary Table 1a**), and the expression of the top 500 changing genes is reported in (**Figure 2.1-1d**). The functional enrichment analysis of the DE genes revealed that downregulated genes are mainly enriched in terms related to cell cycle progression and cell proliferation pathways, DNA replication, DNA damage repair, positive regulation of cell cycle process, in line with the previously known role of ER $\alpha$  as a key transcription factor in the luminal BC cells MCF-7 (Caizzi et al. 2014). Conversely, upregulated genes were mainly enriched in terms related to EMT process, actin cytoskeleton organization, cell movement, cell morphogenesis involved in differentiation, developmental growth and positive regulation of cell migration (**Figure 2.1-1e,f** and **Supplementary Table 1b,c**), supporting that ER $\alpha$  is a key TF important for the maintenance of the epithelial phenotype and for the expression of epithelial genes in the investigated cell lines.

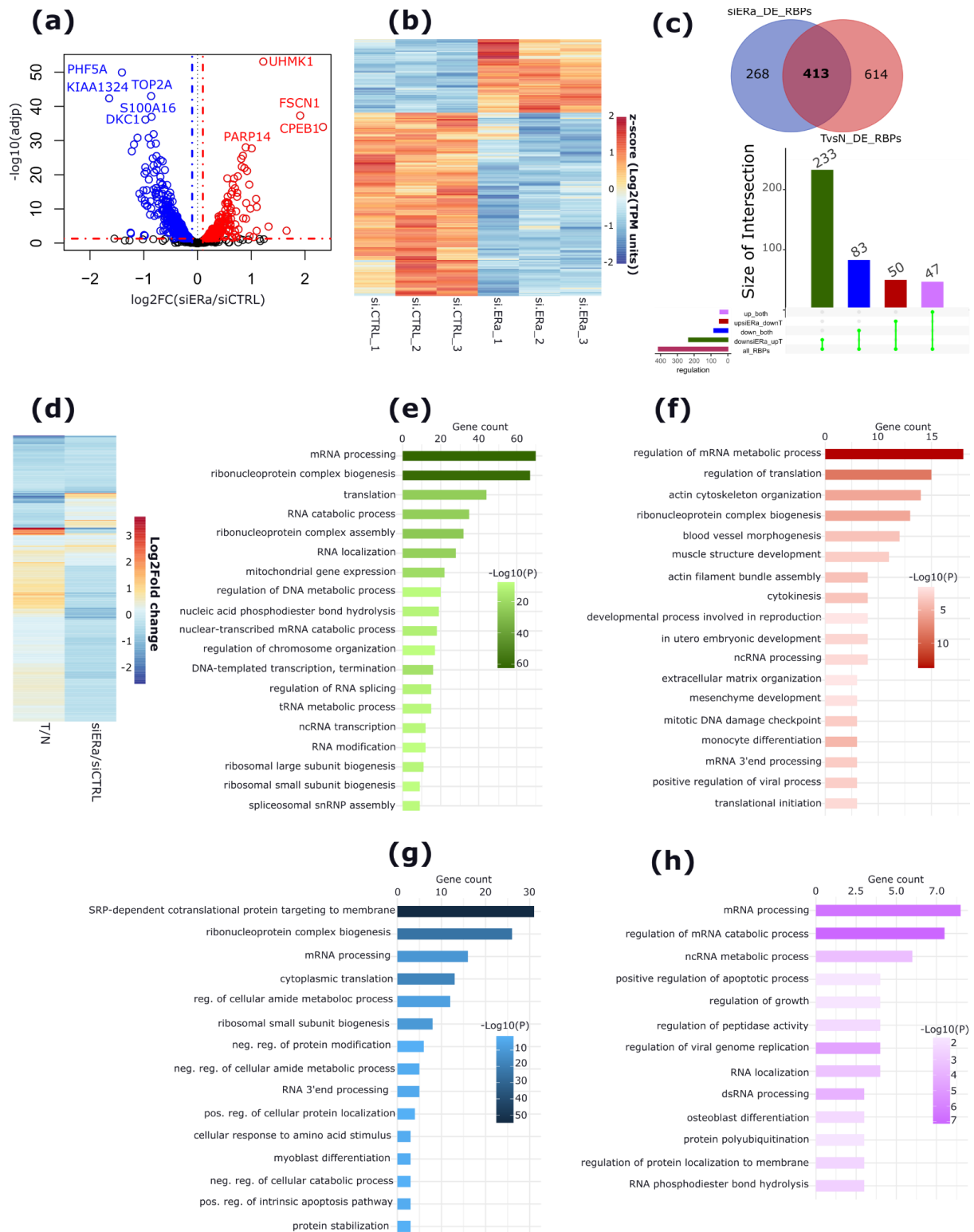


**Figure 2.1-1:** The transcriptional effects of silencing *ERa* gene in MCF-7 BC cells cultured in absence of hormone. (a) PCA plot reporting the two different clusters formed by the replicates of each condition. (b) Dissimilarity matrix between replicates of *siCTRL* and *si.ERa* conditions. (c) Volcano plot showing the  $\log_2\text{FC}$  and significance (*adj-p*) of genes responding to *ERa* gene silencing. In blue color are represented downregulated genes while in red are represented upregulated genes. (d) Heat map the top 500 changing genes upon *ERa* gene silencing. Color intensities correspond to z-score calculated as a difference between mean and variance over samples. Negative and positive z-scores correspond to downregulated and upregulated genes, respectively. (e-f) Bar plots showing the gene ontology (GO) enrichment analysis related to downregulated and upregulated genes, respectively. Bar size represents the number of genes overlapping each enriched GO term and color intensities are proportional to the significance (*p*-value) of the enriched GO terms.



### 2.1.1.1 The Estrogen Receptor alpha signaling pathway controls the expression of RNA binding proteins (RBPs) including splicing factors (SFs) in luminal BCs

Interestingly, the depletion of ER $\alpha$  in MCF-7 induced significant changes in the expression level of RBP genes. Notably, significant expression changes were observed in a total of 681 RBP genes upon ER $\alpha$  silencing, most of which (486, 71%) were downregulated, while only (195, 29%) were upregulated (**Figure 2.1-2a,b** and **Supplementary Table 1d**). Importantly, the comparison of these DE RBP genes to a previously published list of RBP genes reported to be DE between paired normal and ER $\alpha$ + BC tissues (Sebestyén et al. 2016), revealed 413 RBP genes common to both datasets (exact hypergeometric probability,  $p < 0$ ) (**Figure 2.1-2c** and **Supplementary Table 1e**). The overlapping DE RBP genes included 130 (31%) RBP genes coherently regulated in both datasets (83 downregulated and 47 upregulated in both datasets), and 283 (69%) RBP genes showed an opposite gene expression change (233 downregulated in ER $\alpha$  silencing while upregulated in tumors, and 50 upregulated in ER $\alpha$  silencing while downregulated in tumors) (**Figure 2.1-2c,d** and **Supplementary Table 1e**). Moreover, the gene enrichment analysis of the different groups of RBPs revealed distinct GO terms enriched for each group (**Figure 2.1-2e,h** and **Supplementary Table 1f**). Notably, downregulated RBP genes in our dataset and upregulated in tumors were enriched in terms related to pre-mRNA splicing, mRNA modification and processing, splicing and localization, ncRNA splicing regulation, and spliceosomal snRNP complex assembly, while those RBP genes upregulated in our dataset and downregulated in tumors were enriched in terms related mainly to mRNA metabolic process, regulation of translation, actin cytoskeleton organization, actin filament bundles assembly, differentiation, extracellular matrix organization, and mesenchyme development. The third group including downregulated RBP genes in both datasets was enriched in terms mainly related to SRP-dependent cotranslational protein targeting to membrane, ribonucleoprotein complex biogenesis, cytoplasmic translation, regulation of protein modification, localization and stabilization process, positive regulation of intrinsic apoptotic signaling pathway and RNA 3' end processing.

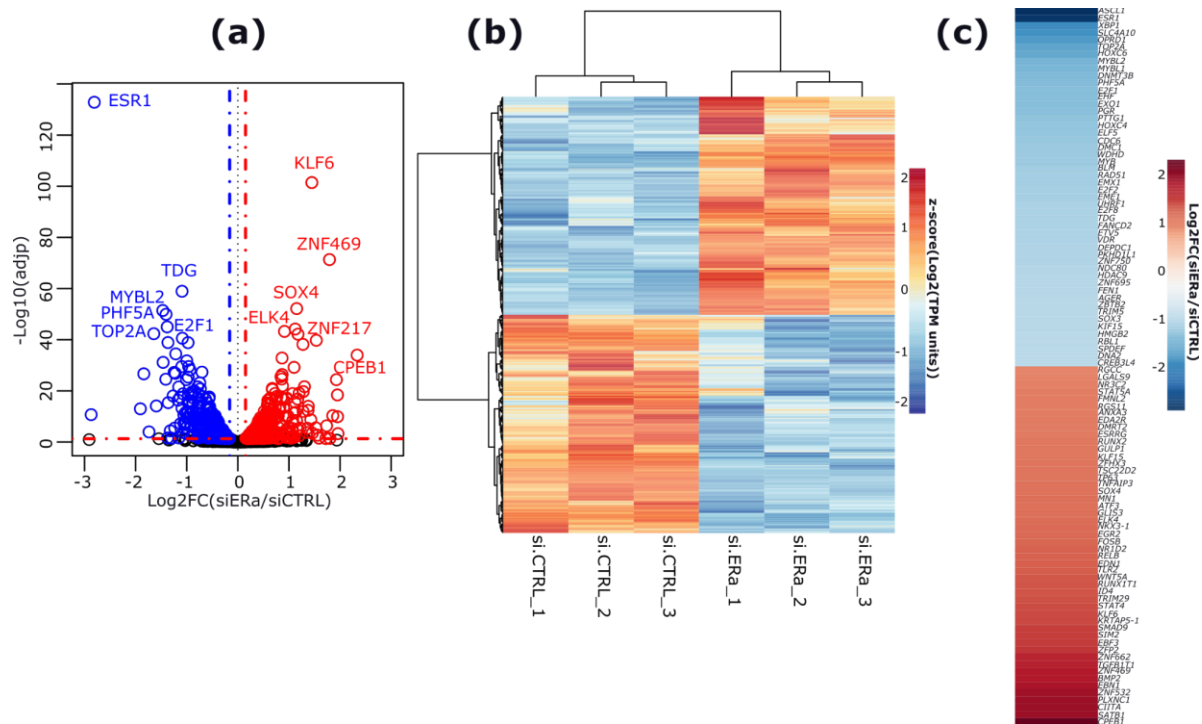


**Figure 2.1-2:** Effects of unliganded-ER $\alpha$  activity depletion on the expression of RBP genes in MCF-7 cells. (a) Volcano plot reporting the  $\log_2FC$  and the significance ( $adj-p$ ) of DE RBP genes. (b) Heat map of the 681 DE RBP genes. Color intensities are proportional to the z-score calculated among samples using RBPs TPM units. Positive and negative z-scores correspond to upregulated and downregulated RBP genes, respectively. (c) top: Venn diagram reporting the number of RBP genes DE upon ER $\alpha$  silencing in MCF-7 and between breast tumor (T) and normal (N) samples, bottom: Upset plot reporting the numbers and type of regulation of overlapping RBP genes. Histograms represent the number of RBPs per each group. (d) Heat map plot reporting the  $\log_2FC$  of overlapping RBP genes. T/N, Tumor/Normal

comparison. (e-h) Bar plots reporting enriched GO terms for different RBP groups classified based on their direction of regulation upon ER $\alpha$  silencing in MCF-7 cells and when comparing tumors versus normal breast tissues, including down\_ER $\alpha$ \_UP\_TvsN, up\_ER $\alpha$ \_down\_TvsN, down\_both, up\_both groups, respectively.

Furthermore, to further support our hypothesis that unliganded-ER $\alpha$  controls EMT in BC by exerting a post transcriptional control of gene expression process, the list of RBP genes identified as DE upon ER $\alpha$  silencing in MCF-7 cells were overlapped with a well-established dataset by (Shapiro et al. 2011) comparing the expression levels of RBP genes between epithelial and mesenchymal BC cell lines. Interestingly, we found that ER $\alpha$  silencing induces an enrichment of RBP genes that are highly expressed in mesenchymal cells and a depletion of RBP genes that are highly expressed in epithelial cells as compared to mesenchymal cells (**Supplementary Table 1g,h**). Moreover, among the DE RBP genes, 84 splicing factor (SFs) genes which represent the key players of pre-mRNA alternative splicing process, as confirmed by (Hegele et al. 2012) (244 in total, **Supplementary Table 2a**), were also DE in our dataset, including 63 downregulated SFs, such as epithelial splicing regulatory protein 1 (*ESRP1*) and 2 (*ESRP2*) genes, core AS regulators in epithelial cells and which are downregulated during EMT process as reported by (Shapiro et al. 2011), and 21 upregulated SFs such as the QKI, KH domain containing RNA binding (*QKI*) gene, and the Splicing factor 3b subunit 1 (*SF3B1*) gene (**Supplementary Table 1i**).

Moreover, ER $\alpha$  silencing not only induced changes in the expression of RBP genes, but also in the expression of TF genes that could potentially be involved in the control of RBP genes expression. The list of TFs with confirmed DNA-binding ability was obtained from a study by (Lambert et al. 2018) (**Supplementary Table 2b**). This analysis revealed 868 DE TF genes upon ER $\alpha$  silencing, including 434 downregulated and 434 upregulated TFs (**Figure 2.1-3a,b** and **Supplementary Table 1j**). The top 100 significant DE TF genes, including top 50 downregulated and top 50 upregulated TF genes are presented in (**Figure 2.1-3c**). In order to explain the observed changes in RBP gene expression, an explanatory network was constructed by integrating ChIP-seq datasets of DE TFs including ER $\alpha$  together with DE results. Therefore, ChIP-Seq datasets for DE TFs in MCF-7 cells under conditions such as with (+E2) or without (-E2) 17- $\beta$  estradiol, in estrogens-enriched medium (full medium, FM) or hormone-deprived condition (HD) were collected from the cistromeDB database (Mei et al. 2017; R. Zheng et al. 2019) and from the ENCODE portal (Davis et al. 2018), except for ER $\alpha$  peaks consisting of highly conservative peak sets were obtained from (Ferrero et al. 2017) and the enhancer-gene interactions in these conditions were retrieved from (Bai et al. 2019). TF binding peaks were searched at the promoter, the body, and at distal regulatory elements of each RBP isoform identified as DE in our dataset. Thus, the final network included a set of regulators represented by TFs including ER $\alpha$  and a set of targets represented by RBPs and edges are defined based on the evidence from the ChIP-Seq datasets. This analysis revealed the presence of ER $\alpha$  peaks at 157 RBP isoforms, including 30 RBP isoforms showing a peak at the promoter, and 127 RBP isoforms showing a peak at the gene body (**Supplementary Table 1k**).



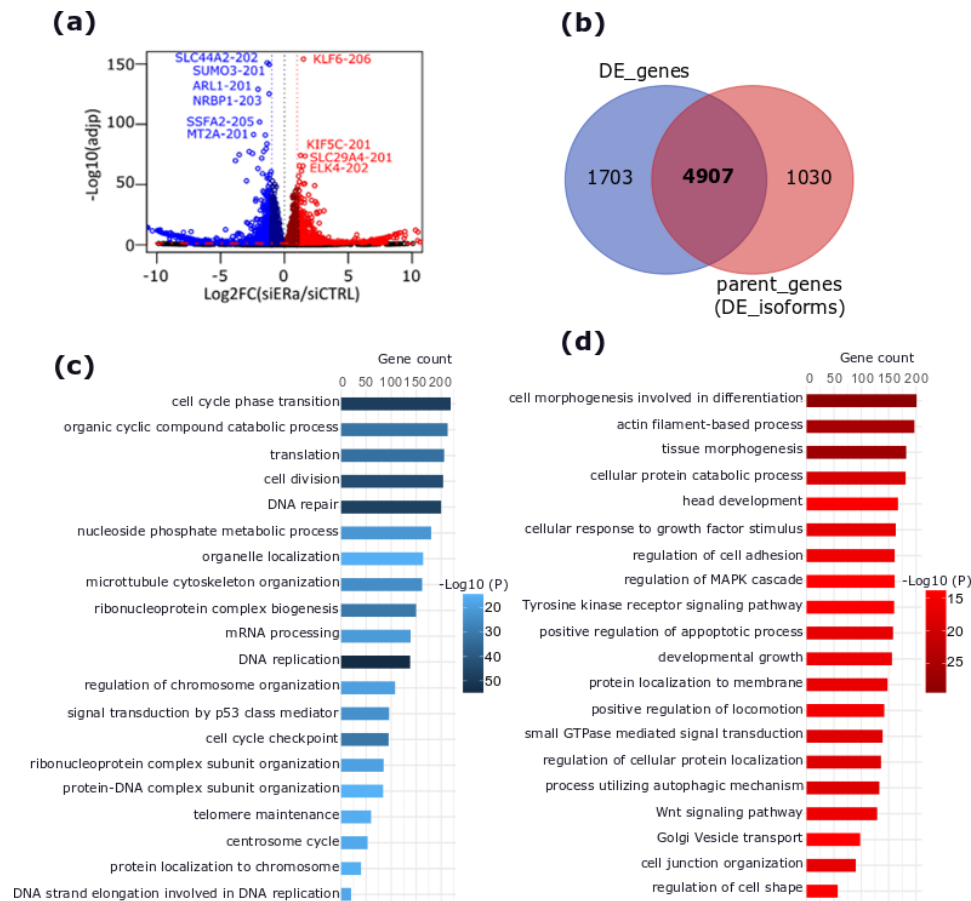
**Figure 2.1-3:** Summary of regulated TFs upon ER $\alpha$  depletion in MCF-7 cells. (a) Volcano plot reporting  $\log_2FC$  and adjusted  $p$ -value of the downregulated (blue) and upregulated (red) TF genes. (b) Heat map plot reporting the expression levels of the 868 DE TF genes. Color bar intensities are proportional to gene expression  $z$ -scores, with blue colors represent negative  $z$ -scores and red colors represent positive  $z$ -scores which correspond to downregulated and upregulated TF genes, respectively. (c) Heat map plot reporting the  $\log_2FC$  of the top 100 DE TF genes ranked based on  $\log_2FC$ . Blue and red colors represent downregulated and upregulated TF genes, respectively.

### 2.1.2 ER $\alpha$ depletion in MCF-7 BC cells induces a differential expression of gene isoforms resulting in differential functional consequences

Since ER $\alpha$  silencing in MCF-7 cells depleted the expression of hundreds of RBP genes including many SF genes, regulators of the AS process, we sought to examine its effects at the isoform level by performing three independent types of analyses: (i) an isoform differential expression (dIE) analysis to identify DE isoforms; (ii) a differential isoform usage (dIU) analysis, consisting of analysing the changes in the relative abundance of individual isoforms and how the contribution of each to the expression of parent gene changes upon ER $\alpha$  silencing; and (iii) by performing a differential local AS analysis in order to identify differentially spliced exons.

The dIE analysis revealed 8742 DE isoforms ( $|\log_2FC| > 0.2$  and  $adjp < 0.05$ ) in our dataset, 4769 downregulated and 3973 upregulated (**Figure 2.1-4a** and **Supplementary Table 3a**). These DE isoforms originated from 5937 genes, of which 4907 (74%) were identified as DE genes according to dGE analysis (**Figure 2.1-4b** and **Supplementary Table 3a**). Functional enrichment analysis of the parent genes of these DE isoforms revealed enriched GO terms similar to those enriched based on dGE analysis. Notably, genes with downregulated isoforms were enriched in terms mainly related to cell cycle progression and cell

proliferation, while genes with upregulated isoforms were mainly related to cell movement and EMT-related processes (**Figure 2.1-4c,d** and **Supplementary Table 3b,c**).



**Figure 2.1-4:** dIE analysis reporting the effect of silencing ER $\alpha$  at isoform level in MCF-7 cells in absence of hormone. (a) Volcano plot of DE isoforms upon ER $\alpha$  depletion in MCF-7 cells. in blue are reported downregulated while in red are reported upregulated isoforms. The top DE isoforms are labeled. (b) Venn diagram reporting the overlap between gene (dGE) and isoform (dIE) differential expression analyses. (c,d) Gene ontology analysis showing the enriched GO terms related to parent genes of downregulated and upregulated isoforms, respectively.

In addition, dIE analysis revealed genes with both upregulated and downregulated isoforms. For instance, 60 isoforms were downregulated while their parent genes were reported as upregulated according to dGE analysis. Similarly, 63 isoforms were reported as upregulated while their parent genes were considered downregulated by dGE analysis (**Supplementary Table 3d,f**). In addition, as shown in (**Figure 2.1-4b**), 1030 genes had a DE isoform while they were considered as not DE by dGE analysis. Such genes could not be identified as regulated by performing a DE analysis at gene level, due to the compensation effect driven by the differential regulation of their isoforms. Therefore, to identify genes with isoform switching events driven by ER $\alpha$  silencing, we performed a differential Isoform Usage (dIU) analysis by running the isoformSwitchAnalyzeR tool (Vitting-Seerup and Sandelin 2017a). Briefly, the tool defines isoform usage as an isoform fraction (IF) by calculating the ratio ( $\exp(\text{isoform})/\exp(\text{parent gene})$ ). Next, the IF ratio of each individual isoform is compared between siCTRL and siER $\alpha$  conditions, resulting in a dIF obtained by subtracting ( $\text{IF}(\text{siER}\alpha) - \text{IF}(\text{siCTRL})$ ) as illustrated in (**Figure 2.1-5a**, details are given in MM section). Interestingly, this dIU analysis revealed 605

genes with isoforms differently responding (e.g. one is induced, the other is repressed) to ER $\alpha$  silencing in MCF-7 BC cells. Specifically, 758 isoforms showed significant switching events ( $\text{adj}p < 0.05$  and  $|\text{dIF}| > 0.05$ ) (**Figure 2.1-5b** and **Supplementary Table 4a**). In particular, the expression of 379 isoforms was repressed while 380 isoforms were induced in the siER $\alpha$  condition (**Figure 2.1-5c** and **Supplementary Table 4a**). Functional enrichment of genes harbouring significant switching isoform pairs revealed GO terms enriched in covalent chromatin organization, regulation of chromosome segregation, response to ionizing radiation, actin filament organization, cell-cell junction organization, stress-activated MAPK cascade, DNA replication, mitotic DNA damage checkpoints, epigenetic regulation of gene expression process, and lactation (**Supplementary Table 4b**). Moreover, we took advantage of the IsoformSwitchAnalyzeR tool (Vitting-Seerup and Sandelin 2017a) to annotate switching isoforms. The tool compares the structural features of upregulated and downregulated isoforms and reports differences such as intron retention (IR) events, exon inclusion/exclusion, alternative transcription start (ATSS) and termination sites (ATTS) usage, presence or absence of poison exons involved in the NMD process, in addition to 3'UTR and 5'UTR regions length. The enrichment of each of these features is based on the expression changes of the isoform populations resulting from opposing features. This analysis showed that ER $\alpha$  silencing results in the enrichment of isoforms with longer 3'UTR, longer 5'UTR, have more protein domains, more IR events, insensitive to NMD process, as well as a switch from non-coding to coding transcripts (**Figure 2.2-5d** and **Supplementary Table 4c**). In particular, in 370 genes, the isoform switching pairs resulted in an enrichment of isoforms with longer 3'UTRs (187 longer vs 57 shorter, proportion  $q$ -value =  $8.85\text{E-}16$ ), longer 5'UTRs (120 longer vs 74 shorter, proportion  $q$ -value =  $2.46\text{E-}3$ ), more protein domains gain (152 gained vs 94 lost domains, proportion  $q$ -value =  $8.36\text{E-}4$ ), IR events gain (57 IR gained vs 30 IR loses, proportion  $q$ -value =  $6.59\text{E-}3$ ), NMD insensitivity (31 NMD insensitive vs 12 NMD sensitive, proportion  $q$ -value =  $6.59\text{E-}3$ ), in addition to enrichment in transcripts with coding potential (101 are coding vs 65 are non-coding, proportion  $q$ -value =  $6.59\text{E-}3$ ) (**Supplementary Table 4c**). Importantly, we sought that genes harbouring isoform switching events with downstream consequences could be of interest. Notably, gene functional enrichment analysis of these genes revealed terms enriched in DNA damage checkpoint, G2 DNA damage checkpoint, regulation of chromosome organization, autophagy, regulation of protein autophosphorylation, actin filament-based process, and constitutive secretory pathway (**Supplementary Table 4d**). Moreover, a protein-protein interaction (PPI) network analysis of these genes using MCODE utility of Metascape (Y. Zhou et al. 2019) revealed a significant enrichment of functionally related protein complexes involved in DNA replication, post-translational protein modification, translational termination, translational elongation, and regulation of intracellular steroid hormone receptor signaling pathway, in addition to RAS protein signal transduction and small GTPase mediated signal transduction pathway (**Supplementary Figure 1** and **Supplementary Table 4d**). A gene functional enrichment analysis was performed on each group of consequences as provided in (**Supplementary Table 5e-k**). Furthermore, the analysis of putative AS events underlying the observed isoform switching events using the spliceR utility of IsoformSwitchAnalyzeR revealed a number of AS events (ASEs) to be enriched. Specifically, this analysis showed that in the ER $\alpha$  silencing condition are enriched IR events ( $n_{\text{UP}}=58$ ,  $n_{\text{DOWN}}=30$ , proportion  $q$ -

value=1.06E-2), as well as a differential usage of TSS (usage of distal promoter) (nUP= 137, nDOWN=191, proportion q-value=1.06E-2) and TTS (usage of distal 3'UTR exon) (nUP=109, nDOWN=232, proportion q-value=3.14E-10) and to lesser extent ES events (nUP=184, nDOWN=147, proportion q-value=4.78E-2) and A5'SS usage (nUP=111, nDOWN=79, proportion q-value=4.49E-2) (**Figure 2.1-5e** and **Supplementary Table 4I**). A clear example gene with switching isoform pairs is presented in (**Figure 2.1-5f**). In this example, the isoform switching pairs involve two *USO1* isoforms that differ by the inclusion of two exons, one of which is encoding for a protein domain. Thus, in this example, ER $\alpha$  silencing results in the upregulation of the exclusion form of the event (*ENST00000514213.6*), while the inclusion form of the event (*ENST00000264904.8*) is downregulated (e.g. two exon skipping (ES) events gained). Consequently, the two protein isoforms encoded by the switching isoform pair differ by the presence/absence of a protein domain (*PF04869*) involved in the dimerization of the globular p115 head of USO1 protein (**Figure 2.1-5g**). The dimer formation of this globular domain mediates its interaction with Rab1 as well as its recruitment on COP II coated vesicles (An et al. 2009; Heo et al. 2020).

Noteworthy, most of the genes (512, 85%) reported to have an isoform switching event were also considered as regulated by dIE analysis. However, dIU analysis revealed more cases where isoform-specific response and the overall gene response to ER $\alpha$  silencing were different. For instance, the expression of 120 isoforms was shown to be repressed by ER $\alpha$  silencing, while their parent genes were considered as upregulated by dGE analysis (**Supplementary Figure 2A**). Similarly, 160 isoforms were reported to be induced by ER $\alpha$  silencing, while their parent genes were considered as downregulated by dGE. On the other hand, other genes were considered as not regulated by dGE analysis, but were significantly regulated at the isoform level only as shown in the selected examples reported in (**Supplementary Figure 1B**). These results suggest that ER $\alpha$  not only controls GE transcriptionally, but also controls the expression of specific RNA isoforms through the control of their co-transcriptional and post-transcriptional processing events.

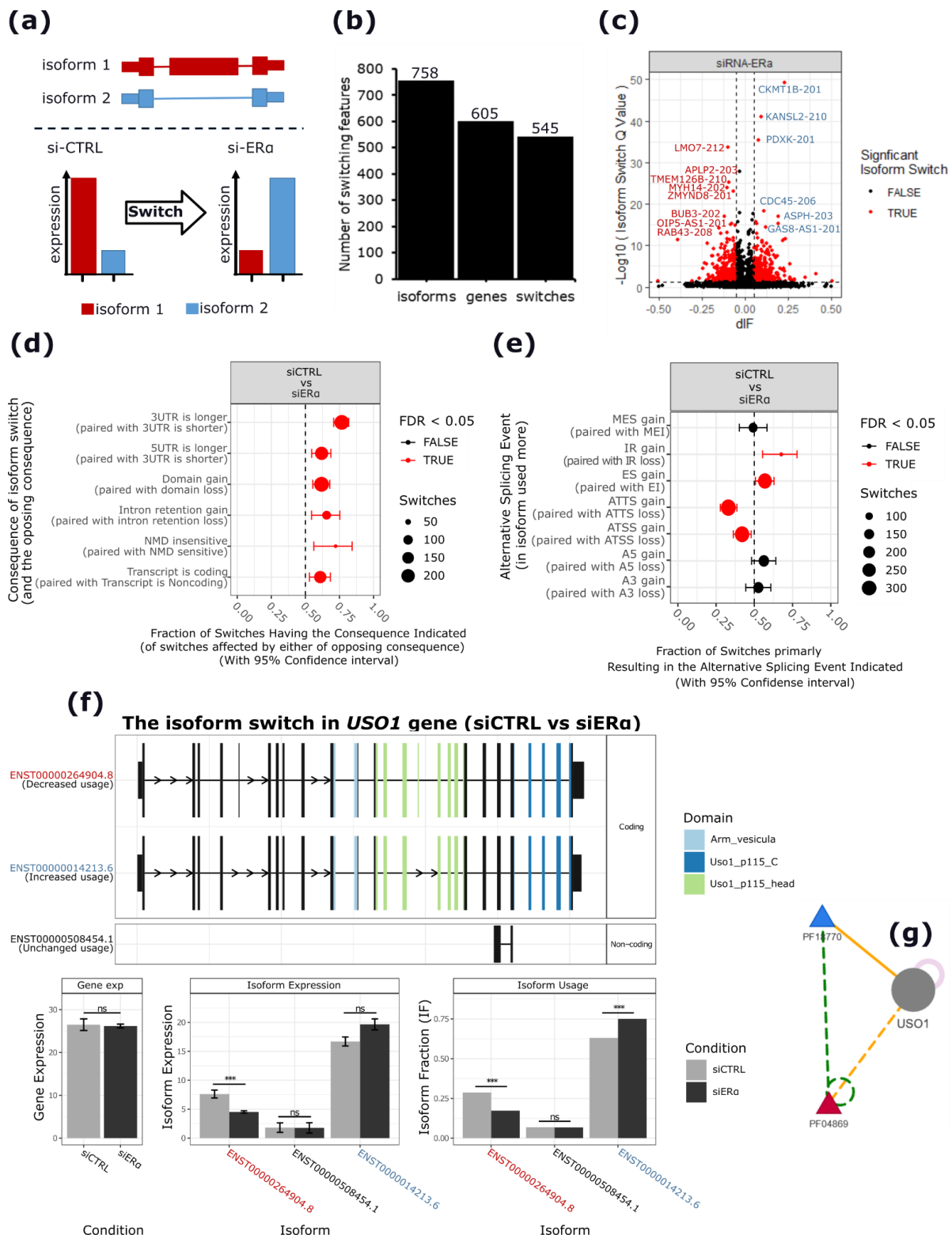


Figure 2.1-5: Isoform switching events observed upon ERa gene silencing in MCF-7 in absence of hormone. (a) Scheme depicting the principle behind differential isoform usage (dIU) analysis implemented in IsoformSwitchAnalyzeR tool (Vitting-Seerup and Sandelin 2019). ERa silencing induces a switch in the relative abundance of the two isoform pairs 1 and 2 as compared to control condition. (b) Bar plots reporting the number of significantly switching genes and isoforms involved in the switching events. (c) Volcano plot reporting the dIF and relative significance ( $-\log_{10}(\text{adj } p\text{-value})$ ) of switching isoform pairs. The top significant switches are labeled accordingly. (d) consequences enrichment analysis reporting the enrichment of specific isoform features resulting from switching isoform pairs. The x-axis indicates the



relative proportion of isoform switches having the consequence indicated on the y-axis. A relative proportion equal to 0.5 means no specific enrichment of that feature. A relative Proportion  $< 0.5$  means higher number of isoforms with the consequence indicated are enriched in the control condition. A relative Proportion  $> 0.5$  means higher number of isoforms with the consequence indicated are enriched in the siER $\alpha$  condition. Proportions statistically different are shown in red color. (e) ASEs enrichment analysis reporting the fraction of switches primarily resulting from each specific ASE. MES/MEI, multiple exon skipping/inclusion; ES/EI, exon skipping/inclusion; A5/A3, alternative 5'/3' splice sites; IR, intron retention; ATTS/ATSS, alternative transcription termination/start sites. (f) Isoform switching plot reporting the significant isoform switching pairs in *USO1* gene induced by ER $\alpha$  silencing. Upper panel shows the isoforms involved in the switch and their 5'UTR, 3'UTR, exons, and relative protein domains they encode for. The histograms show the gene and isoform expression levels in (normalized TPM units) and their DE status at both gene and isoform levels (ns, not significant, \*\*\*,  $p < 0.0001$ ). (g) downstream effects of the isoform switching event observed in *USO1* gene. Triangles represent the protein domains of the *USO1* protein (grey circle), with the domain encoded by the skipped exon 15 is highlighted in red and its partner protein domain is highlighted in blue. Dashed green and yellow lines represent the suppressing effects of the exon 15 skipping event at protein level.

### 2.1.3 ER $\alpha$ depletion in MCF-7 induces internal alternative splicing events

In order to further support our findings from the dIU analysis, we added an additional layer of information by performing a differential AS analysis using rMATS (Shen et al. 2014). rMATS calculates AS changes by considering the number of sequencing reads spanning on exon-exon junctions and provides a more precise quantification of local ASEs. Interestingly, rMATS analysis revealed 825 unique ASEs upon hormone-independent ER $\alpha$  activity depletion in MCF-7 BC cells. Specifically, 546 (65%) ASEs were classified as ES, followed by 145 (17%) IR, 73 (9%) A3, 45 (5%) A5', and 37 (4%) Mutually Exclusive Exons (MX) events (**Figure 2.1-6a,b** and **Supplementary Table 5a-f**). A density plot reporting the differential inclusion levels (dPSI, for differential percent spliced-in index) for the significant ASEs identified is reported in (**Figure 2.1-6c** and **Supplementary Table 5a-f**). This plot shows that the dPSI of most of the significant ASEs falls within the range (-0.2 to 0.2) with the exception in case of IR events where the dPSI of most of the events falls within the range (-0.1 to 0.1). Interestingly, the top100 significant ASEs are ES events of which the top50 are shown in (**Figure 2.1-6c**), most of which are induced upon ER $\alpha$  depletion. Moreover, the functional enrichment analysis of genes harbouring ASEs showed significant enrichment in terms related to different processes, depending on the ASEs type. For instance, genes harbouring an ES event are related to mitotic cell cycle phase transition, chromosome segregation process, phospholipid metabolic process, actin cytoskeleton organization, and actin filament-based movements. Conversely, genes harbouring an IR event were related to terms enriched in DNA mismatch repair, double-strand repair, spindle organization, and transcription by polymerase I (**Figure 2.1-6e** and **Supplementary Table 5g-k**). Selected examples of genes with the most significant ES events including *APLP2*, *LMO7*, *MYH14*, *SCUBE2*, *USO1*, and *APBB2* are reported in (**Figure 2.1-6f**).

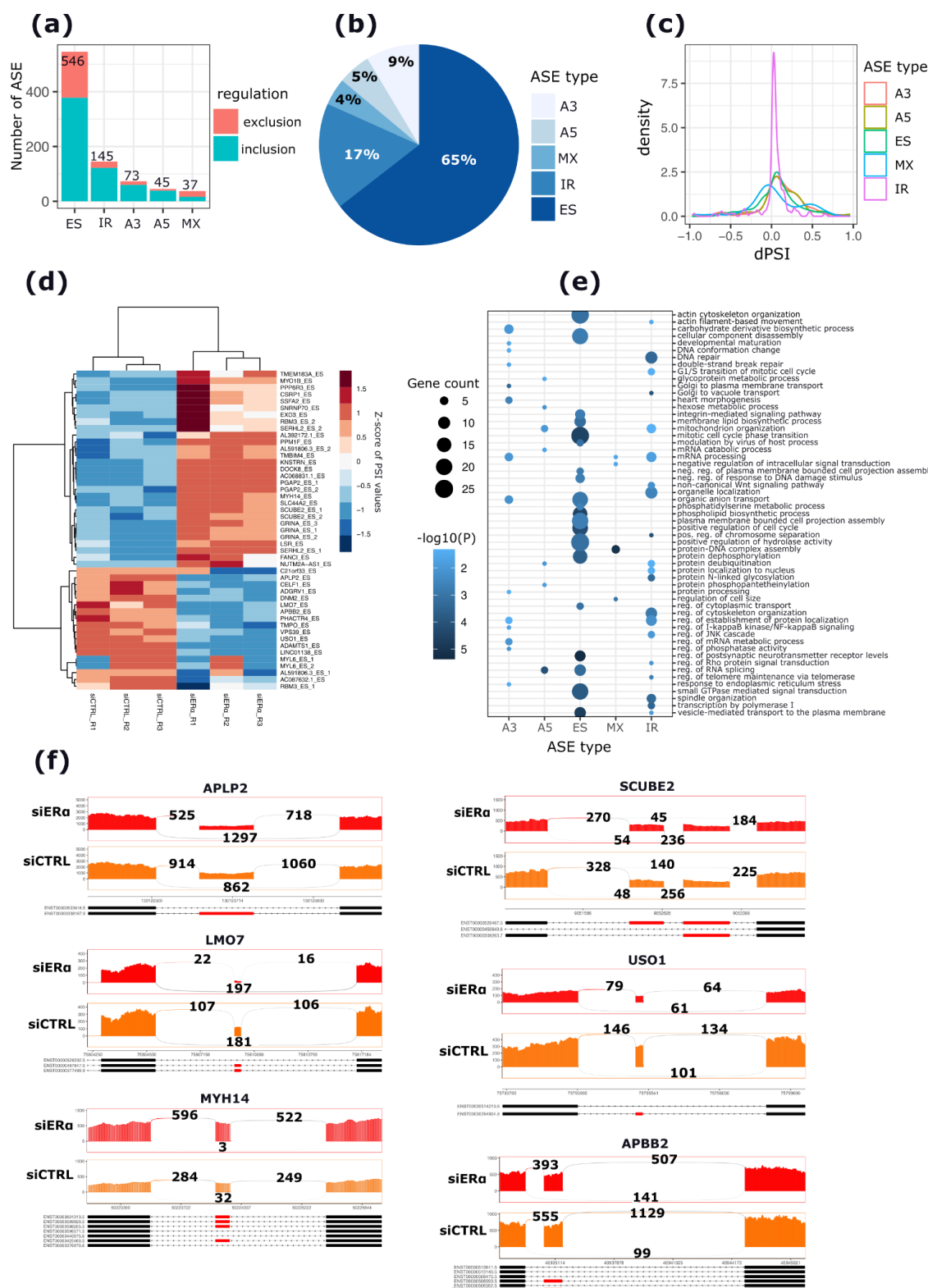


Figure 2.1-6: General overview of differential AS changes occurring upon hormone-independent ER $\alpha$  activity depletion in MCF-7 cells. (a) Stacked bar plots representing the number of significant ASEs divided based on the AS type (x-axis) and the regulation type. Red and green colors represent the number of included and repressed ASEs, respectively. (b) Pie-chart plot representing the percentage of ASE types. (c) Density plot representing the dPSI of the significant ASEs reported in plots (a) and (b). (d) Heatmap plot reporting the dPSI values of the top 50 significant ASEs per each replicate. Color bar intensities are proportional to the inclusion level of each event. (e) Dot plot representing the GO enrichment analysis of genes harbouring significant ASEs. The x-axis represents the different ASE types. The size of the dots is

proportional to the number of genes enriched per each enriched GO term. The color of the dots is proportional to the significance of the enrichment ( $-\text{Log}_{10}(P)$ ). (f) Sashimi plots reporting selected examples of the top significant ASEs. Alternative exons and their flanking constitutive exons involved in each event are reported. The numbers above junctions indicate the total number of reads supporting either inclusion or exclusion of the ASE. At the end of each plot are represented the isoforms affected by the ASE, with exons involved are highlighted in red.

Next, we asked the question whether the observed AS changes induced by silencing ER $\alpha$  are enriched for a specific cellular phenotype. Therefore, we selected to compare the significant ASEs occurring in our dataset with a list of ASEs identified by (Shapiro et al. 2011) as differentially expressed between epithelial and mesenchymal BC cell lines. Interestingly, 26 genes overlapped between the two compared datasets (exact hypergeometric test p-value,  $p < 5.82\text{E-}8$ ) (**Supplementary Figure 3** and **Supplementary Table 6a,b**). In particular, 26 ES and 2 MX events were common to both datasets, of which 19 ES and 1 MX events were coherently regulated and 10 ES events were incoherently regulated between the two datasets. For example, the most significant ES event upon ER $\alpha$  silencing was a skipping event of the 7th exon of the Amyloid Beta Precursor Like Protein 2 (*APLP2*) gene (dPSI = -0.176; adjp = 0) ranked as the top third significant event in the (Shapiro et al. 2011) study and was also skipped (dPSI = -0.53; adjp =  $2.35\text{E-}119$ ) during the EMT process. Similarly, the exon skipping event (dPSI = -0.25; adjp =  $3.70\text{E-}8$ ) in the *USO1* Vesicle Transport Factor (*USO1*) gene due to ER $\alpha$  silencing was also repressed (dPSI = -0.95; adjp =  $6.07\text{E-}16$ ) during the EMT process in (Shapiro et al. 2011). Both events, among others, were confirmed by qRT-PCR in (Shapiro et al. 2011). Furthermore, since the ER $\alpha$ -regulated ASEs seemed to be involved in EMT, we overlapped our list of ER $\alpha$ -regulated ASEs with a previously published high confidence list of ASEs occurring in a 7-days' time course EMT RNA-seq (total) dataset (Yueqin Yang et al. 2016). In this study, authors induced EMT in epithelial cells and measured AS changes over a 7-days' time course to define a list of significantly changing ASEs (Yueqin Yang et al. 2016). Interestingly, this analysis revealed a higher number of overlapping ASEs with our dataset (exact hypergeometric test p-value,  $p < 2.95\text{E-}25$ ), including 60 ES, 10 MX, 27 IR, 5 A3, and 3 A5 events. Of this, 75 (71%) and 30 (29%) ASEs were coherently and incoherently regulated, respectively (**Supplementary Figure 4** and **Supplementary Table 6c,d**). Importantly, similar to our results, genes with strong AS changes during EMT process in both (Shapiro et al. 2011) and (Yueqin Yang et al. 2016) were strongly related to the regulation of the actin cytoskeleton, cell-cell junctions, regulation of cell migration and wound healing. Noteworthy, although most of the gene functions and biological processes affected at the splicing and gene levels were largely similar, the sets of genes undergoing a splicing-level and gene-level changes did not overlap more than expected by chance (**Supplementary Table 7**), suggesting that the phenotypical changes occurring at the splicing-level upon silencing of ER $\alpha$  in MCF-7 cells, notably EMT, occurs in a manner that is parallel to the transcriptional program, and that both gene-level and splicing-level changes may cooperatively and coordinately drive changes to the cell morphology.

Similarly, since ER $\alpha$  depletion strongly hampers the expression of many cell cycle related genes as previously shown by gene-level analysis, we sought to determine among the AS changes driven by ER $\alpha$  silencing in MCF-7 cells, those linked to cell cycle progression and proliferation. Therefore, the set of significant ASEs identified

in our dataset were overlapped with a set of ASEs that were reported as linked to cell cycle progression in HeLa cells by (Dominguez et al. 2016). Dominguez and colleagues identified a set of cell cycle-dependent AS changes affecting approximately 1300 genes that were significantly enriched in cell cycle control. Interestingly, the overlap revealed 69 genes undergoing significant AS pattern changes in both datasets and which were strongly related to cell cycle control terms. This includes the SR kinase protein (*CLK1*) gene, the Anaphase promoting complex subunit 11 (*ANAPC11*) gene, the Cell Division Cycle 25C (*CDC25*), the MDM2 proto-oncogene (*MDM2*) and the Protein kinase, membrane associated tyrosine/threonine 1 (*PKMYT1*) gene. These results suggest the existence of a link between splicing-level regulation and cell cycle control and that this interplay between splicing-level changes and cell cycle progression is potentially under the control of hormone-independent activity of ER $\alpha$  in MCF-7 cells. Furthermore, although a small set of cell cycle related genes underwent both gene-level and splicing-level changes, the genes functions regulated at both levels are similar and converge towards the control of the same biological process, as previously stated.

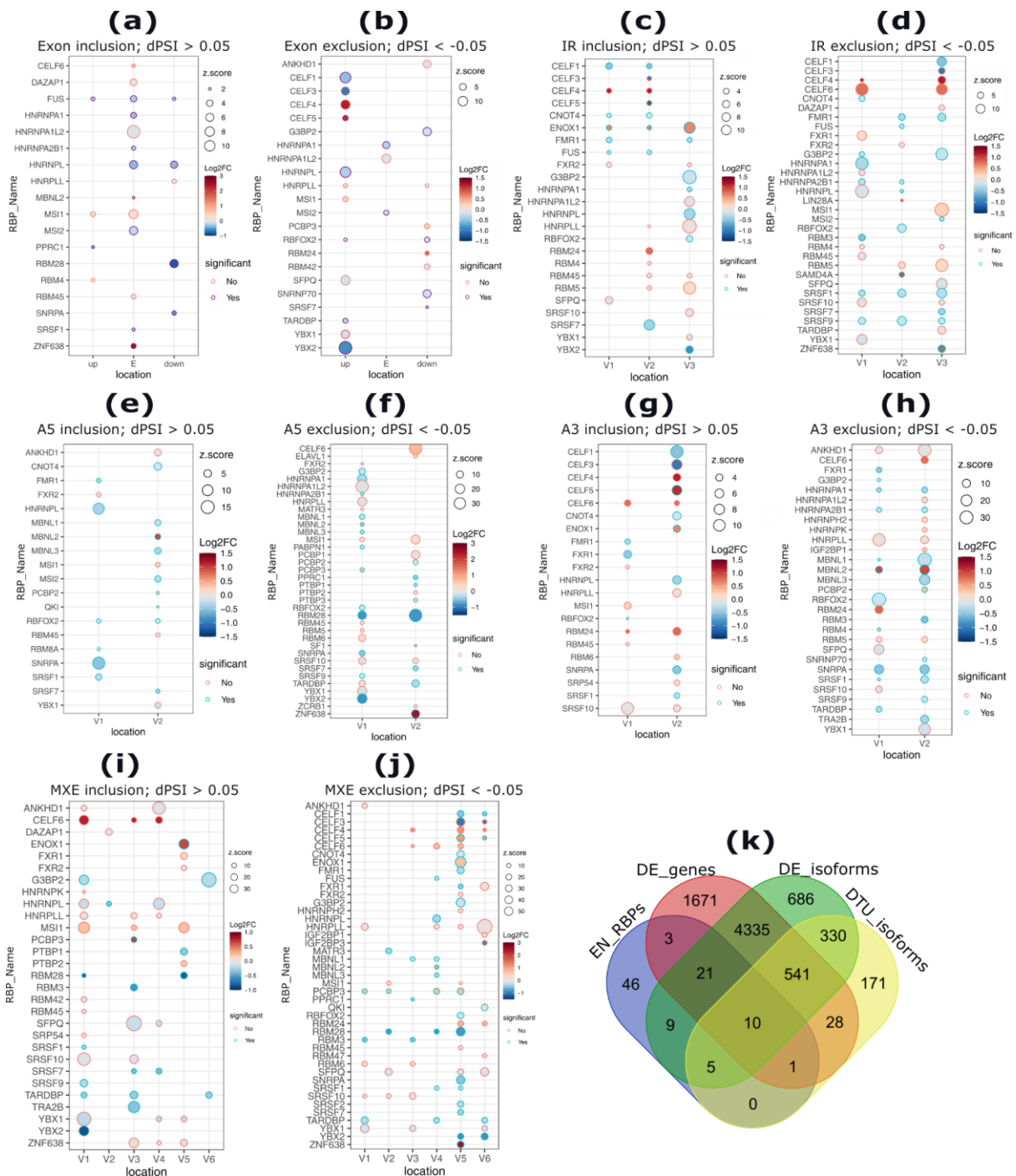
### 2.1.3.1 Identification of putative RBPs regulating the ER $\alpha$ -regulated ASEs

We further analysed the list of the ER $\alpha$ -regulated ASEs to determine SFs that could be potentially involved in the observed AS changes upon ER $\alpha$  silencing. Therefore, a differential RBP-binding motif enrichment analysis was performed by considering each event type, the genomic region involved, and direction of regulation (e.g. dPSI > 0.05, dPSI < -0.05). MoSEA (Sebestyén et al. 2016) was used in this analysis and binding motifs were inferred from position weight matrices (PWM) collected for each SF from the RNAcomplete study (Ray et al 2009, 2017). This analysis revealed 95 enriched SF-binding motifs, including 37 enriched for the ES events (**Figure 2.1-7a,b**), 41 for IR (**Figure 2.1-7c,d**), 49 for A5 events (**Figure 2.1-7e,f**), 61 for A3 events (**Figure 2.2-7g,h**), and 91 SFs enriched for MXE events (**Figure 2.1-7i-j**). Importantly, while showing a preferential binding depending on the direction of regulation and the type of ASEs analysed, most of the enriched SFs were common to different ASE types, with the exception of MX events, where 17 SFs were exclusively enriched. Furthermore, the binding motifs of three SFs (TUT1, TIA1, and TIAL1) were exclusively enriched in A3 events. A summary of this analysis is presented in (**Supplementary Table 8**).

Among the 95 enriched SFs, 49 were DE upon ER $\alpha$  silencing, either at gene or isoform level (**Figure 2.1-7k** and **Supplementary Table 9**). For example, in the case of ES events, binding motifs of 10 significantly DE SFs were enriched in case of inclusion events (n=377) and 12 DE SFs-binding motifs were enriched in the case of exclusion events (n=167) (**Figure 2.1-7b** and **Supplementary Table 9**). Importantly, most of the predicted SFs in both inclusion and exclusion events were significantly downregulated in our dataset, suggesting that the observed AS changes are a consequence of ER $\alpha$  depletion. On the other hand, the enriched SF motifs in case of ES events were prevalently predicted upstream of the exons, while in case of inclusion events, the enriched SF motifs were prevalent within the spliced exons, coherently with a position-dependent effect, as previously reported (Sebestyén et al. 2016). The top 5 most significantly enriched SFs in case of exon inclusion events showed an exonic enrichment. This SF set includes HNRNPA1L2 (z-score = 11.42; not DE) which was previously reported to be overexpressed in ER $\alpha$ + BC tumors as compared to normal adjacent samples

(Sebestyén et al. 2016), MSI1 (z-score > 5.97; log2FC = 0.22; adjp = 0.05) and MSI2 (z-score = 5.06; log2FC = -0.20; adjp = 0.01) both of which were DE at isoform level only; RBM28 (z-score = 4.97; log2FC = -0.82; adjp = 1.88E-23), and hnRNPL which was also downregulated at isoform level only (z-score = 4.87; log2FC = -0.36; adjp = 0.04). Conversely, the top 5 most significantly enriched SFs in case of ES events showed an upstream intronic enrichment and included YBX2 (z-score = 14.0; log2FC = -0.86; adjp = 6.95E-11), followed by CELF1 (z-score = 9.91) whose isoforms (but not the gene) were differentially regulated by ER $\alpha$  silencing (CELF1-204: log2FC = -0.52; adjp = 2.617E-07; CELF1-202: log2FC = 3.47; adjp = 4.89E-07; CELF1-214: log2FC = -0.40; adjp = 0.0027), hnRNPL (z-score = 9.60; log2FC = -0.36; adjp = 0.04), and SFPQ (z-score = 7.52; not DE) (**Supplementary Table 8**).

Noteworthy, among the enriched SFs, 16 showed a significant isoform switching event resulting in significant downstream consequences. In particular, for the splicing factor CELF1, the isoforms showed a significant isoform switching event (CELF1-204: log2FC = -0.52; adjp = 2.617E-07; CELF1-202: log2FC = 3.47; adjp = 4.89E-07; CELF1-214: log2FC = -0.40; adjp = 0.0027). This isoform switching event (CELF1-204 downregulated and CELF1-202 upregulated) had as significant downstream consequences an intron retention loss and a shortening 3'UTR region (Supplementary Table 4a-c). Furthermore, the differential AS analysis using rMATS revealed a significant ES event of the region located at the beginning position of the longest 3'UTR region of the downregulated isoform CELF1-204 (chr11:47471997-47472102, dPSI = -0.167; adjp = 7.69E-07), further supporting the 3'UTR shortening and the IR loss. Noteworthy the two ES events in CELF1-204 isoform reported by rMATS and by dIU analysis were also confirmed using a different AS tool PSIsigma (K.-T. Lin and Krainer 2019) which reported two events involving the shortening of the 3'UTR region of CELF1-204 (event region 1: chr11:47470556-47472190; dPSI=-15.76%, adjp=0.01; event region 2: chr11:47468892-47472190, dPSI = -15.76, adjp = 0.01) and an IR event resulting in the loss of retained intron (IR event region: chr11:47465944-47472102; dPSI = -10%, adp = 0.03) (Supplementary Table 5l). Other examples of isoform switching events involving SF-coding transcripts, included those of MSI2 transcripts showing significant switching events resulting in the downregulation of a non-coding isoform and upregulation of a coding isoform (**Supplementary Table 4a-c**).

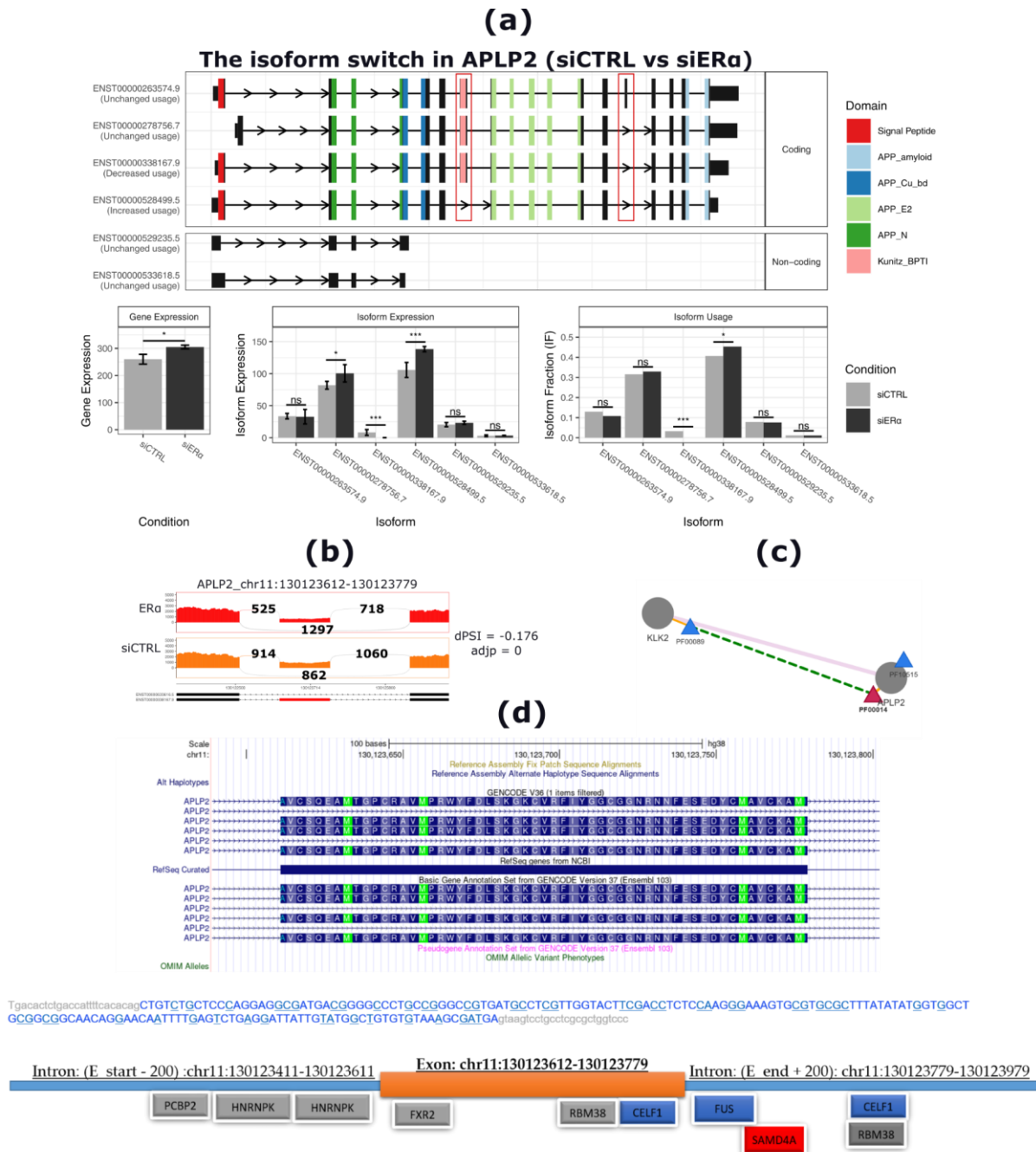


**Figure 2.1-7:** Overview of the RBP-binding motifs enrichment analysis performed on ASEs reported upon ER $\alpha$  silencing in MCF-7 cells. (a-j) dot plots reporting the enriched RBP motifs for ES, RI, A5 'SS, A3 'SS, and MXE events, respectively. The name of enriched RBP is provided on the y-axis of the plot. The x-axis represents the positions where the binding motifs predicted to be enriched (up: upstream intron; E: exon, down: downstream intron). The color intensities represent the DE status of enriched RBPs (red for upregulated and blue for downregulated). The significance of the DE status of the RBP is represented by the border color of the dot. The size of the dot is proportional to the enrichment z-score (sig:  $z > 1.96$ ). (V1 to V3) in (c,d) represent upstream, within the intron, or downstream exon, respectively. (k) Venn Diagram representing the number of enriched RBPs and whether they are or not regulated at gene or isoform level.

Noteworthy, the differential AS and dIU analyses agreed on a significant number of events. For instance, 98 genes harbouring an ES event were considered as significantly regulated by both analyses. For example, the

top significant ES event (*APLP2*\_chr11:130123611-130123779; dPSI = -0.176; adjp = 0) was also confirmed by dIU analysis, which reported a significant increase of isoform *APLP2*-210 (lacking the exon) and a decrease of isoform *APLP2*-201 (including the exon). Moreover, the two switching isoforms differ by the differential inclusion of another exon which was also reported as differentially spliced by rMATS analysis (exon 14, event: *APLP2*\_chr11:130137255-130137291; dPSI = -0.05; adjp = 9.46E-05). Interestingly, further annotation of this event reveals that the exon (*APLP2*\_chr11:130123611-130123779) encodes for the Kunitz/Bovine Pancreatic Trypsin Inhibitor protein domain (*Kunitz\_BPTI*) that is involved in a protein-protein interaction of *APLP2* protein with Kallikrein related peptidase 2 (*KLK2*) protein (**Figure 2.1-8a,b** and **Supplementary Table 9**). Moreover, the annotation of this skipped exonic region using DIGGER database (Louadi et al. 2021) reveals that it encodes for the residues that lie on the interaction interface of the two proteins (**Figure 2.1-8c**).

Furthermore, to completely illustrate the putative ER $\alpha$ -mediated regulation of the exon skipping event in the *APLP2* gene and exemplify the regulatory mechanisms of other ASEs, we integrated the RBP-binding motif enrichment analysis together with the gene-level and splicing-level changes of this event. Notably, RBP-binding motif enrichment analysis revealed a number of RBP binding sites as enriched within and around the spliced exon of *APLP2* gene. The top significant predictions revealed PCBP2 and hnRNPK to have the closest binding motifs at the upstream intron (200 nt upstream of exon start), FXR2 was enriched at the start of the exon, followed by RBM38 and CELF1 which had the binding at the end of the exon, FUS, SAMD4A, CELF1, and RBM38 at the downstream intron (200 nt downstream of the exon end) (**Figure 2.1-8d**). The enrichment results including the position, region having predicted binding motif, score and significance of the predictions for the top ranked motifs are summarized in (**Table 2.1**). Importantly, among the top 5 enriched SFs for this event, three were DE upon ER $\alpha$  silencing, including SAMD4A which was upregulated, Fus which was downregulated at gene-level and CELF1 which had a significant isoform switching and whose binding motif had the highest motif prediction score. The genome browser representation of the genomic region of *APLP2* skipped exon and flanking introns, with the annotation of predicted SFs is reported in Figure 2.2-8d. This example of *APLP2* exon skipping illustrates a possible mechanism through which ER $\alpha$  controls the post transcriptional events of a target gene.



**Figure 2.1-8:** Post-transcriptional events involving APLP2 gene as reported by IsoformSwitchAnalyzeR (a) and by rMATS (b). Functional consequences of post-transcriptional events at the protein level as reported by DIGGER analysis are reported in (c). The downstream consequences of the switching APLP2 isoform pairs at protein level. Dashed green line in (c) highlights the affected PPI due to the skipping of the exon. Pink solid line highlights the PPI interaction evidenced from the DIGGER database. (d) Top: Genome browser representation of the spliced exon and the flanking introns. The sequence of the exon is shown in blue in the middle of the plot and first bases of introns in grey. Bottom: A predictive model hypothesising the putative SFs regulating the spliced exon based on the RBP-binding motif enrichment analysis. In grey, blue and red colors are highlighted the not DE, downregulated, and upregulated SF genes upon ERa silencing, respectively.



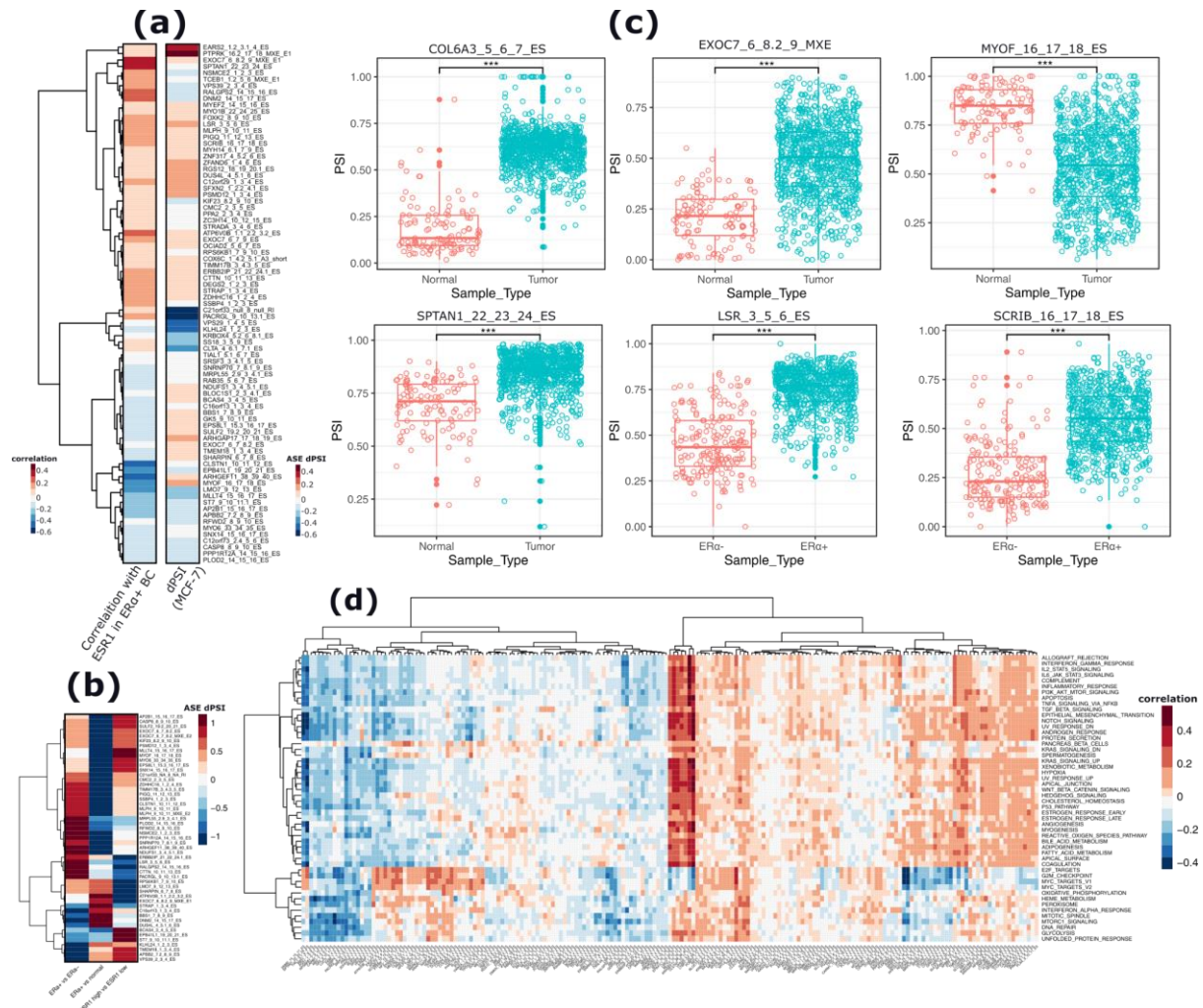
Table 2.1-1: Top significant RBP binding motifs predicted around the spliced in APLP2 gene.

RBP	Position	Location	Score	adj.p-value	Motif
PCBP2	102-108	Up intron	10.63	7.06E-05	CCTCCCC
HNRNPK	108-114	Up intron	8.95	0.000276	CCAGCCC
HNRNPK	113-119	Up intron	10.20	8.35E-05	CCATCCC
FXR2	22-28	Exon	9.26	0.00029	TGACGGG
RBM38	151-157	Exon	9.55	0.0002	CTGTGTG
CELF1	152-158	Exon	10.95	6.12E-05	TGTGTGT
FUS	12-18	Down intron	8.98	0.000383	CTCGCGC
SAMD4A	17-23	Down intron	10.14	0.0001	GCTGGTC
RBM38	82-88	Down intron	9.54	0.0002	CTGTGTG
CELF1	83-89	Down intron	10.82	9.73E-05	TGTGGTG

### 2.1.3.2 ER $\alpha$ -regulated exons are also differentially expressed in primary tumor tissues and correlates with ER $\alpha$ expression status

The set of ER $\alpha$ -regulated ASEs were further explored in 965 primary breast tumor samples including 773 ER $\alpha$ +, 192 ER $\alpha$ - and 113 adjacent normal samples (**Supplementary Table 10a**). First, the events were explored using the TCGASpliceSeq database (M. Ryan et al. 2016) which holds information on spliced exons and their inclusion (PSI) level in every TCGA sample for 33 different tumor types and when available adjacent normal samples (**Supplementary Table 10b-d**). This step resulted in the identification of 228 (28%) ASEs also expressed in breast tumors and normal samples (**Supplementary Table 10b-d**). Interestingly, among the identified events, 81 ASEs were significantly correlated with *ESR1* gene expression in the ER $\alpha$ + tumor samples, including 50 ASEs positively correlated and 38 negatively correlated (**Figure 2.1-8a** and **Supplementary Table 10e**). The most significantly correlated events include an ES event in the Calsyntenin-1 gene (*CLSTN1*) (event: CLSTN1\_10\_11\_12\_ES; Spearman rho correlation = -0.44; Spearman p-value = 4.82E-37; dPSI = -0.05), and an ES event in the myoferlin (*MYOF*) gene (event MYOF\_16\_17\_18\_ES; Spearman rho correlation = -0.36; Spearman p-value = 4.28E-24; dPSI = 0.153). A summary of ASEs showing a significant correlation with *ESR1* gene expression in ER $\alpha$ + tumors is reported in (**Figure 2.2-8a** and **Supplementary Table 10e**). Furthermore, ER $\alpha$ + patients were classified using the median expression value of *ESR1* gene as a threshold into (i) highly *ESR1* expressing patients (*ESR1* high, n = 386), and (ii) lowly *ESR1*

expressing patients (*ESRI* low, n =387). Then, a dPSI value representing the differential inclusion level of each ER $\alpha$ -regulated ASE was calculated among the different groups of patients. This analysis revealed 140 ASEs as characterized by significantly different inclusion/exclusion levels ( $|\text{dPSI}| > 0.05$ ; Wilcoxon p-value < 0.05) among the different comparisons, including 53 ASEs were differentially expressed between ER $\alpha$ <sup>+</sup> and ER $\alpha$ <sup>-</sup>, 72 ASEs between tumor and normal samples, and 15 ASEs between *ESRI* high and *ESRI* low patients (**Supplementary Table 10f**). Moreover, the inclusion/exclusion level of 12 ASEs was significantly different in all the three comparisons (**Figure 2.1-8b** and **Supplementary Table 10f**), all of which were significantly correlated with ER $\alpha$  gene expression in ER $\alpha$ <sup>+</sup> tumors (**Supplementary Table 10f**). In particular, in the ER $\alpha$ <sup>+</sup> tumor versus normal samples comparison, the sixth exon (chr2:238287279-238287878) of the collagen type VI alpha 3 chain (*COL6A3*) gene involved in the ES event (*COL6A3\_5\_6\_7\_ES*), exhibited a significant differential inclusion level between the two groups of subjects, where it was more included in tumors (dPSI = 0.49; Wilcoxon p-value = 2.34E-62), followed by an ES event (*COL6A3\_2\_4\_5\_ES*) involving the same gene *COL6A3*. This event corresponds to the skipping of the third (chr2:238303230-238303847) and fourth (chr2:238296225-238296827) exons of the *COL6A3* gene; by connecting second to the fifth exon. The event was more included in tumors as compared to normal samples (dPSI = 0.21; Wilcoxon p-value = 1.54E-59). Conversely, the top third significant event in this comparison was an ES event (*CLSTN1\_10\_11\_12\_ES*) of exon 11 (chr1:9797556-9797612) of the Calsyntenin 1 (*CLSTN1*) gene, which was more included in normal samples (dPSI = -0.32; Wilcoxon p-value = 4.12E-53). Furthermore, among the 53 ASEs that exhibited differential inclusion/exclusion levels between ER $\alpha$ <sup>+</sup> and ER $\alpha$ <sup>-</sup> patients, 29 ASEs were also significantly correlated with *ESRI* gene expression (**Supplementary Table 10f**). Finally, the 15 ASEs which were significantly differentially expressed between samples high *ESRI* and low *ESRI* samples, were also significantly correlated with *ESRI* expression in ER $\alpha$ <sup>+</sup> tumors (**Supplementary Table 10f**). Selected examples of these ASEs are reported in **Figure 2.2-9c** and the full list is provided in (**Supplementary Table 10f**).



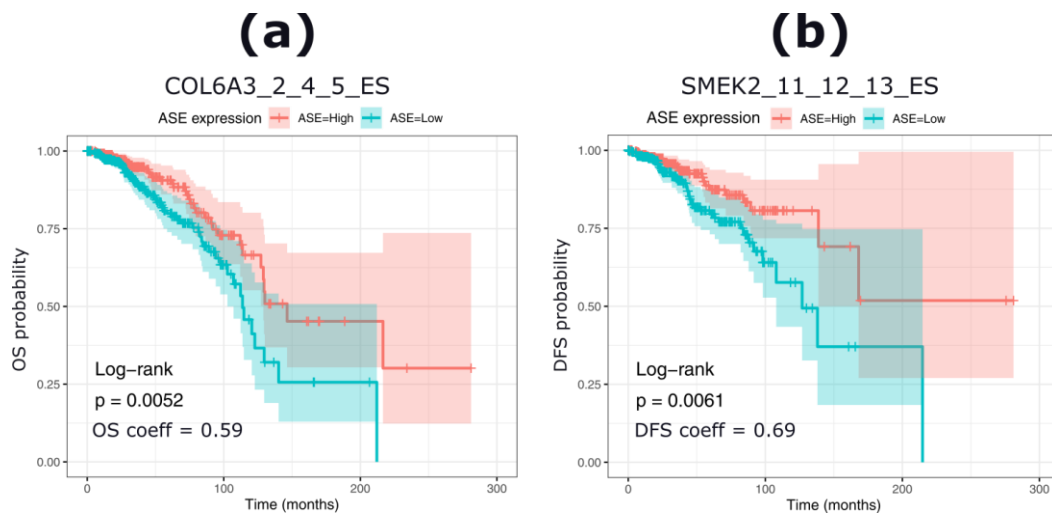
**Figure 2.1-9:** Overview of the comparative analysis of the expression of ER $\alpha$ -regulated AEs in normal and tumor breast tissues. (a) Heat map plot reporting the dPSI levels in MCF-7 as well as the correlation coefficient of the of AEs expression (PSI value) with ESR1 gene expression in ER $\alpha$ + tumor samples retrieved from the BRCA TCGASpliceSeq database. The events are labeled as gene name, exons involved in the event (regulated and flanking exons) and the type of event. (b) Heat map plot reporting dPSI values of AEs compared among (i) ER $\alpha$ + vs ER $\alpha$ -, (ii) ER $\alpha$ + vs normal and (iii) high ESR1 vs low ESR1 expressing patients. The plot reports 51 AEs which were significant based on ( $p$ -value < 0.05) in all comparisons. (c) Box plots reporting the PSI values of selected AEs whose inclusion levels are different among compared groups (Tumor vs Normal) and (ER $\alpha$ + vs ER $\alpha$ -) patients. The names of AEs are shown on the top of each boxplot, Wilcoxon  $p$ -value (\*\*\*,  $p$  < 0.00001). (d) Heat map plot representing the correlation of AEs PSI values of ER $\alpha$ -regulated AEs with enriched hallmarks as reported using PEGASAS algorithm (Phillips et al. 2020).

Moreover, to get a functional enrichment of the ER $\alpha$ -regulated AEs with respect to molecular pathways, an analysis with PEGASAS (Phillips et al. 2020) was performed considering the expression levels of regulated exons represented by PSI values measured in TCGA BRCA RNA-Seq samples. Interestingly, this analysis revealed two main clusters of molecular pathways. Particularly, the first main cluster was related mainly to EMT-related terms such as TGFB signaling pathway, EMT, apical junction, KRAS signaling\_up, estrogen response and cholesterol homeostasis pathways. Conversely, the second cluster was mainly related to cell cycle-related and metabolism and included terms such as DNA repair, E2F\_targets, G2M checkpoint, Myc

targets V1 and V2, mitotic spindle, in addition to metabolism-related terms such as oxidative phosphorylation, and glycolysis. Interestingly, a particular group of 7 ES events including (PLOD2\_14\_15\_16\_ES, MYOF\_16\_17\_18\_ES, EPB41L1\_19\_20\_21\_ES, LMO7\_9\_12\_13\_ES, MLLT4\_15\_16\_17\_ES, ARGEF11\_38\_39\_40\_ES, and CLTSTN1\_10\_11\_12\_ES) were exclusively positively correlated with molecular pathway terms of the first cluster (e.g. EMT related terms, TGF $\beta$  signaling early and late estrogen response, angiogenesis, adipogenesis) while negatively correlated with terms of the second cluster (e.g. E2F targets, G2M checkpoints, Myc\_targets\_V1 and Myc\_targets\_V2), with the exception of some terms of this cluster including DNA repair and glycolysis molecular pathways with which those events were positively correlated. Noteworthy, these ASEs exhibited both a differential inclusion/exclusion level between tumor and normal samples and significantly correlated with *ER $\alpha$*  gene expression. Conversely, the two splicing events DN2M2\_14\_15\_17\_ES and SPTAN1\_22\_23\_24\_ES while negatively correlated with all terms of cluster 1, they showed a positive correlation with the four cell cycle-related terms of cluster 2 (E2F targets, G2M checkpoints, Myc\_targets\_V1 and Myc\_targets\_V2) (**Figure 2.1-9d and Supplementary Table 10e**). Altogether, this data supports that *ER $\alpha$*  controls the EMT and cell cycle processes in BC not only by a gene-level control, but also by a splicing-level regulation.

### 2.1.3.3 Prognostic value of ASEs regulated by the hormone-independent activity of *ER $\alpha$* in MCF-7 and their association with clinical breast cancer subtypes

To determine the prognostic value of the significant ASEs induced by *ER $\alpha$*  silencing, we sought next to explore whether they are associated with overall survival (OS) and disease-free survival (DFS) times in *ER $\alpha$* <sup>+</sup> patients. Therefore, overall and disease-free survival information of 773 *ER $\alpha$* <sup>+</sup> samples were retrieved from TCGA GDC portal (Grossman et al. 2016b), together with PSI values, representing the expression levels of the ASEs, which were retrieved from the TCGASpliceSeq database (M. Ryan et al. 2016). Samples with a PSI value greater than the median were classified as highly expressing the ASE and samples with PSI value less than this threshold were classified as lowly expressing the ASE. This analysis revealed 12 ASEs (11 ES and 1 A3'SS events) as significantly associated with overall survival of the patients (Supplementary Table 10h). In particular, a higher inclusion of the exon (chr2:238296225-238296827) of the *COL6A3* gene is significantly associated with longer patient OS time (OS coefficient = 0.59; log-rank p-value = 0.0052) (Figure 2.2-10a and Supplementary Table 10h). Similarly, a higher inclusion of the remaining 11 ASEs events is also associated with a longer patient OS. Noteworthy, two ES events in the *COL6A3* gene (*COL6A3\_2\_4\_5\_ES*, *COL6A3\_5\_6\_7\_ES*) were associated with a longer patient OS. On the other hand, 7 ASEs (6 ES and 1 A3'SS events) significantly associated with patients DFS (**Supplementary Table 10h**). In particular, 4 ASEs associated with a better DFS and 3 ASEs associated with a worse DFS. Notably, a higher inclusion level of the exon (chr2:55573619-55573777) of the Protein Phosphatase 4 Regulatory Subunit 3B (*PPP4R3B/SMEK2*) gene was associated with both better patient DFS (HD = 0.69; log-rank p-value = 0.0061) as well as a longer OS (HD = 0.42; log-rank p-value = 0.04) (**Figure b and Supplementary Table 10h**).



**Figure 2.1-10:** Survival analysis plots reporting the top significant ASEs associated with overall survival (a) and disease free (DFS) survival (b) of ER $\alpha$ + patients, respectively. Patients are divided into highly expressing (ASE=high) and lowly expressing the event (ASE=low) based on the median PSI calculated among all patients. ASE, Alternative Splicing Event; ES: Exon Skipping. OS, Overall Survival; DFS, Disease Free Survival. Event names are provided on the top of each plot.

### 2.1.4 Discussion

In the present study, hormone-independent ER $\alpha$  activity was explored in MCF-7 cells at both transcription and posttranscriptional levels. In addition, an important set of relevant AS-level ER $\alpha$ -mediated regulation events were also confirmed in tumor tissues and their prognostic value in BCs was investigated. On one hand, the crucial role of ER $\alpha$  in maintaining the expression of epithelial genes, and promoting cell cycle progression in the investigated cells was evidenced through the analysis of a high quality paired-end RNA sequencing experiment in triplicates. Clearly, ER $\alpha$  silencing significantly hampered the expression of cell cycle-related genes, essential for cell proliferation and survival, while promoting an increase in the expression and activity of a number of mesenchymal markers, which may result in a more mesenchymal-like phenotype in surviving cells. Interestingly, ER $\alpha$  silencing significantly repressed the expression of RBP genes, including SF genes, which resulted in a significant regulation at the level of isoforms. Notably, in a number of cases, ER $\alpha$  silencing had a different effect on the expression of isoforms than that at the gene level as shown by dIE analysis at the level of isoforms, by dIU analysis using IsoformSwitchAnalyzeR (Vitting-Seerup and Sandelin 2019) and by differential AS analysis using the event-based approach, rMATS (Shen et al. 2014) which further supported dIU analysis. On the other hand, our applied computational pipeline sheds light on the importance of considering the analysis at the level of isoforms, rather than limiting the attention at the gene level, as previously reported (Soneson, Love, and Robinson 2015). Moreover, making use of publicly available databases such as DIGGER (Louadi et al. 2021), and APPRIS (Rodriguez et al. 2018) databases, the effects of ER $\alpha$  depletion in MCF-7 cells were explored at the protein isoform level, revealing that complex mechanisms at the level of RNA transcripts such as differential AS, differential promoter usage, and differential last exon usage, drive the expression of specific protein isoforms that may functionally be different.

Analysis at gene level shows that hormone-independent activity of ER $\alpha$  is crucial for cell proliferation and for maintaining an epithelial-like luminal phenotype of MCF-7 cells. Notably, several cell cycle related genes such as the E2F transcription factor 1 (*E2F1*) gene, the checkpoint kinases 1 (*CHEK1*) and 2 (*CHEK2*) genes, Cyclin dependent kinases 1 (*CDK1*), 2 (*CDK2*), 4 (*CDK4*), 6 (*CDK6*), and 7 (*CDK7*), and Minichromosome Maintenance Complexes 3 (*MCM3*), 4 (*MCM4*), 5 (*MCM5*), 6 (*MCM6*), 7 (*MCM7*) and 10 (*MCM10*) were significantly depleted by silencing ER $\alpha$ . Noteworthy, all of the aforementioned genes were induced under 17 $\beta$ -estradiol stimulation of MCF-7 cells, as reported in (Duoqiao Chen et al. 2021) and in (Dago et al. 2015). On the other hand, other genes known to be involved in EMT such as the Tumor Growth Factors Beta 1 (*TGFB1*), 2 (*TGFB2*), 3 (*TGFB3*) and their receptors type 1 (*TGFBR1*), 2 (*TGFBR2*) and 3 (*TGFBR3*), the CD44 molecule (Indian blood group) (*CD44*) gene, the collagen type V alpha 1 chain (*COL5A1*) gene, type VI alpha 1 (*COL6A1*) and 2 chain (*COL6A2*), and the Filamin A (*FLNA*) gene were significantly induced by silencing ER $\alpha$  (Vasaikar et al. 2021). Such a transition from epithelial to mesenchymal phenotype was indeed reported in a study by (Bouris et al. 2015) showing that stably transfected MCF-7 cells by knocking down ER $\alpha$  gene using specific shRNA lentiviral particles results in a potent EMT induction and changes in the expression and activity of matrix macromolecules, and that the loss of ER $\alpha$  promotes BC cell migration and invasion.

Therefore, our experiment consisting of a transient transfection of MCF-7 cells with ER $\alpha$ -specific siRNAs in hormone-deprived culture conditions mimicked the effects of the stable transfection in (Bouris et al. 2015). In addition, our attention was particularly attracted by observing that among the ER $\alpha$ -regulated genes, a significant number of RBP genes including SF genes, key players at the post transcriptional level and inducers of cellular phenotype, were also regulated. Thus, ER $\alpha$  activity at the splicing-level was investigated. Indeed, an important number of studies compared the transcriptome of human breast tumors versus healthy matched tissues and revealed that approximately 50% of the protein-coding genes have an altered transcript variant expression (Wen et al. 2015; Stricker et al. 2017). These observations were further supported by recent *in vitro* findings that show a significant switch in the splicing pattern during EMT that is accompanied with a specific EMT splicing signature (Shapiro et al. 2011). Importantly, this switch in splicing pattern was correlated to the concentration level and activity of specific SF genes, especially in cancer (S. Park et al. 2019). In particular, ER $\alpha$  silencing repressed the expression of epithelial-specific RBP genes including the PHD finger protein 5A (*PHF5A*), previously identified as an oncogene frequently upregulated and associated with poor survival in BC (Y.-Z. Zheng et al. 2018). knocking down this gene significantly suppressed cell proliferation and increased apoptotic signalling by promoting the expression of a short truncated Fas-activated serine/threonine kinase isoform, enabling Fas-mediated apoptosis in BC cells (Y.-Z. Zheng et al. 2018). Furthermore, *PHF5A* also acted as an oncoprotein in lung carcinoma (Yan Yang et al. 2018), and promoted lung cancer progression by regulating AS (Mao et al. 2019). ER $\alpha$  silencing also significantly decreased the expression of the nucleolar-related dyskerin pseudouridine synthase 1 (*DKCI*) encoding gene which was proposed as a prognostic marker in BC patients and associated with poor patient outcomes (Elsharawy et al. 2020). Other groups have reported that *DKCI* overexpression conferred a more aggressive phenotype and increased intrinsic ribosomal activity in cells derived from normal breast epithelium (Guerrieri et al. 2020). Furthermore, the depletion of hormone-independent ER $\alpha$  activity in hormone-deprived MCF-7 cells caused a significant decrease in the expression of an important number of SF genes related to epithelial phenotype. Among others, this includes the two Epithelial Splicing Regulatory Protein 1 (*ESRP1*) and 2 (*ESRP2*) genes, the core regulators of AS in epithelial cells (Warzecha et al. 2010; Shapiro et al. 2011; Yae et al. 2012). Since these two factors were reported to play a crucial role during EMT, we particularly analyzed the binding of ER $\alpha$  at the vicinity of these genes and found that ER $\alpha$  indeed binds these two genes in absence of estrogenic stimuli and these binding results in the expression of both genes (Supplementary Figure 6 and Further discussion in the next chapter). Other examples of downregulated SF genes included the Muscleblind-like 1 (*MBNL1*) gene which was found to act as a tumor suppressor in BC (Fish et al. 2016) by controlling AS, translation and RNA decay through binding of transcripts at their 3'UTRs (Masuda et al. 2012; Batra et al. 2014). Similarly, ER $\alpha$  silencing decreased the expression of *MBNL3* gene, which is downregulated during EMT of epithelial BC cells (Shapiro et al. 2011; Lu et al. 2015). The expression of the Transformer 2 beta homolog (*TRA2B*) gene which was previously characterized to act as an oncogene in BC, and regulates the splicing pattern of *CD44* gene involved in EMT (Watermann et al. 2006) was also decreased by silencing ER $\alpha$ .

On the other hand, ER $\alpha$  silencing induced the expression of 30 SF genes previously reported to be higher expressed in mesenchymal phenotype in BC (Shapiro et al. 2011). This includes the Cytoplasmic Polyadenylation Element Binding Proteins 1 (*CPEB1*) gene (top significant DE SF gene in our dataset), 2 (*CPEB2*), and 4 (*CPEB4*), the eukaryotic translation initiation factor 4E family member 3 (*EIF4E*) gene, essential factors for RNA translation by controlling polyadenylation tails and 3'UTR length of certain EMT- and metastasis-related genes (Nagaoka et al. 2016). Similarly, a significant increase was also observed in the expression of the Splicing factor 3b subunit 1 (*SF3B1*), frequently mutated in ER $\alpha$  positive BCs and associates with aberrant splicing and poor patient's prognosis (Maguire et al. 2015; X. Fu et al. 2017). Other upregulated RBP genes affecting other aspects of RNA processing included the Programmed cell death 4 (*PDCD4*) gene which was previously reported to be downregulated in BCs by miR-21 and its low levels associated with aromatase inhibitor resistance, with HER2-positive status, and a poor prognosis in ER $\alpha$ -positive BCs (Z. Chen et al. 2015), and was later reported to be regulated by miR-421 (Y. Wang, Liu, and Shen 2019). ER $\alpha$  silencing also induced the expression of other interesting genes involved in RNA splicing process such as the Pre-mRNA Processing Factor 3 (*PRPF3*) and 6 (*PRPF6*) genes which encode for nuclear proteins core components of the U4/U5 and U4/U6&U5 spliceosomal complexes (Lauber et al. 1997; S. Liu 2006), and *PRPF8* which was identified as essential for the BRCA1-mediated homologous recombination (Onyango, Lee, and Stark 2017). ER $\alpha$  silencing also induced the expression of QKI, KH domain containing RNA binding (*QKI*) gene, which was previously considered as the most important RBP regulating the expression of linear and circular RNA transcripts during EMT in human mammary epithelial cells (Conn et al. 2015). *QKI* was also found to positively correlate with the expression of EMT markers and its high expression associated with worse overall and disease free survival times in BC patients (S. Gu et al. 2019). Taken together, our findings support a crucial role of the hormone-independent activity of ER $\alpha$  in maintaining the luminal epithelial phenotype in BC by promoting the expression of epithelial SF genes and preventing the expression of mesenchymal SF genes. Specifically, ER $\alpha$  silencing in MCF-7 cells induced not only a gene-level regulation of SF genes but also a splicing-level regulation. Notably, ER $\alpha$  silencing induced a splicing pattern switch in 20 SF genes. In particular, an isoform switching event in *CELF1* isoform pairs resulted in the upregulation of the isoform with shorter 3'UTR and downregulation of the isoform with longer 3'UTR. Noteworthy, although no regulation at the gene level was observed, four *CELF1* isoforms were regulated and responded in opposite directions to ER $\alpha$  silencing, which explains the none overall change at the gene level. Strikingly, analysis of ChIP-seq data revealed that the spliced out (intron + 3'UTR) region of *CELF1* transcript overlaps with the binding of several transcription factors including ER $\alpha$ , CTCF, TRIM24, SPDEF, AHR, DNMT3A, RARG, and TP63 (Supplementary Figure 5), most of which are DE upon ER $\alpha$  silencing. Moreover, the switching *CELF1* isoform pairs encode two protein isoforms that differ in their sequence by an hydrophobic Alanine (A p.104) residue at position 104, which overlaps with a splice site (Ota et al. 2004). Interestingly, RBP-binding motif enrichment analysis revealed strong enrichment of the *CELF1* binding motif in 320 ASEs, in particular ES events (Supplementary Table 11). Moreover, predicted binding sites on 179 ER $\alpha$ -regulated ASEs overlapped with *CELF1* binding peaks reported by an CLIP-Seq experiment in HeLa cells (Le Tonquèze et al. 2016).



Noteworthy, this comparison confirmed the predicted binding site of CELF1 at the alternatively spliced exon 7 of *APLP2* gene presented in earlier sections (**Supplementary Table 11**). The genomic distribution of the predicted CELF1 binding sites on our list of ASEs revealed that there are about twice as many intronic than exonic bindings of CELF1, in line with the genomic distribution of the binding clusters reported by (Le Tonquèze et al. 2016). Thus, ER $\alpha$ -mediated regulation of CELF1 at both gene and splicing levels could explain in part the observed AS changes identified in our dataset. Furthermore, Le Tonquèze and colleagues identified a high number of CELF1 binding sites within the 3'UTR regions of target transcripts, one of the significantly regulated events in our dataset. We identified that 69 genes showing a 3'UTR shortening/lengthening event overlapped with the binding sites on 3'UTRs identified by the CLIP-Seq experiment in HeLa cells (Le Tonquèze et al. 2016), further suggesting that CELF1 could be premoninly involved in the splicing-level events identified in our dataset. CELF1 was reported to be upregulated upon TGF $\beta$  treatment of epithelial MCF10A and MCF-7 cells and regulated a set of EMT drivers post transcriptionally by binding to GRE-rich elements, typical binding sites for CELF1 and MBNL1, on their 3'UTRs (Chaudhury et al. 2016; Van Nostrand et al. 2020).

Moreover, isoform switching analysis on our RNA-seq dataset revealed different aspects of post transcriptional and splicing-level regulation exerted by ER $\alpha$  on target genes. In particular, the 3'UTR lengthening/shortening events showed as the most significant posttranscriptional events induced by ER $\alpha$  silencing. The functional importance of such mechanisms has been previously reported in the context of cancer (Stumpf et al. 2013), notably in epithelial BC cells under EMT-inducing treatments (Chaudhury et al. 2016), as well as in other cancer types, where it was considered as a recurrent event involved in cancer development and progression (Xue et al. 2018). Recent studies have reported 3'UTR lengthening events as a novel mechanism involved in the process of cellular senescence, and that genes in senescent cells tended to use distal APA sites, leading to a global 3'UTR lengthening. These global 3'UTR lengthening events were associated with a decreased gene expression of senescence-associated genes (M. Chen et al. 2018). Other recent studies have shown that 3'UTR length differs among ER $\alpha$ <sup>+</sup> and ER $\alpha$ <sup>-</sup> BC subtypes and that 3'UTR shortening events contributes to tumor growth by interfering with the stability of an endogenous competitive RNA (ceRNA) network in ER $\alpha$ <sup>-</sup> tumors, especially in association with the aggressive and metastatic phenotypes (Z. Fan et al. 2020). In line with (M. Chen et al. 2018), silencing ER $\alpha$  in MCF-7 cells under hormone deprivation induced a global 3'UTR lengthening than shortening events. Importantly, genes with a 3'UTR lengthening event were significantly enriched in terms related to regulation of cellular response to stress (13 genes), DNA checkpoint (6 genes), positive regulation of cell cycle (8 genes), cell junction organization (11 genes), and protein localization to membrane (11 genes). However, no clear association was observed between 3'UTR lengthening events and gene expression changes (92 upregulated and 89 downregulated). In contrast, genes harbouring a 3'UTR shortening event tended towards a decreased expression (21 downregulated vs 9 upregulated). Moreover, genes harbouring a 3'UTR shortening event were enriched in terms related to actin filament organization, RNA splicing, and developmental maturation. Therefore, we foresee that this posttranscriptional events involving

3'UTRs could be an important mechanism underlying isoform regulation by ER $\alpha$  in BC and further investigation of these events could reveal important aspects of RNA biology in BC.

Furthermore, using MCF-7 and siRNA-mediated silencing of ER $\alpha$  and RNA sequencing experiments, we successfully identified a set of ASEs whose expression is dependent on the expression of ER $\alpha$  in MCF-7 cells. Interestingly, a number of ASEs were confirmed to be differentially spliced in BCs. For instance, the top significant ASE regulated in MCF-7 cells corresponds to the ES event of the 7<sup>th</sup> exon of the amyloid beta precursor like protein 2 (*APLP2*). The same exon was previously reported by AS-sensitive microarrays to be differentially spliced between MCF-7, MDA-MB-231 in one hand and the human mammary epithelial cells (HMEC) on the other hand, showing a more inclusion level in MCF-7 as compared to other cell lines under 2D and 3D cell cultures and in vivo mice models (C. Li et al. 2006). However, the mechanisms involved in these differences were not illustrated. The *APLP2* gene encodes for a transmembrane glycoprotein associated with Alzheimer's disease (Suzuki et al. 1997) and is recently shown to be dysregulated in a number of cancers, including neuroblastoma (Adlerz et al. 2003), colon cancer, where it induces an increase in cell proliferation (Moss, Doran, and Macmathuna 2007), in Ewing's sarcoma (Peters, Yan, and Solheim 2013), and melanoma (Tuli et al. 2009), and was shown to increase cell growth and migration in pancreas cancer (Pandey et al. 2015). The 7<sup>th</sup> alternative exon of *APLP2* gene is a 168 nucleotides exon encoding for the Kunitz protease inhibitor domain (Sandbrink, Masters, and Beyreuther 1994) and its differential AS results in two protein isoforms either lacking or containing the entire Kunitz protease inhibitor domain. The ES events of this exon and the one of exon 14 were reported to be recurrently misspliced in a panel of cancers including LIHC, KIRC, LUAD, and LUSC when matched to their adjacent normal samples, as reported in (Danan-Gotthold et al. 2015; Sebestyén et al. 2016). Our differential AS analysis indicates that this 168 nucleotides exon is skipped upon silencing of ER $\alpha$ , suggesting that ER $\alpha$  when expressed maintains the expression of the isoform including the exon, thus favouring the expression of the protein isoform containing the Kunitz protease inhibitor domain. The domain-containing protein isoform is involved in 5 different protein-protein interactions, including the interaction with APBB3, DAB2, APBB2, APBB1, APBA3, DAB1, and KLK2 (Huttlin et al. 2015, 2017; Louadi et al. 2021). The Kunitz protease inhibitor domain is located on the extracellular region of the protein and harbours intramolecular disulfide bounds and an O-linked glycosylation site and is involved in the interaction with KALK2 protein, which is completely removed as consequence of skipping the exon (Louadi et al. 2021). Clearly, further investigations are needed to unveil the downstream consequences of this ES event of the 7<sup>th</sup> alternative exon of *APLP2* gene upon hormone-independent ER $\alpha$  activity and whether it drives a specific cellular phenotype in MCF-7 cells. Similarly, a number of ASEs identified in our dataset were also reported to be recurrently misspliced in different cancers, including, among others, the ES events of the 4<sup>th</sup> (12 nucleotides length) and the 14<sup>th</sup> exons (21 nucleotides) of *USO1* gene which was found to be misspliced in 2 cancer types including LIHC, the ES event of the 12<sup>th</sup> exon (30 nucleotides) of *LMO7* gene which was recurrently misspliced in 6 tumors including BRCA, LUAD, PRAD, KIRC, HNSC, and COAD, the ES event of the 7<sup>th</sup> exon (24 nucleotides) of the *MYH14* gene which was reported to be highly expressed in BRCA, LUAD, and LUSC samples as compared to adjacent normal tissues, the ES event of the 6<sup>th</sup> (174 nucleotides

length) and the MXE events of the 8<sup>th</sup> (84 nucleotides length) and 9<sup>th</sup> exons (127 nucleotides length) of *PGAP2* gene which were reported to be misspliced in BRCA, COAD, LIHC, LUAD, and LUSC (Danan-Gotthold et al. 2015; Sebestyén et al. 2016).

Furthermore, by performing an RBP-binding motif enrichment analysis we identified putative SFs involved in the observed AS changes. Notably, this analysis revealed a high number of ASEs showing a predicted binding motif for the splicing factor CELF1, that underwent a splicing-level regulation in our dataset. For instance, the ES event involving the *LMO7* gene showed a strong enrichment of CELF1 binding motif (5'-TGTGTGT-3') downstream the exon, in line with the previously reported results showing that the inclusion of exon 10 (chr13:75809153-75809183) of *LMO7* gene, a marker gene of breast cancer, especially metastatic cancer cells (Perou et al. 2000) is positively regulated by *CELF1* (Xia et al. 2017). The integration of CELF1 CLIP-Seq dataset by (Le Tonquèze et al. 2016) further supported most of the predicted binding sites of CELF1 in our dataset.

Analysis of the association of the identified ASEs with clinical BC phenotypes revealed a number of events to be significantly associated with the overall and disease-free survival of ER $\alpha$ + patients. This includes 12 ES events involving 11 genes that were highly expressed in BCs and positively associated with a better overall survival time, and 7 events including 6 of type ES and one of type A3'SS, of which four events positively associated with better disease-free survival time while three negatively associated with it. Noteworthy, the 19 events were confirmed using an independent differential AS analysis tool PSIsigma (Supplementary Table 5). In particular, the ASEs involving exons E3, E4, and E6 in the *COL6A3* gene showed a good association with patient overall and disease-free survival time of ER $\alpha$ + patients. These same events were particularly investigated in other cancer types including pancreatic ductal carcinoma (PDA) (Arafat et al. 2011; C. Y. Kang et al. 2014), in colon cancer (Lian et al. 2020; W. Liu et al. 2018), in gastric cancer, bladder as well as in prostate cancer (Thorsen et al. 2008), indicating that the splicing change may play a role in multiple cancer types. The *COL6A3* gene encodes for the  $\alpha 3$  chain of the COL6 protein, and is formed by a short triple helical (TH) non collagenous domain of 200 repeating amino acids, 5 C-terminal domains (C1-C5), and 10 (N1-N10) tandem globular N-terminal modules similar to the von Willebrand factor type A (vWF-A) domain, each encoded by a single exon. The tumor-specific AS of E3, E4, and E6 results in the production of protein isoforms either including or lacking the N7, N9 or N10 domains. The expression of the three exons (E3, E4 and E6) was tumor-specific in PDA tissues (Arafat et al. 2011). In CRC, high expression of the gene is associated with poor prognosis, and its knockout significantly reduces cell proliferation and invasion, and induces apoptosis signaling in cancer cells (Lian et al. 2020; W. Liu et al. 2018). In particular, higher inclusion levels of the E5-E6 junction was specifically associated with better overall survival time of CRC patients. In line with these studies, we confirm the tumor-specific expression of these three exons in ER $\alpha$ + BCs supported by TCGASpliceSeq AS information analysis and we provide evidence that these high expressions are associated with a better overall survival time. Similarly, we identified a set of regulated ASEs upon ER $\alpha$ + silencing with higher expression levels associated with a better disease-free survival time, which could be used as a prognostic marker in BC.

Taken together, these data suggest the functional role of hormone-independent ER $\alpha$  activity in controlling post-transcriptional events by controlling the expression of RBP genes acting at different levels, including AS, intron retention, and polyadenylation (polyA) tail and polyA site selection. Such control exerted at this level is clearly to maintain cell proliferation and survival pathways and to prevent cellular transition of epithelial cells towards a more mesenchymal aggressive phenotype. Therefore, this data will further increase our understanding of the ER $\alpha$ -regulatory mechanisms exerted at the post transcriptional level and will help to identify putative target nodes and hubs acting downstream of ER $\alpha$  in BC.

## 2.2 The epithelial ESRP1 and ESRP2 factors control AS in ER $\alpha$ + BCs and their expression is tightly controlled by hormone-independent activity of ER $\alpha$

This chapter reports in four different sections the findings related to the gene-level and splicing-level effects of the combined silencing of *ESRP1* and *ESRP2* (*ESRP1/2*) genes in the ER $\alpha$ + MCF-7 BC cell line. The first section of the chapter, (Section 2.2.1), presents the findings related to the expression and copy number alterations of *ESRP1/2* in different BC subtypes and their association with specific clinical and molecular features of BC subjects. In addition, this section reports evidence that the expression of *ESRP1/2* in MCF-7 BC cells is strongly dependent on ER $\alpha$  expression in both cell lines and primary BC tissues. In line with this, a clear correlation between the expression of the three genes *ESR1*, *ESRP1* and *ESRP2* in BC tissues is reported. In section 2.2.2, the effects of the combined silencing of *ESRP1/2* genes at both gene- and splicing-level in MCF-7 cells are reported. Therefore, the results of differential isoform expression (dIE), differential isoform usage (dIU) and differential alternative splicing (AS) analyses are presented. In section 2.2.3, the expression of significant ASEs regulated by *ESRP1/2* are explored in BC tissues and their correlation with the expression of both genes is reported. The prognostic value of the identified ASEs in the context of the BC is presented. Finally, we foresee that *ESRP1/2* are downstream effectors of the ER $\alpha$  signaling pathway in ER $\alpha$ + BC. Thus, post-transcriptional events found common to ER $\alpha$  and *ESRP1/2* silencing experiments are discussed in the last section (section 2.2.4).

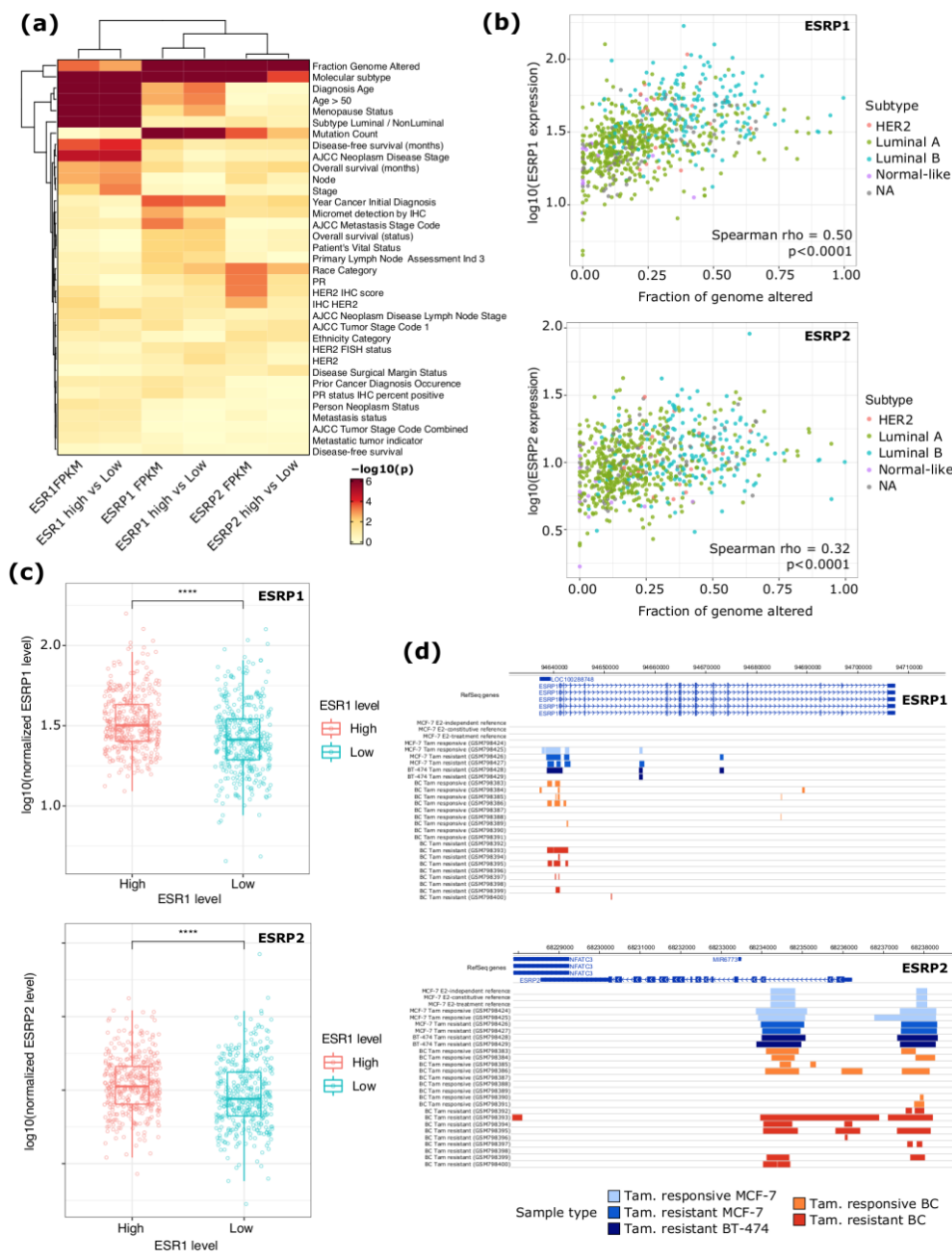
### 2.2.1 ESRP1 and ESRP2 expression is altered in ER $\alpha$ + BCs and is regulated by ER $\alpha$

The two Epithelial Splicing Regulatory Proteins 1 (ESRP1) and 2 (ESRP2) represent the core regulators of AS patterns in epithelial cells. They are crucial for maintaining the epithelial phenotype of cells by inhibiting the EMT process (Shapiro et al. 2011; Ishii et al. 2014; Hayakawa, Saitoh, and Miyazawa 2016). EMT induction by treating epithelial cells with TGF $\beta$  generally downregulates these factors (Shapiro et al. 2011; Horiguchi et al. 2012). The observed downregulation of these factors was reported to be mediated by EMT-related transcription factors such as SNAI1 and ZEB1, which repress *ESRP1* expression at the transcriptional level (Reinke, Xu, and Cheng 2012; Larsen et al. 2016), and ZEB2 which represses the expression of *ESRP2* (Horiguchi et al. 2012). Other regulatory mechanisms of both factors at the posttranscriptional level were recently discovered. For instance, the oncogenic AML4-ELK kinase activity repressed the expression of ESRP1 and ESRP2 in normal and lung cancer cells (Voena et al. 2016). The arkadia (ring finger protein 111, *RNF111*) is involved in the ubiquitination of ESRP2, thus it regulates its expression post-transcriptionally in clear-cell renal cancer cells (Mizutani et al. 2016). Conversely, the overexpression of these factors in mesenchymal cells induces a splicing pattern switch promoting the expression of epithelial-like transcripts, consequently resulting in a mesenchymal to epithelial transition (MET) (Warzecha et al. 2009). Furthermore, increasing evidence was recently reported that *ESRP1/2* genes are both overexpressed in many cancer types including ovarian (Jeong et al. 2017) and breast cancers (Sebestyén et al. 2016), where the overexpression was frequently associated with copy number variations affecting both genes. Recently, several studies have shown that the expression of both *ESRP1/2* genes is regulated by the activity of hormone receptor signaling pathways.

In prostate cancer, the androgen receptor directly regulates *ESRP2* expression promoting the onset and progression of the disease by controlling a splicing program in prostate cancer (Munkley et al. 2019). In BC, a recent work of Gökmen-Polar and colleagues demonstrated that *ESRP1* expression correlates with worse ER $\alpha$ + BC patient overall survival and endocrine treatment outcome, showing also *ESRP1* overexpression in two different endocrine-resistant BC cellular models (Gökmen-Polar et al. 2019).

In this study, the depletion of hormone-independent activity of ER $\alpha$  in MCF-7 cells strongly decreased the expression of both *ESRP1/2* genes (log<sub>2</sub>FC = -0.50 and -0.55 for *ESRP1* and *ESRP2*, respectively; adj-p <0.0001) (**Supplementary Table 1a** and **Supplementary Figure 6**), suggesting that ER $\alpha$  regulates the expression of these genes independently of estrogenic stimuli. Furthermore, to investigate the relationship between ER $\alpha$ , *ESRP1* and *ESRP2* gene expressions in BCs, *ESRP1/2* Copy Number Variation (CNV) status and RNA levels were analyzed in 773 ER $\alpha$ + BC samples from TCGA (Supplementary Table 12a). On one hand, this analysis revealed that the genomic region harboring the *ESRP1* locus was characterized by frequent copy number gain (60% of samples), while *ESRP2* locus was characterized by a frequent heterozygous deletion (62% of samples) in the analysed samples (**Supplementary Figure 7a**), in line with previously reported results (Sebestyén et al. 2016). The analysis of *ESRP1/2* gene expression levels revealed a set of tumor molecular and clinical features related to these genes (**Figure 2.3-1a** and **Supplementary Table 12b**). Specifically, the expression of both genes was correlated with the estimated fraction of genome altered (p<0.0001), particularly for *ESRP1* (Spearman rho = 0.50, p<0.0001) (**Figure 2.3-1b**) and was higher in luminal B tumors compared to the other molecular subtypes (p<0.0001) (Supplementary Figure 7b). The expression of both genes was also lower in patients presenting micrometastasis (p<0.01 and p<0.05 for *ESRP1* and *ESRP2*, respectively) while only the expression of *ESRP1* was related to the menopause status (p<0.05), and worse OS (p<0.05) (**Supplementary Figure 7c** and **Supplementary Table 12b**), as previously observed (Gökmen-Polar et al. 2019).

This analysis also highlighted a significantly higher expression of *ESRP1/2* genes in tumors characterized by the highest ER $\alpha$  mRNA level (p<0.001) (Figure 2.3-1c) as confirmed by a correlation analysis between the expression of these genes (Spearman rho = 0.30 and 0.20 for *ESRP1* and *ESRP2*, respectively; p<0.0001) (Supplementary Figure 7d). Analysis of public ER $\alpha$  ChIP-Seq experiments performed in BC cell lines and primary tissues confirmed ER $\alpha$  chromatin binding in proximity of both genes (Figure 2.3-1d). Specifically, considering *ESRP2* gene, ER $\alpha$  binding was observed at gene promoter and second intron by integrating in the analysis public ChIP-Seq data of ER $\alpha$  chromatin binding in estrogens-treated or estrogens-deprived MCF-7 cells (Ferrero et al. 2017). Moreover, ER $\alpha$  binding sites were also detected in tamoxifen-responsive and resistant cell lines (MCF-7, BT474) and primary BCs (Ross-Innes et al. 2012). Conversely, ER $\alpha$  binding at *ESRP1* gene was observed at gene promoter but only in a subset of drug-responsive MCF-7 and primary BC tissues. These data suggest that *ESRP1/2* genes are overexpressed in BCs, correlate with the expression of ER $\alpha$  gene in ER $\alpha$ + tumors, and significantly associate with a set of specific molecular and clinical features of these tumors. Importantly, their expression in the ER $\alpha$ + MCF-7 BC cell lines is strongly dependent on ER $\alpha$  gene expression and occurs independently of estrogenic stimuli as shown by RNA-Seq and ChIP-seq data analysis.

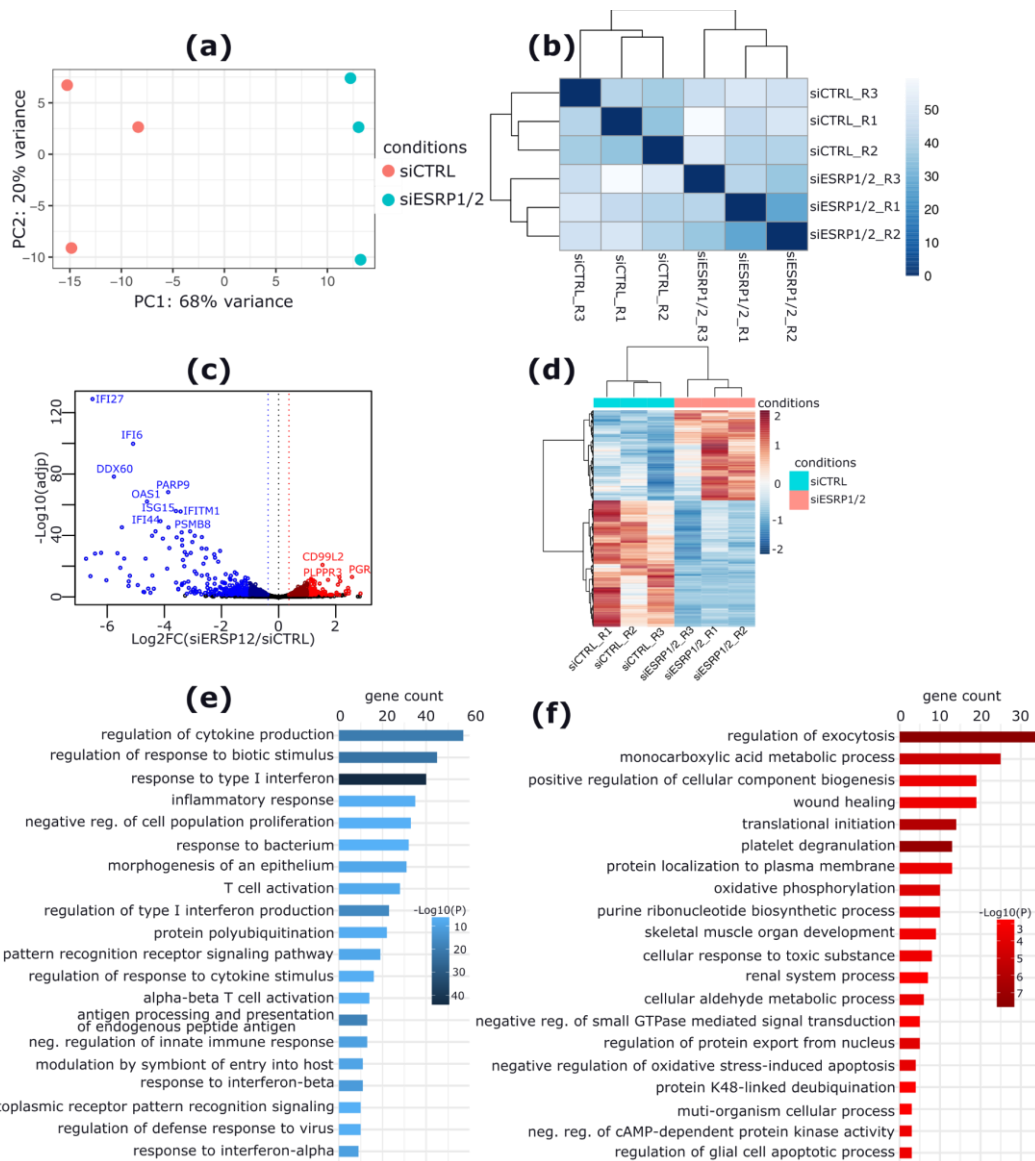


**Figure 2.2-1:** (a) Heatmap reporting the  $p$ -value obtained by the analysis of TCGA BC patients clinical data with respect to ESR1, ESRP1, and ESRP2 expression. The analysis was performed by considering continuous expression levels (first, third, and fifth columns) or by dividing subjects into high and low groups based on the gene expression. PR, Progesterone Receptor; IHC, Immuno-histochemistry; HER2, human growth factor receptor-2. (b) Scatter plots reporting the expression level of ESRP1 (top) and ESRP2 (bottom) in relation with the degree of genome alteration of TCGA BC samples. Samples are color-coded based on their associated BC molecular subtype. (c) Box plots reporting the ESRP1 (top) and ESRP2 (bottom) expression levels in TCGA BC patients divided based on the ESR1 expression levels.  $p$ -value by Wilcoxon Rank-Sum test. \*\*\*\*,  $p$ -value  $< 0.0001$ . (d) Genome browser view of ESRP1 (top) and ESRP2 (bottom) loci with indication of the ER $\alpha$  binding sites detected by ChIP-Seq BC cell lines and primary BC tissues. Data of Tamoxifen (Tam) sensitive cell lines and tissues are color-coded in light-blue and orange, respectively. Data obtained in Tam-resistant cell lines/tissues were color-coded in blue (MCF-7), dark-blue (BT-474), and red (primary BC tissues).

### 2.2.2 ESRP1 and ESRP2 regulate metabolism, cell proliferation, in addition to EMT pathways in MCF-7 cells

Since their expression was tightly dependent on the expression of ER $\alpha$ , the activity of ESRP1 and ESRP2 on the ER $\alpha$ + BC transcriptome was explored by performing a triplicate paired-end RNA-seq, consisting of transfecting MCF-7 cells with siRNAs specific to *ESRP1* and *ESRP2* mRNAs (combined) (siESRP1/2) (Supplementary Figure 8) or control siRNA. The resulting RNA-Seq dataset (**Figure 2.2-2a,b**) was analysed for both gene-level and splicing-level changes induced by siESRP1/2. Notably, dGE analysis revealed 915 DE genes of which, 479 downregulated and 436 upregulated (**Figure 2.2-2c,d** and **Supplementary Table 13a**). Functional enrichment analysis showed that downregulated genes were mainly enriched in terms related to type I interferon signalling pathways, regulation of cell adhesion, morphogenesis of an epithelium, and regulation of cell population proliferation. Conversely, upregulated genes were enriched in terms such as regulation of exocytosis, metabolism, oxidative phosphorylation, wound healing, regulation of vesicle mediated transport, extracellular matrix organization, and response to oxidative stress (**Figure 2.2-2e,f** and **Supplementary Table 13b,c**). Furthermore, a hallmark gene sets enrichment analysis revealed terms differentially enriched for upregulated and downregulated genes, including EMT, apical junction, KRAS signaling, androgen response, and early estrogen response hallmarks, related to downregulated genes. Conversely, late estrogen response (top significant hallmark), adipogenesis, oxidative phosphorylation, glycolysis, hypoxia, UV response\_up, DNA repair, and reactive oxygen species enriched for upregulated genes (**Supplementary Figure 9** and **Supplementary Table 13d,e**). These results suggest that the expression of ESRP1 and ESRP2 splicing factors in the ER $\alpha$ + MCF-7 BC cell line tends to regulate the expression and activity of genes related to EMT, intracellular signaling, metabolism and, most importantly, to estrogen response.





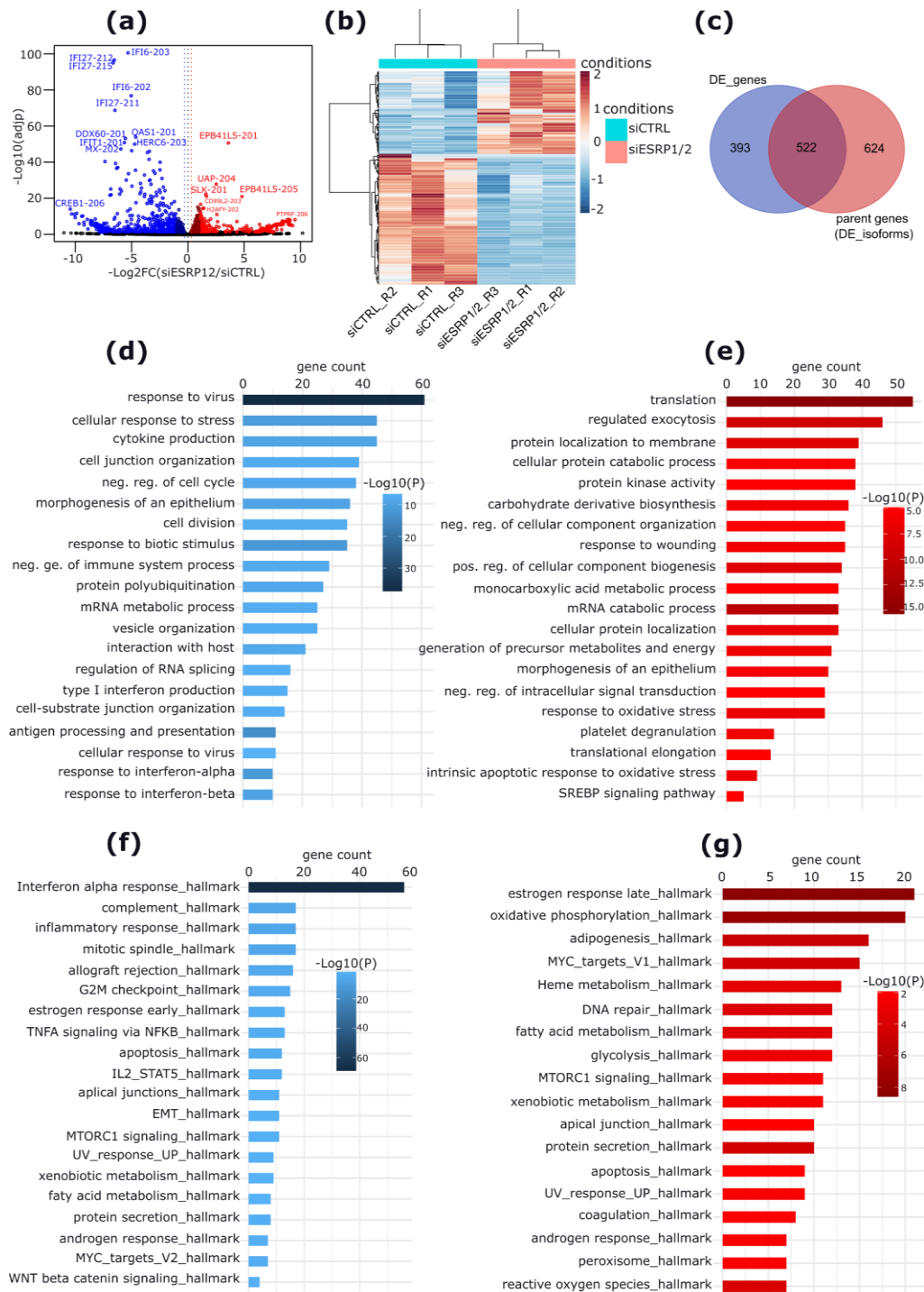
**Figure 2.2-2:** General overview of the gene-level changes induced by the combined silencing of *ESRP1* and *ESRP2* genes with specific siRNAs. (a) PCA plot showing the separation of the analyzed RNA-Seq replicates. (b) dissimilarity matrix heat map showing the distance between the RNA-seq experiment replicates. (c) Volcano plot reporting the  $\text{Log}_2\text{FC}$  (x-axis) and the significance ( $\text{adjp}$ -value) (y-axis) of DE genes responding to siESRP1/2. Labels are accordingly provided for top DE genes. (d) Heatmap plot reporting the expression levels in (TPM units) of the top 500 changing genes. Color intensities representing the z-scores are proportional to the expression level of genes, with red and blue colors indicating upregulated and downregulated genes, respectively. (e-f) bar plots reporting the top 20 enriched GO clusters for downregulated (blue) and upregulated (red) genes, respectively. The bar lengths are proportional to the number of genes per each GO cluster, and color intensities are proportional to the significance ( $\text{log}_{10}$  transformed p-value) of the enrichment.

Similarly, to identify an isoform-specific response upon the combined silencing of *ESRP1* and *ESRP2* genes, a DE analysis at the level of isoforms (dIE) was performed. This analysis revealed the differential expression of 1461 isoforms, including 756 and 705 downregulated and upregulated, respectively (**Figure 2.2-3a,b** and **Supplementary Table 14a**). Notably, the overlap with dGE analysis revealed a total of 522 overlapping genes which were considered as DE at both gene- and isoform-levels (**Figure 2.2-3c**). However, a set of DE isoforms

originated from 624 genes that were considered as not DE at the gene-level (**Figure 2.2-3c**), suggesting that isoform-level analysis captures more regulatory events occurring either at the transcriptional or post transcriptional levels and improves the detection of these hidden events. Furthermore, this also implies that an isoform-level regulation event does not potentially drive an overall gene expression change, further supporting the outperformance of an isoform-level over a gene-level analysis.

In order to identify potentially affected by DE isoforms, genes harboring a DE isoform were subjected to an GO enrichment analysis. In addition to confirming the terms enriched for genes regulated at gene-level, the isoform-level analysis revealed the enrichment of other terms that were not enriched by the dGE analysis. For instance, in the case of downregulated isoforms, we observed significant enrichment of GO terms related to regulation of cell cycle and cell division, regulation of mRNA metabolic process, regulation of RNA splicing, cell-substrate junction organization, and vesicle organization. Similarly, novel terms were enriched for upregulated isoforms, including mRNA catabolic process, and translation which appeared as the top significant term (**Figure 2.2-3d,e** and **Supplementary Table 14b,c**). In addition, hallmark gene sets enrichment analysis at isoform level confirmed the terms enriched at the gene-level and revealed further enriched terms. In particular, G2M checkpoint, MYC\_targets\_V2, and protein secretion terms were exclusively enriched for downregulated isoforms (**Figure 2.2-3f,g** and **Supplementary Table 14d,e**).

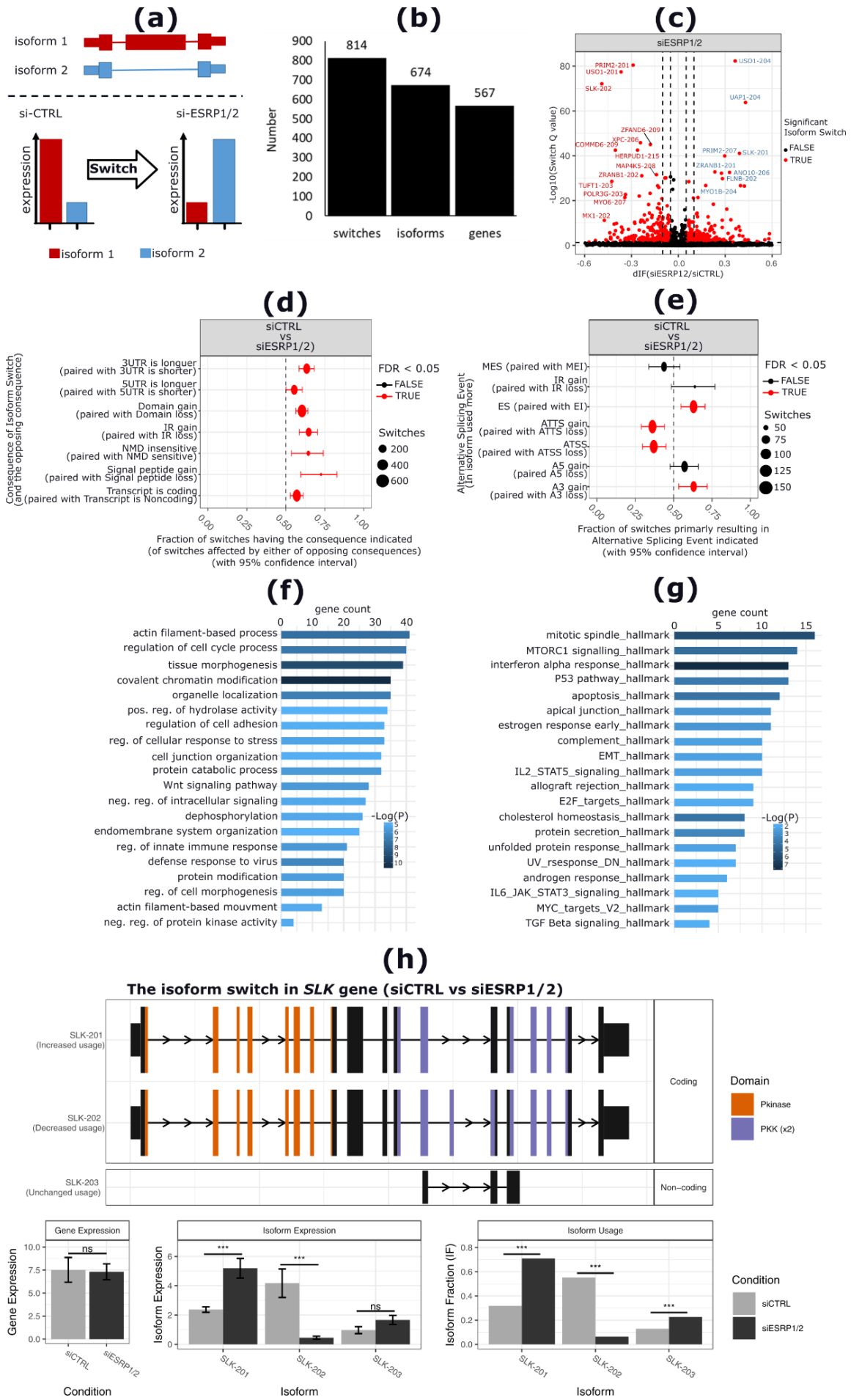
On the other hand, the comparison of gene-level and isoform-level analyses revealed a number of gene isoforms pairs that responded differently to the silencing of *ESRP1/2* genes, as compared to their parent genes. For instance, 4 isoform pairs of 4 genes (*MYO1B*, *ANP32E*, *DNMT3B*, and *CYB561A3*) were downregulated while their parent genes were upregulated. Similarly, 5 isoform pairs of 5 genes (*CD99L2*, *AP1B1*, *MYL6*, *OTUB1*, and *FLNB*) were downregulated while their parent genes were upregulated (**Supplementary Table 14a**).



**Figure 2.2-3:** Overview of isoform-level changes induced by the combined silencing of ESRP1 and ESRP2 genes with specific siRNAs in MCF-7 cells. (a) Volcano plot representing the log<sub>2</sub>FC (x-axis) and significance (adjp, y-axis) of DE isoforms. The name of the top DE isoforms is reported. (b) Heat map plot reporting the expression levels (in TPM units) of the top 500 changing isoforms. Color intensities representing the z-score are proportional to the expression level of genes, with red and blue colors indicating upregulated and downregulated genes, respectively. (c) Venn diagram reporting the overlap between gene-level and isoform-level analysis. (d-e) Bar plots reporting the enriched biological process for downregulated (d) and upregulated isoforms (e), respectively. (f-g) Bar plots reporting the enriched hallmark gene sets related to downregulated (f) and upregulated isoforms (g), respectively.

In order to identify those events where isoforms responded differentially to the combined silencing of *ESRP1/2* genes and to analyse their downstream effects with ease, a differential isoform usage (dIU) analysis reporting isoforms switching events was performed using IsoformSwitchAnalyzeR (Vitting-Seerup and Sandelin 2019) (**Figure 2.2-4a**). This analysis revealed a set of isoform switching events not captured by dGE and dIE analyses. Notably, 674 isoforms (of 567 genes) were involved in 814 significant switches, resulting in the induction of 338 isoforms and repression of 336 isoforms upon *ESRP1/2* silencing (**Figure 2.2-4b,c** and **Supplementary Table 15a**). The 567 switching genes were enriched in terms such as interferon-alpha response, mitotic spindle, mTORC1 signaling, apoptosis, cholesterol homeostasis, p53 pathway, early estrogen response, MYC targets, TGFB signaling and EMT hallmarks (**Figure 2.2-4f,g** and **Supplementary Table 15b**).

The identified isoform switching events resulted in significant downstream consequences distinguishing induced from repressed isoforms. Notably, switching isoform pairs differ by 3'UTR and 5'UTR lengths, by type and number of protein domains, by IR events, by sensitivity to NMD process, by presence/absence of signal peptide sequence, and by their coding potential (**Figure 2.2-4d** and **Supplementary Table 15c**). In particular, the combined silencing of *ESRP1/2* genes resulted in the expression of isoforms characterized by longer 3'UTRs (169 longer vs 99 shorter, proportion q-value = 2.5E-5), longer 5'UTRs (134 longer vs 104 shorter, proportion q-value = 0.05), having more protein domains (244 protein domain gains vs 161 loss, proportion q-value = 4.6E-5), more IR events (94 gain vs 60 loss, and 4 IR switches, proportion q-value = 7.8E-3), less sensitive to NMD process (43 NMD insensitive vs 23 NMD sensitive, proportion q-value = 0.02), with more coding potential (176 coding vs 162 non coding, proportion q-value = 0.4) and more signal peptides (28 gains vs 13 loss, proportion q-value = 0.02) (**Figure 2.2-4d** and **Supplementary Table 15c**). Furthermore, this analysis revealed specific ASE types enriched in the silencing as compared to the control condition. Notably, switching isoform pairs involved more ES events (249 ES vs 197 EI; proportion q-value = 0.02), more IR gains than losses (93 IR gained vs 61 IR lost; proportion q-value = 0.02), more A3 gains than losses (171 A3 gained vs 137 A3 lost; proportion q-value = 0.05), less ATTS (usage of distal 3'exon) (153 gains vs 241 losses, proportion q-value = 9.3E-5), and more ATSS (usage of proximal promoter) losses than gains (174 gained vs 238 lost; proportion q-value = 0.0007) (**Figure 2.2-4e** and **Supplementary Table 15d**).

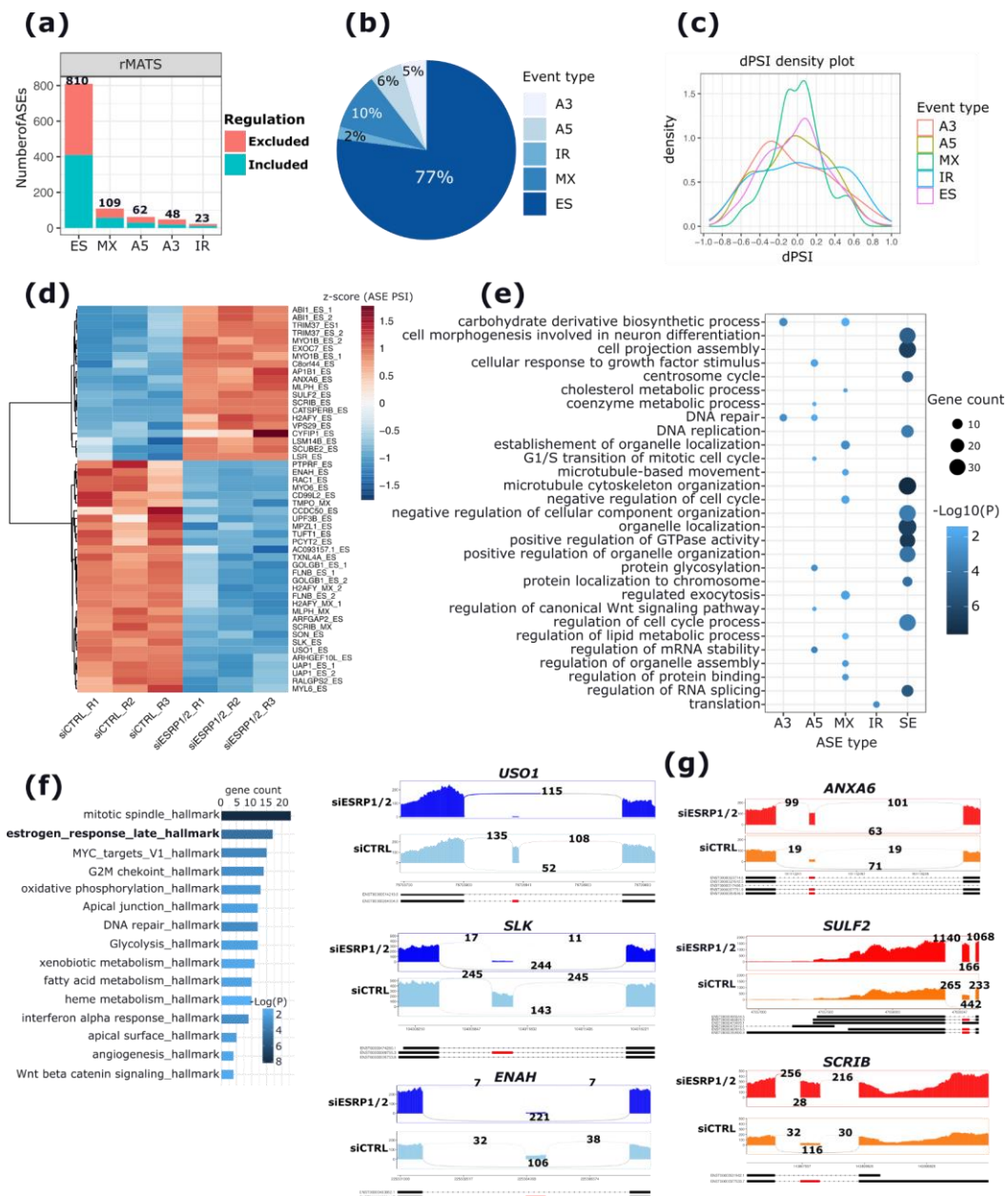


**Figure 2.2-4:** Overview of the differential isoform usage (dIU) events induced by the combined silencing of *ESRP1* and *ESRP2* genes. (a) The principle behind the isoform switching analysis. Relative abundances of the example isoforms 1 and 2 are compared among *siCTRL* and *siESRP1/2* conditions. (b) Bar plots reporting the number of significant switching events, and the isoforms and genes involved. (c) Volcano plot reporting the differential isoform fraction/usage (dIF) on x-axis and their relative significance (*Q* value) on y-axis. Top significantly switching isoforms are labeled accordingly. For example, *USO1* gene has two switching isoform pairs, *USO1-201* repressed and *USO1-204* induced, by the combined *ESRP1/2* silencing. (d) Functional consequences of switching events (y-axis) and the proportion of switches (x-axis) having either of the opposing consequences. (e) Alternative splicing events (ASEs) (y-axis) predicted by *IsoformSwitchAnalyzeR* to be enriched and the proportions (x-axis) of switches resulting from either opposing ASEs. *MES*: multiple exon skipping, *MEI*: multiple exon inclusion, *ES*: exon skipping, *EI*: exon inclusion, *A5*: alternative 5' splice site; *A3*: alternative 3' splice site; *ATTS*: alternative transcription termination site; *ATTS*: alternative transcription start site. (f-g) Bar plots reporting enriched biological processes (f) and gene set hallmarks (g) related to switching genes, respectively. (h) Isoform switch plot reporting the example of *SLK* gene having two switching isoforms *SLK-201* (induced), and *SLK-202* (repressed). Protein domains resulting from each exonic region are shown in the top panel. The expression of the gene and its isoforms in both conditions are reported in the histograms in the bottom panel. (ns, not significant,  $*** < \text{adj-}p < 0.00001$ ).

Furthermore, this analysis confirmed the observed changes reported by dIE analysis. Notably, the most significant switching isoform pairs in dIU analysis were also reported by dIE analysis to respond in opposite directions to the combined silencing of *ESRP1/2* genes. This includes switching events in the *USO1* gene isoforms, where the isoform *USO1-201* is repressed (dIF = -0.36, adjp = 3.79E-78) while the isoform *USO1-204* is induced (dIF = 0.36, adjp = 5.04E-83); in the *PRIM2* gene isoform pairs, where *PRIM2-201* is repressed (dIF = -0.29, adjp = 3.53E-81) and *PRIM2-207* is induced (dIF = 0.29, adjp = 1.19E-41.); in the *SLK* gene isoform pairs, where the isoform *SLK-202* is repressed (dIF = -0.49, adjp = 7.60E-73), while *SLK-201* isoform is induced (dIF = 0.39, adjp = 8.45E-42); in the *UAP1* gene isoform pairs, where the isoform *UAP1-204* is induced (dIF = 0.43, adjp = 1.71E-64) while *UAP1-201* isoform is repressed (dIF = -0.28, adjp = 4.66E-5); and in the *FLNB* gene where the isoform *FLNB-202* is induced (dIF = 0.27, adjp = 6.56E-33), while the isoform *FLNB-201* is repressed (dIF = -0.10, adjp = 8.04E-13) (**Figure 2.2-4c** and **Supplementary Table 15a**). Importantly, all these switching isoform pairs involved an ES/EI event targeting an exon that in several cases encodes for a protein domain. For instance, as shown in **Figure 2.2-4h**, the switching isoform pairs of *SLK* gene differ by the inclusion/exclusion of the 13th exon, and a consequent effect of this switch is that the isoform including the exon (*SLK-202*) is repressed while the one lacking the exon (*SLK-201*) is being induced by the combined silencing of *ESRP1/2* genes. As a result of this switch, the encoded PKK protein domain is lost in the silencing condition and the number of PKK protein domains is decreased (**Figure 2.2-4h**). However, other similar events but targeting exons not encoding for protein domains were also reported. For instance, the switching isoform pairs of *FLNB* gene differ by the inclusion/exclusion of the 16th and 30th exons, both are repressed by the silencing of *ESRP1/2* genes. While the 16th exon encodes for a protein domain (filamin), the 30th exon is not annotated to encode for a protein domain (**Supplementary Figure 10**). Noteworthy, the skipping of this same exon, (30th exon), was reported to be sufficient for EMT induction in MCF-7 and MDA-

MB-231 BC cell lines, and that this ES event resulted in releasing the transcription factor FOXC1 in BC cell lines (J. Li et al. 2018).

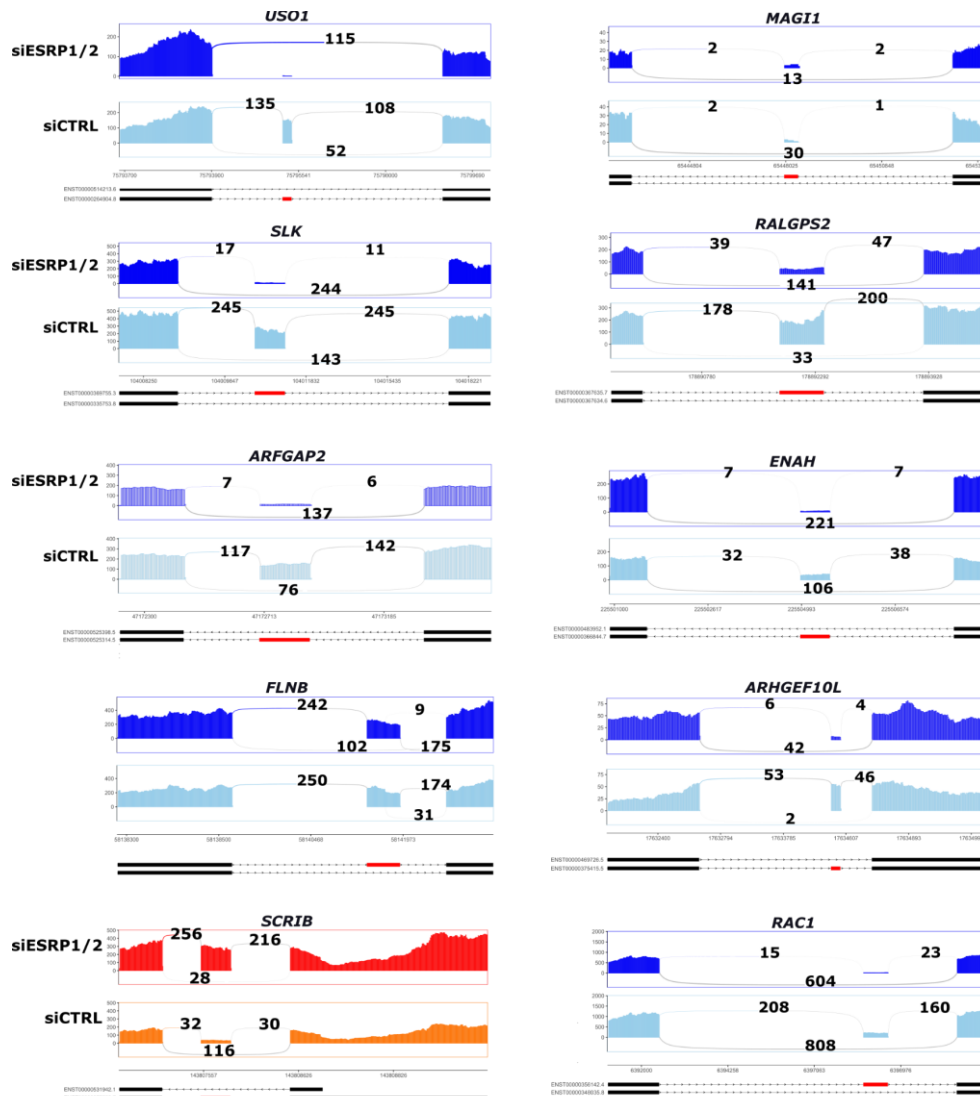
Furthermore, to further support our findings reported by the dIU analysis and their putative downstream effects at protein isoform level, local ASEs were explored in our siESRP1/2 RNA-Seq dataset by performing a differential alternative splicing (AS) analysis using the event-based approach, rMATS (Shen et al. 2014). Indeed, this analysis revealed that ESRP1/2 knock-down induces changes in the splicing pattern of 788 genes. A total of 1,052 ASEs were significantly ( $p < 0.05$ ) differentially regulated upon ESRP1/2 silencing, including 810 (77%) ES/EI in 650 genes, 109 (10%) MX in 91 genes, 62 (6%) A5 in 58 genes, followed by 48 (5%) A3 in 45 genes, and 23 (2%) IR events involving 21 genes (**Figure 2.2-5a-c** and **Supplementary Table 16a-e**). The top 50 significant ASEs in this analysis correspond to 45 ES events and 5 MX events including 27 excluded and 23 included exons by ESRP1/2 silencing (**Figure 2.2-5d** and **Supplementary Table 16a-e**). Enrichment analysis on the ASEs harbouring-genes showed an overrepresentation of terms such as microtubule cytoskeleton organization, organelle localization, cell projection assembly, positive regulation of GTPase activity, in addition to regulation of cell cycle process (**Figure 2.2-5e** and **Supplementary Table 16f**). These results were enriched in genes harboring ES/EI events, while genes harbouring MX events were overrepresented in terms related to metabolic processes, including carbohydrate derivative biosynthetic process, cholesterol biosynthetic process, as well as EMT process (**Figure 2.2-5e** and **Supplementary Table 16f**). Moreover, a hallmark gene sets enrichment analysis of significantly spliced genes reported enrichment of terms related to mitotic spindle, estrogen response, MYC targets, G2M checkpoint, DNA repair, angiogenesis, interferon-alpha response, cell junction organization, in addition to metabolism-related terms such as oxidative phosphorylation, glycolysis, heme metabolism, and fatty acid metabolism (**Figure 2.2-5f** and **Supplementary Table 16g**). Selected examples of the top ES/EI events induced by the combined silencing of *ESRP1* and *ESRP2* genes are shown in **Figure 2.2-5g**.



**Figure 2.2-5:** Overview of differential AS changes induced by the combined silencing of ESRP1 and ESRP2 genes as reported using rMATS (Shen et al. 2014). (a) Stacked bar plots reporting the number of ASEs classified by ASE type and type of regulation (e.g. inclusion/exclusion). (b) Pie-chart reporting the proportions of each ASE type reported to be significantly differentially spliced by rMATS. (c) Density plot reporting the distribution of dPSI values for each ASE type. Only dPSI of significant ASEs are reported. (d) Heat map plot representing the PSI value of the top 50 ASEs differentially spliced upon ESRP1 and ESRP2 silencing. (e) Dot plot reporting the enriched biological processes for each ASE type. Number of genes per ASE type is proportional to dot size. Color intensities represent the significance (p-value) of the enrichment. (f) Bar plot representing the enriched gene set hallmark terms related to all significantly spliced genes. Bar length is proportional to the number of genes overlapping each enriched term. Color intensities represent the significance (p-value) of the enrichment. (h) Sashimi plots showing selected examples of ES events induced or repressed by the combined silencing of ESRP1 and ESRP2 genes. First panel represents skipping events (blue color). Second panel reports examples of inclusion events (red color). Normalized read counts supporting either inclusion or exclusion of regulated exons are reported near to corresponding junctions. Isoforms representing the inclusion and exclusion forms of the events are reported, with exons involved in the events are highlighted in red.



Given that ESRP1 and ESRP2 are two factors linked to epithelial phenotype and play a crucial role during the EMT process, our list of significant ASEs was explored to identify EMT-related splicing patterns previously reported in literature. Indeed, this analysis confirmed in our dataset a number of ASEs affecting the splicing patterns of ESRP-regulated genes involved in EMT as previously reported (Yueqin Yang et al. 2016; Shapiro et al. 2011; Warzecha et al. 2010, 2009), including *USO1* (ES), *SLK* (ES), *ARGFAP2* (ES), *FLNB* (ES and A5'SS), *SCRIB* (ES), *MAGI1* (ES), *RALGPS2* (ES), *ENAH* (ES), *ARHGEF10L* (ES), and *RAC1* (ES) (**Figure 2.2-6** and **Supplementary Table 16a**).



**Figure 2.2-6:** Sashimi plots reporting the AS changes involving the top10 EMT-related splicing patterns. For each ASE, inclusion and exclusion levels in the control and ESR1/2 silencing conditions are given. Numbers above each junction indicate the normalized read counts supporting either inclusion or exclusion of the ASE. ES events are indicated in blue color, while EI events in the *SCRIB* gene are indicated in red color. Selected genes are based on literature in (Yueqin Yang et al. 2016; Shapiro et al. 2011; Warzecha et al. 2010, 2009).

Furthermore, besides EMT-related ASEs, *ESRP1/2* combined silencing induced significant switches in the AS patterns of genes involved in cell signaling for motility and proliferation, including the Rac family small GTPase 1 (*RAC1*), *KRAS* proto-oncogene, GTPase (*KRAS*) and Annexin A6 (*ANXA6*) genes. We observed

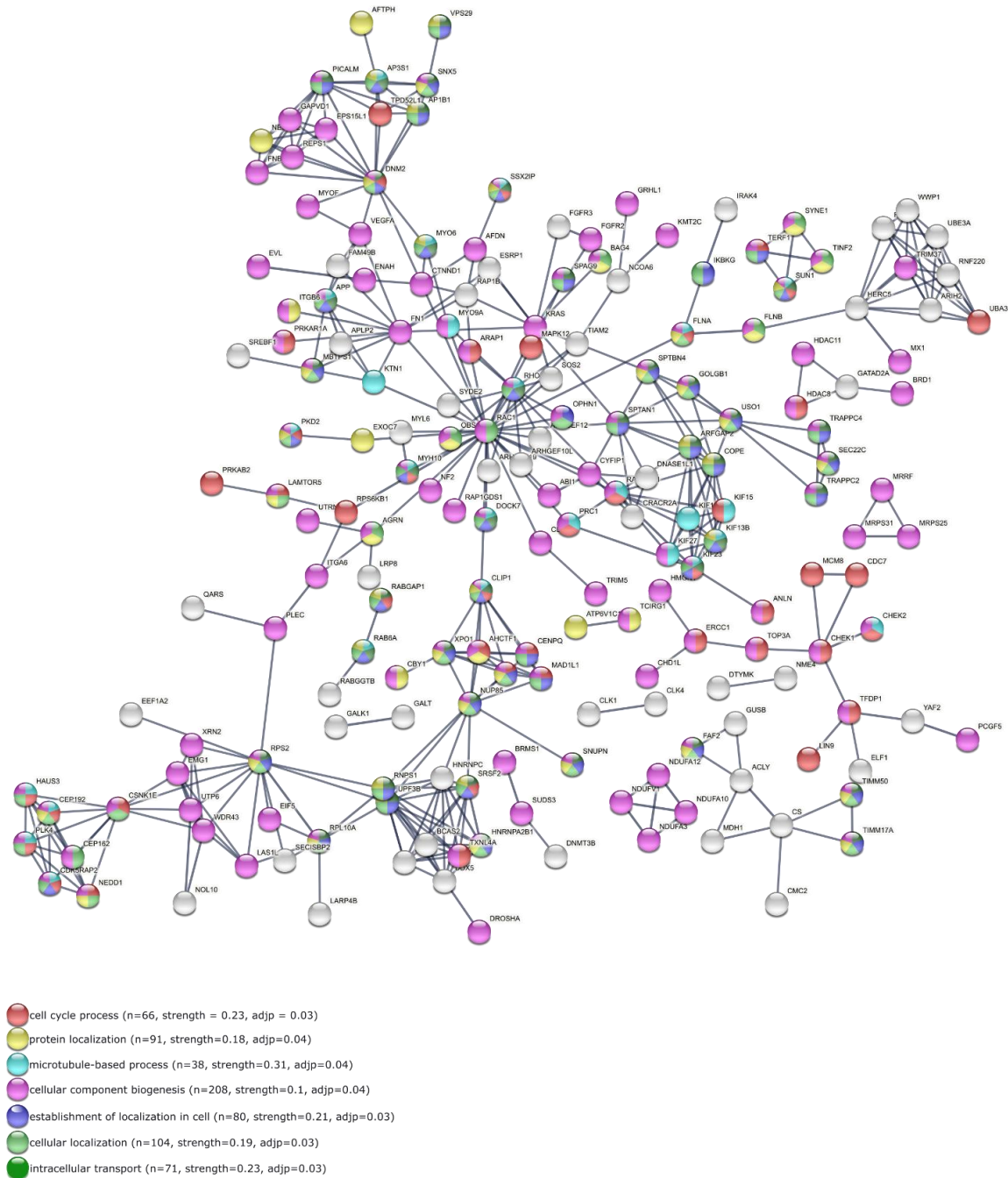
from dIU and differential AS analysis a significant switch in *RAC1* gene isoforms, resulting in the significant decrease in the relative abundance of Rac 1 b (Rac 1b) isoform and a significant increase in the expression of the canonical Rac 1 isoform, lacking exon 4 (also known as exon 3b) (**Supplementary Figure 11**). The Rac 1b isoform which is more expressed in cells treated with siRNA control includes exon 4 and has been described to be constitutively active and highly expressed in different cancers including breast, thyroid, colorectal and lung tumors (Schnelzer et al. 2000; Gonçalves et al. 2014; C. Zhou et al. 2013). Our results are in line with Eiden and coworkers showing a higher expression of Rac 1b isoform and low expression of the canonical RAC1 in epithelial BCs including MCF-7 and patients biopsies, and that Rac 1b/RAC1(canonical) ratio is inverted in undifferentiated BC tissues and mesenchymal cell lines (Eiden and Ungefroren 2021). Moreover, the top significant ASE induced by *ESRP1/2* combined silencing is an exon inclusion (EI) event of exon 21 of *ANXA6* gene (dPSI = 0.44, adj-p= 0) (**Supplementary Table 16a**), in line with dIU analysis showing a significant increase in the *ANXA6* gene canonical isoform that includes exon 21 (**Supplementary Figure 12** and **Supplementary Table 15a**). Noteworthy, changes in the aforementioned ASEs were confirmed using three independent differential AS analysis tools, including rMATS (Shen et al. 2014), PSISigma (K.-T. Lin and Krainer 2019), and Whippet (Sterne-Weiler et al. 2018) algorithms, all of which indicated the exon inclusion event of the *ANXA6* exon as among the top significant events in our dataset (**Supplementary Table 16e-f**). The *ANXA6* gene has been characterized for its important and diverse functional roles in a wide range of cancers including different BC subtypes (Qiet al. 2015). In addition, *ANXA6* was reported as a potential marker for the detection of BC and prediction of survival (Koumangoye et al. 2013).

### 2.2.2.1 *ESRP1/2* knockdown-induced ASEs cooperatively alter functional protein-protein interaction networks in MCF-7 BC cells

To understand downstream consequences of the AS changes induced by *ESRP1/2* silencing, both a protein-level and exon-level analyses were performed. Specifically, genes harbouring an ES/EI event were subjected to protein-protein interaction (PPI) enrichment analysis using the STRING database (Szklarczyk et al. 2019). Next, exons involved in each ES/EI event were explored using the results from the dIU analysis to identify whether an alternatively spliced exonic region encodes for a protein domain. The identified protein domains were then explored using the DIGGER database (Louadi et al. 2021), to identify their interacting partners (e.g. domain-domain and protein-protein interactions). Interestingly, functionally related PPI networks were enriched for this group of ES/EI events (**Figure 2.3-7** and **Supplementary Table 17a**). In particular, a high confidence, experimentally supported PPI network (selecting highest confidence 0.9, and setting active interaction sources to experiments, curated databases) was significantly enriched in biological processes, including cell cycle, microtubule-based process, intracellular transport, protein localization, and cellular localization processes (number of edges=399, expected number of edges=306, PPI enrichment p-value=2.01E-7). The enriched PPI network includes 66 nodes (e.g. proteins) involved in cell cycle, 91 in protein localization, 80 in cell localization, 71 in intracellular transport, and 201 in cellular component biogenesis (**Figure 2.3-7** and **Supplementary Table 17a**). Moreover, a physical PPI network, where interacting proteins form

---

functional protein complexes, was enriched (Supplementary Figure 13 and Supplementary Table 17b), suggesting that AS changes observed upon the combined silencing of *ESRP1/2* genes may target functional protein complexes by splicing in/out exonic regions encoding protein domains. On the other hand, analysis of the exonic regions involved in ES/EI events using DIGGER (Louadi et al. 2021) revealed that ASEs have a combinatorial effect at the protein level, by either suppressing or inducing a number of functionally related protein domains. For instance, 17 ES events involving the Rac family small GTPase (*RAC1*) gene resulted in the complete removal of the protein domains involved in the mutual interactions between RAC1 protein and its partners. In particular, the skipped exon (chr7:6398661-6398718) of *RAC1* encodes for the Ras family protein domain (*Ras*, PF00071), involved in 169 PPIs, including its interaction with fibronectin (FN1). On the other hand, the skipped exon (chr2:215380810-215381080) of *FNI* gene encodes for the protein domain Fibronectin type III domain (fn3, PF00041), mediating 89 protein interactions, of which 69 interactions are completely removed upon the skipping of the exon, including the interaction with RAC1 (**Figure 2.3-7 and Supplementary Table 17c**), 6 interactions are partially affected (20% -75% suppressed) and 3 are unaffected by the skipping of that exon. Altogether, these data suggest that, in addition to measuring local AS changes, one has to consider a protein-level analysis to decipher downstream targets of the switching AS patterns, and not only limiting the attention to common AS events (e.g. differentially spliced genes). This is especially relevant when comparing different experimental contexts, since different ASEs targeting different genes could result in the alteration of the same PPI network.

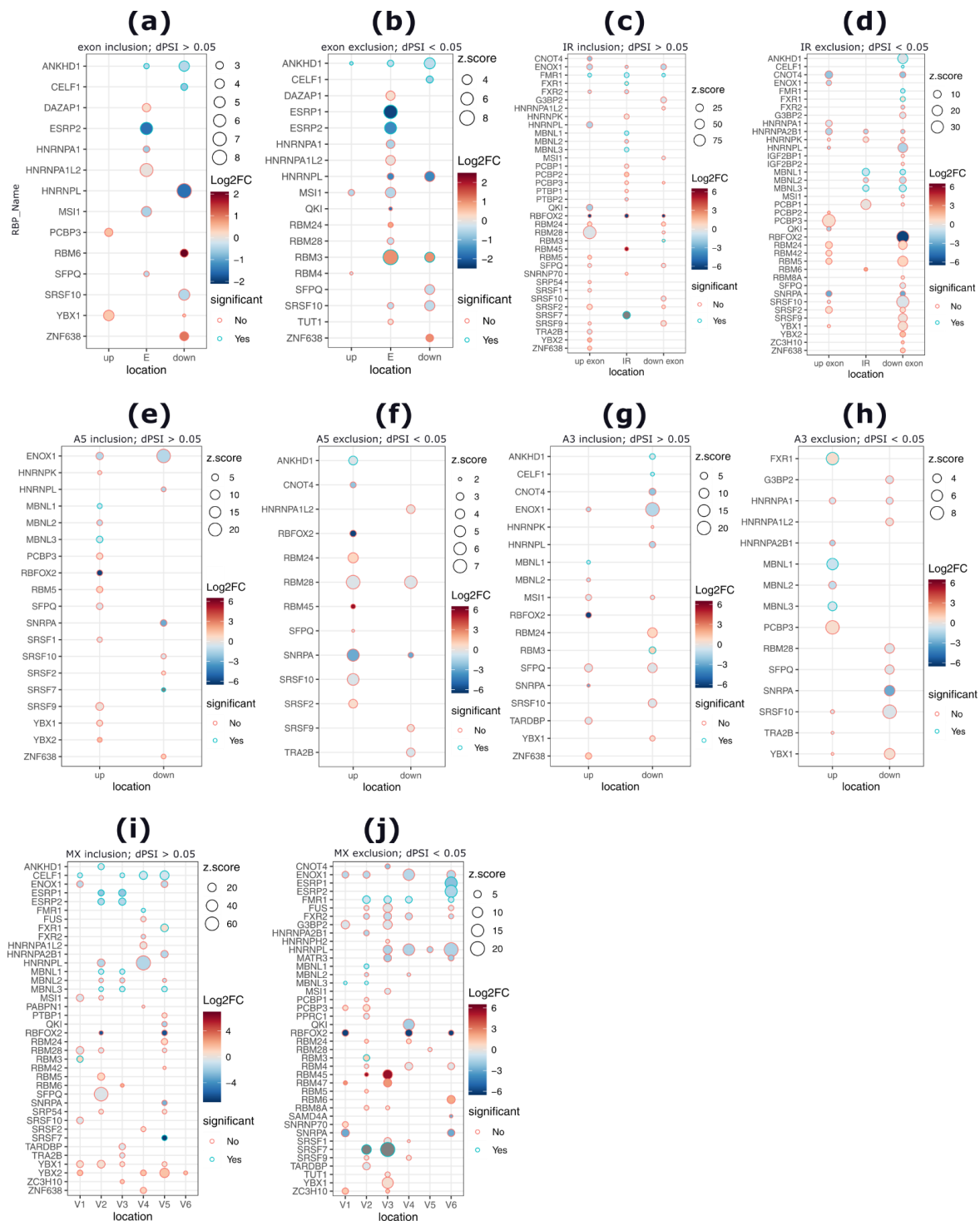


**Figure 2.2-7:** Protein-protein interactions (PPIs) network enriched for genes harbouring an ES event induced by the combined silencing of *ESRP1/2* genes. Nodes of the network represent the potential proteins encoded by spliced genes and the edges are experimentally supported interactions in STRING database (Szklarczyk et al. 2019). Colors indicate enriched biological processes related to ES harbouring genes. A node with different colors indicates that the node is involved in more than one biological process.

### 2.2.2.2 ESRP1 and ESRP2 binding motifs are enriched in the ASEs induced by the combined knock down of both genes

To determine whether ESRP1 and ESRP2 splicing factors are directly involved in the control of the observed AS changes, the ASEs were further analysed for the presence of RBP-binding motifs. Therefore, an RBP-binding motif enrichment analysis was performed for 105 SFs (122 binding motifs) including ESRP1 and

ESRP2. First, ASEs were classified based on event type, and direction of regulation (included/excluded). Second, the sequences of the genomic regions involved in the ASEs and 200 nucleotides on both sides were scanned for RBP binding motifs, with the exception of MX events which were extended by 100 nucleotides on both sides. As expected, ESRP1 and ESRP2 binding motifs were predicted in 694 significant ASEs in 457 genes, consisting of 540 (51.33%) ES/EI events (274 skipped, 266 included) in 364 genes, 37 (3.52%) A5 (19 skipped, 18 included) in 34 genes, 28 (2.66%) A3 (17 skipped, 11 included) in 26 genes, 70 (6.65%) MX events (34 skipped, 36 included) in 57 genes, 18 (1.71%) RI events (8 skipped, 10 included) in 16 genes (**Supplementary Table 18a,b**). ESRP1/2 binding motifs were predicted at both exons and their flanking regions involved in the ASE. For example, among the 274 ES events, 85 binding motifs were exonic, 99 upstream and 122 were predicted downstream of spliced exons. Similarly, among the 266 EI events, 87 predicted motifs were exonic, 118 upstream and 114 downstream of spliced exons. Furthermore, an RBP-binding motif enrichment analysis performed between the exons involved in the regulated ASEs and a control set of expressed exons showing no regulation upon *ESRP1/2* gene silencing ( $|dPSI| < 0.01$ ,  $adj-p > 0.05$ ) revealed, as expected, an enrichment of ESRP1 and ESRP2 binding motifs. In particular, ESRP1 and ESRP2 binding motifs were enriched in ES events, with binding motifs prevalently predicted within exons (**Supplementary Table 18a,b**). Interestingly, functional enrichment of genes harbouring an ASE showing a predicted binding motif for ESRP1 and ESRP2 factors were enriched in terms related to estrogen response, angiogenesis, interferon alpha response, mitotic spindle, DNA repair, in addition to EMT and metabolism-related processes such as protein localization, cholesterol homeostasis and adipogenesis (**Supplementary Table 18c**). Noteworthy, only a subset of other RBPs for which a binding motif was predicted to be enriched for the ASEs were also DE upon the combined silencing of *ESRP1/2* genes. For instance, CELF1, ANKHD1, MBNL1, MBNL3, and RBM3 were both enriched and showed significant regulation at either gene or isoform level (**Supplementary Table 18d-k**).



**Figure 2.2-8:** Dotplots reporting the RBP-binding motifs enrichment results induced by the combined silencing of *ESRP1* and *ESRP2* genes. (a-b) Enriched RBP-binding motifs in case of exon inclusion (EI) (a) and exon skipping (ES) events, respectively. up, E, and down (x-axis) refer to intronic upstream, exonic, and downstream intronic genomic regions showing a predicted motif for the indicated RBP (y-axis). (c-d) Enriched RBP-binding motifs in case of IR events. (e-f) Enriched RBP-binding motifs in case of A5 'SS. up and down refer to upstream and downstream of splice sites, respectively. (g-h) Enriched RBP-binding motifs in case of A3 'SS events. up and down refers to upstream and downstream of splice sites, respectively. (i-j) Enriched RBP-binding motifs in case of MX events. V1-V6 refer to genomic locations, upstream and downstream of mutually exclusive exons, respectively. Color bar intensities represent the DE status of

enriched RBPs. Dot size is proportional to the enrichment z-score. Dot border color indicates the significance of the DE status of the RBP.

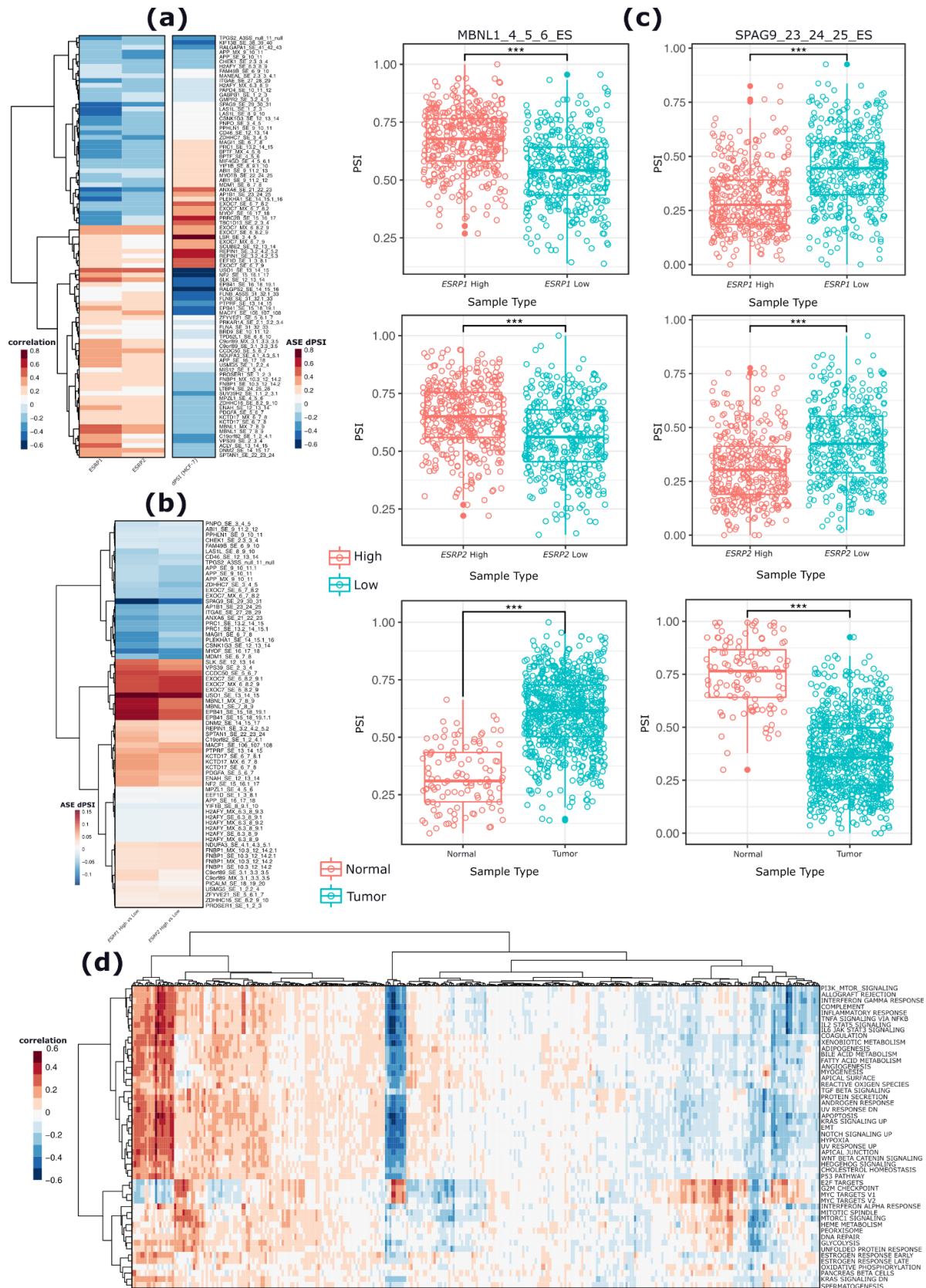
### 2.2.3 ESRP1 and ESRP2 splicing factors control a set of ASEs in ER $\alpha$ + BCs

To verify the importance of the set of ASEs detected upon the combined silencing of *ESRP1/2* genes in the ER $\alpha$ + MCF-7 BC cells, a differential inclusion/exclusion analysis was performed also in primary BC samples by exploring the ASEs PSI data from TGCASpliceSeq (M. Ryan et al. 2016). This analysis revealed 351 out of 1,052 ASEs (33%) detectable in BC samples including 773 ER $\alpha$ +, 192 ER $\alpha$ - tumors, in addition to 113 normal breast tissues (**Supplementary Table 19a-c**). Among these events, 175 and 107 ASEs showed in ER $\alpha$ + BC samples a significant correlation with the expression of *ESRP1* and *ESRP2* genes ( $|\text{Spearman } \rho| > 0.1$ , p-value  $< 0.05$ ), respectively, of which 90 ASEs coherently correlated with the expression of both genes (**Figure 2.2-9a and Supplementary Table 19d**). In particular, 82 ASEs positively and 93 ASEs negatively correlated with *ESRP1* gene expression, respectively. Similarly, 57 ASEs positively and 52 ASEs negatively correlated with *ESRP2* gene expression (**Figure 2.2-9a and Supplementary Table 19d**). Furthermore, 40 ASEs were differentially expressed relatively to the expression level of *ESRP1* (*ESRP1* high versus *ESRP1* low) and *ESRP2* genes (*ESRP2* high versus *ESRP2* low) (**Figure 2.2-9b and Supplementary Table 19e**). Importantly, most of the correlation between the ASEs PSI values and the expression of the investigated genes in ER $\alpha$ + BCs mirrored the induced effect of silencing the genes in MCF-7 cells (**Figure 2.2-9a**). Among the ASEs correlating with *ESRP1/2* mRNA levels in ER $\alpha$ + BC samples, the most significant positive correlation was observed between an ES event in the splicing factor *MBNL1* gene (exon: *MBNL1*\_chr3:152446703-152446757) and *ESRP1* and *ESRP2* gene expression (Spearman  $\rho=0.5$ , p-value=2.42E-48; Spearman  $\rho=0.30$ , p-value=6.38E-17, for *ESRP1* and *ESRP2*, respectively) (**Figure 2.2-9c and Supplementary Table 19d**). Consistently with our findings, the exon involved in this ES event is also strongly repressed during EMT process (Shapiro et al. 2011), and was found to play a functional role in the proper cellular localization of MBNL1 proteins within cells (Gates, Coonrod, and Berglund 2011). The MBNL1 isoforms containing this exon were identified as recurrently associated with different tumors such as colorectal, lung, and breast cancers (Danan-Gotthold et al. 2015). Moreover, MBNL1 regulates the splicing of its own pre-mRNA by binding on the intron upstream of the spliced exon, resulting in the upregulation of isoforms lacking the exon (Terenzi and Ladd 2010). The higher inclusion of this exon in many cancers, including BC as compared to normal tissues (**Figure 2.3-9c and Supplementary Table 19a**) is believed to be driven by the frequent downregulation of the MBNL1 protein in cancer (Gates, Coonrod, and Berglund 2011). Noteworthy, a second alternative exon (chr3:152455541-152455577) of *MBNL1* gene, important for MBNL1 protein homodimerization (Tabaglio et al. 2018), was induced upon combined silencing of *ESRP1* and *ESRP2*, and negatively correlated with *ESRP1* mRNA levels in ER $\alpha$ + BC samples (Spearman  $\rho=-0.13$ , p-value=0.0005) (**Supplementary Table 19d**). The expression of this exon is higher in cancer cell lines as well as in several solid tumors including BC as compared to adjacent normal tissues (Tang, Zhao, and Kong 2019). MBNL1 isoform lacking this exon was reported to act as a tumor suppressor in prostate and breast cancers by regulating the AS pattern and transcript abundance of genes involved in DNA repair, cell cycle, migration, and suppresses breast metastasis by regulating

transcript abundance of certain genes involved in these pathways (Tang, Zhao, and Kong 2019). Coherently with local AS analysis using rMATS, both dIU and dIE analyses revealed a differential regulation of MBNL1 isoforms including or lacking the differentially spliced exons (**Supplementary Figure 14**). In contrast, an exon inclusion event in the *SPAG9* gene inversely correlated with *ESRP1/2* gene expressions in ER $\alpha$ + BCs (**Figure 2.3-9c** and **Supplementary Table 19d**), where higher inclusion of this event associates with a lower expression of *ESRP1* and *ESRP2* genes (Spearman rho=-0.48, p-value=8.59E-46; Spearman rho=-0.34, p-value=2.39E-22, for *ESRP1* and *ESRP2*, respectively). In line with the siESRP1/2-induced effects in our cellular model, the inclusion of *SPAG9* exon 24 was reported to increase during EMT, with exon inclusion is induced by hnRPM and is inhibited by ESRP1 (Harvey et al. 2018). Moreover, higher inclusion of this exon was reported to associate with poorer patients survival in BC, and Harvey and coworkers showed that genes positively correlated with higher inclusion of this exon in mesenchymal cells, not expressing ESRP1, were all EMT-related genes (Harvey et al. 2018). In addition, the genomic blockade by silencing *SPAG9* gene using specific siRNAs targeting exons shared among all *SPAG9* isoforms was reported to suppress cell growth and inhibit invasive potential in TNBC cell lines (Sinha et al. 2013). However, expression analysis revealed a 20 to 52 folds higher expression of *SPAG9* mRNA and protein levels in a panel of BC cell lines, including MCF-7, MDA-MB231, BT20, and SK-BR-3, relative to that in normal mammary epithelial cells (Sinha et al. 2013).

In order to decipher which core molecular pathways are associated with the observed ASEs changes upon *ESRP1/2* gene silencing, a correlation analysis between the inclusion level (PSI) of each ASE in ER $\alpha$ + BCs and molecular pathways was performed using PEGASAS algorithm (Phillips et al. 2020). This pathway-guided enrichment analysis revealed a differential enrichment of ASEs in molecular pathways, resulting in two distinct clusters (**Figure 2.2-9d** and **Supplementary Table 19f**). In particular, the first cluster contained terms mainly related to interferon response, TGF $\beta$  signaling, EMT and metabolism, while the second cluster contained terms mainly related to proliferation, estrogen response, and cell cycle progression (**Figure 2.2-9d** and **Supplementary Table 18f**). Furthermore, unsupervised hierarchical clustering of ASEs PSI levels revealed subsets of ASEs differentially correlated with the two main clusters (**Figure 2.2-9d**). Notably, 20 ASEs were positively correlated with terms of the first cluster (e.g. PI3K-AKT-mTOR signaling, TGF $\beta$  signaling, EMT, interferon response and immune response pathways), while they were negatively correlated with terms of the second cluster (e.g. cell cycle, proliferation, estrogen response, and G2M checkpoint pathways). A second subset of ASEs, while they were negatively correlated with terms of the first cluster, they positively correlated with terms of the second cluster (**Figure 2.3-9d** and **Supplementary Table 18f**).



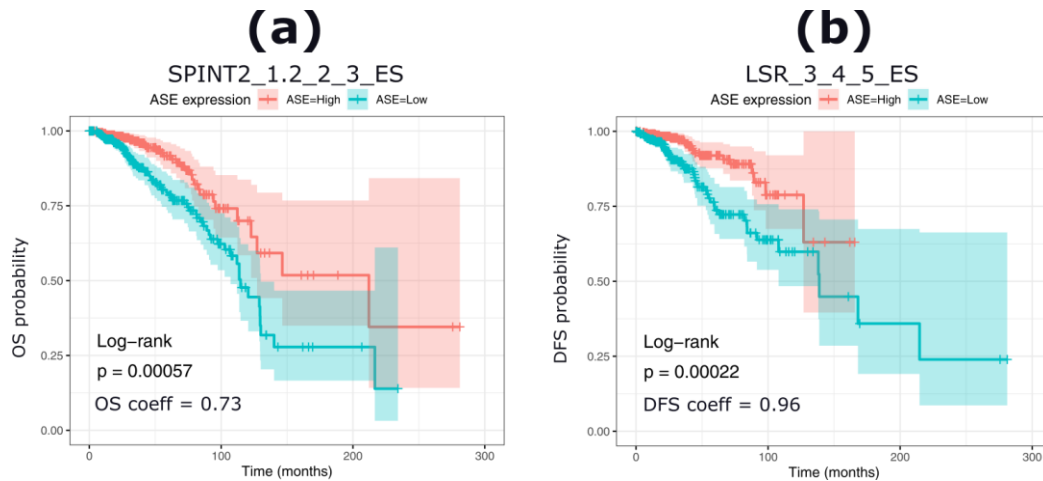


**Figure 2.2-9:** Overview of the expression analysis of ASEs detected in breast tumor data. (a) Heat map plot reporting the correlation between ASEs with the expression of ESRP1 and ESRP2 genes in ER $\alpha$ + tumors and The dPSI representing the changes of the correlating ASEs upon silencing ESRP1 and ESRP2 in MCF-7 cells. (b) Heat map plot reporting the dPSI values of ASEs as a function of ESRP1 and ESRP2 gene expressions in ER $\alpha$ + tumors. Only the ASEs significant in both comparisons (ESRP1 high, ESRP1 low and ESRP2 high, ESRP2 low) are reported. (c) Boxplots reporting the

expression level (PSI) of 2 selected ASEs as a function of *ESRP1* expression (top), *ESRP2* expression (middle), and in tumors compared to normal samples (bottom) (Wilcoxon p-value, \*\*\* < 0.001). (d) Heatmap plot showing the correlation between the expression of ASEs regulated by *ESRP1/2* silencing in MCF-7 cells and related molecular pathways in ER $\alpha$ + tumors.

### 2.2.3.1 The splicing factors *ESRP1* and *ESRP2* regulate the expression of a set of ASEs with a prognostic value in ER $\alpha$ + BCs

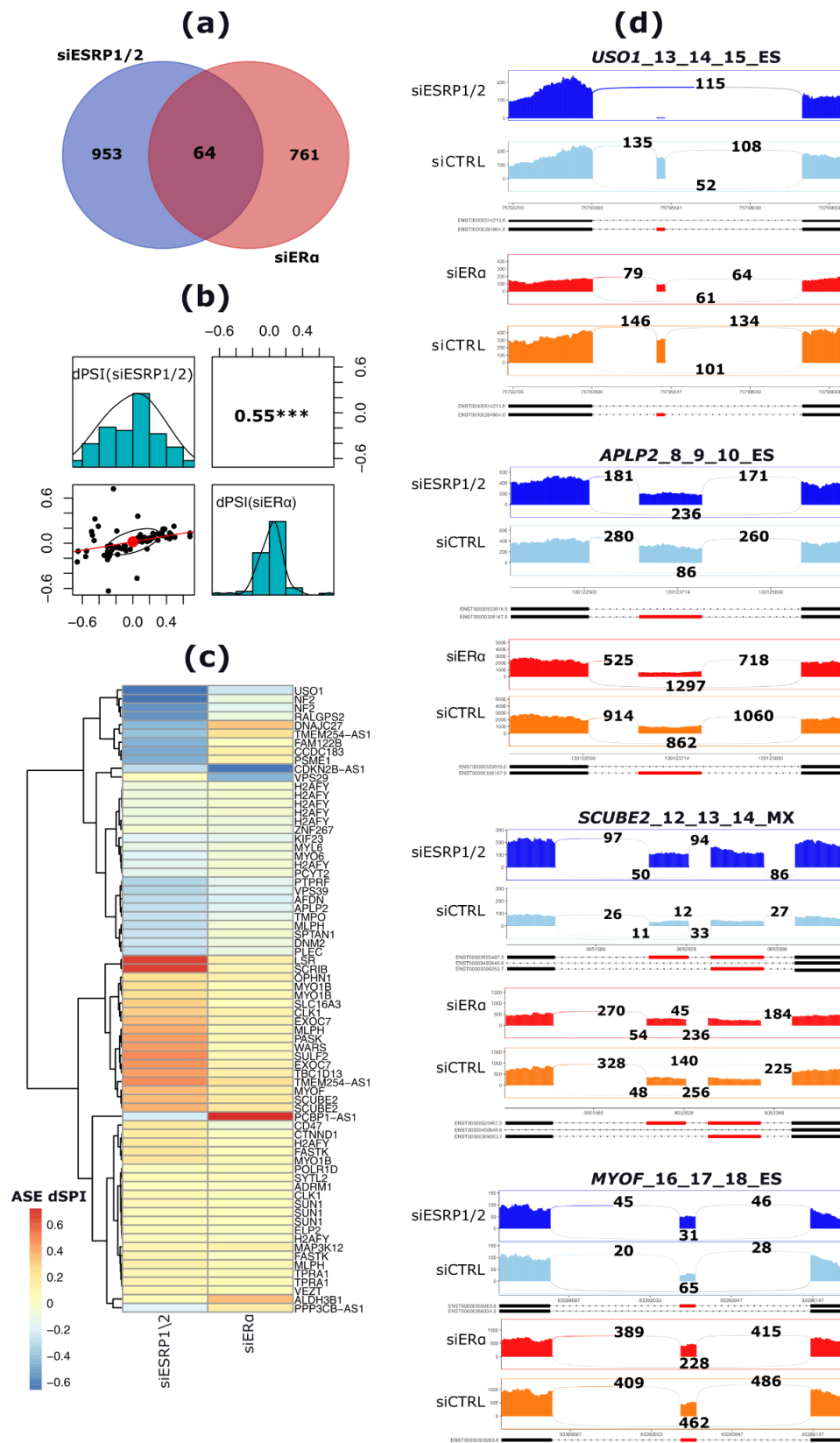
To determine the prognostic value of the ASEs significantly changing upon the combined silencing of *ESRP1* and *ESRP2* genes in the ER $\alpha$ + MCF-7 BCs, the association of the PSI level of each ASE, retrieved from TCGASpliceSeq (M. Ryan et al. 2016), with the overall and disease-free survival times, retrieved from the TCGA GDC data portal (Grossman et al. 2016b), was explored. This analysis revealed 36 ASEs significantly associated with overall survival (OS) or disease-free survival (DFS) times of ER $\alpha$ + BC subjects. In particular, 20 ASEs including 16 and 4 were positively and negatively associated with OS time, respectively. Similarly, 16 ASEs including 8 and 8 were positively and negatively associated with DFS, respectively. The top significantly associated event with patient OS is an EI of exon 2 of the *SPINT2* gene, where higher expression of the exon associates with longer OS time (HR= 0.73; log-rank p-value =5.73E-4) (**Figure 2.2-10a** and **Supplementary Table 19g**), while the top significantly associated event with DFS was an ES of exon 4 of *LSR* gene, where higher inclusion of the exon associates with better DFS (HD= 0.96; log-rank p-value =2.22E-4) (**Figure 2.3-10b** and **Supplementary Table 19g**). Noteworthy, higher expression of the EI event in the *ANXA6* gene is also associated with a longer patient OS (HD= 0.42, log-rank p-value = 0.04) (**Supplementary Table 19g**). In contrast, higher expression of an EI event in the *FLNA* gene involving exon 32 is associated with both shorter OS and worse DFS in ER $\alpha$ + BCs (HR=-0.63, log-rank p-value=0.0021; HR=-0.57, log-rank p-value=0.02, for OS and DFS, respectively) (**Supplementary Table 19g**). Noteworthy, the annotation of the exon involved in this event revealed that it encodes for a protein domain (filamin/ABP238 repeats, PF00630) involved in the direct physical interaction of FLNA protein with ER $\alpha$ , androgen receptor (AR), and the Nuclear Receptor Subfamily 2 Group C Member 2 (NR2C2) proteins (Louadi et al. 2021; Tarallo et al. 2011; Cheng, Chang, and Chen 2010). In addition, dIU analysis also revealed a switching behaviour in the relative abundance of *FLNA* gene isoform pairs, where the the inclusion form of the event was induced while the exclusion form of the event was repressed (**Supplementary Figure 15**). Taken together, these data suggest that *ESRP1*, notably *ESRP2*, regulate the expression levels of ASEs that could serve as prognostic biomarkers in the context of BC, some of which some encode for protein domains physically interacting with ER $\alpha$  at the protein level.



**Figure 2.2-10:** Kaplan Meier survival plots reporting top ASEs significantly associated with overall survival (OS) (a) and disease-free survival (DFS) times (b) in ER $\alpha$ + subjects, respectively. Patients are divided into highly expressing (ASE=High) and lowly expressing the event (ASE=Low) based on the median inclusion level (PSI) calculated among all patients. ASE, alternative splicing event; ES, exon skipping; OS, overall survival, DFS, disease-free survival. Events names are provided above the plots accordingly.

#### 2.2.4 A set of ESRP1/2 regulated ASEs occurs also upon silencing ER $\alpha$ in hormone-deprived MCF-7 cells

In this study, we provide evidence that unliganded-ER $\alpha$  regulates the expression of *ESRP1* and *ESRP2* genes independently of estrogenic stimuli in hormone-deprived MCF-7 BC cells. Therefore, we sought to identify those AS changes occurring in both ER $\alpha$  and *ESRP1/2* silencing experiments. The overlap between the two experiments revealed 64 commonly dysregulated events (exact hypergeometric test p-value,  $p < 0.001$ ), particularly classified as ES (50, 78%), MX (5, 7.80%), A5 (4, 6.25%), A3 (2, 3.13%), and finally 3 (4.68%) IR events (**Figure 2.2-11a** and **Supplementary Table 20a**). The analysis of the ASEs dPSI values obtained in the two experiments confirmed a significantly high correlation (Spearman rho = 0.55,  $p < 0.0001$ ), that was particularly high for ES/EI events (Spearman rho = 0.89,  $p < 0.0001$ ) (**Figure 2.2-11b**) with 46 events (86%) showing a coherent dysregulation (**Supplementary Table 20a**). The top four significant ASEs overlapped and coherent between the two experiments included the skipping of exon 14 of *USO1* gene, skipping of exon 7 of *APLP2* gene, the inclusion of exon 13 of *SCUBE2* gene, and the inclusion of exon 17 of *MYOF* gene (**Figure 2.2-11c** and **Supplementary Table 20a**).



**Figure 2.2-11:** An overview of AEs regulated upon both si-ERα in MCF-7 BC cells and si-ESRP1/2. (a) Venn diagram reporting the number of AEs overlapping both datasets. (b) Correlation plot showing the correlation between the dPSI values of the 64 overlapping events shown in (a). (c) Heat map showing the direction of splicing changes (dPSI values) of the overlapping AEs. (d) sashimi plots of selected examples of AEs regulated in both datasets. Numbers representing

the normalized read counts supporting either inclusion or exclusion of regulated exons (highlighted in red colors) are reported. Isoforms concerned by the ASEs are also shown.

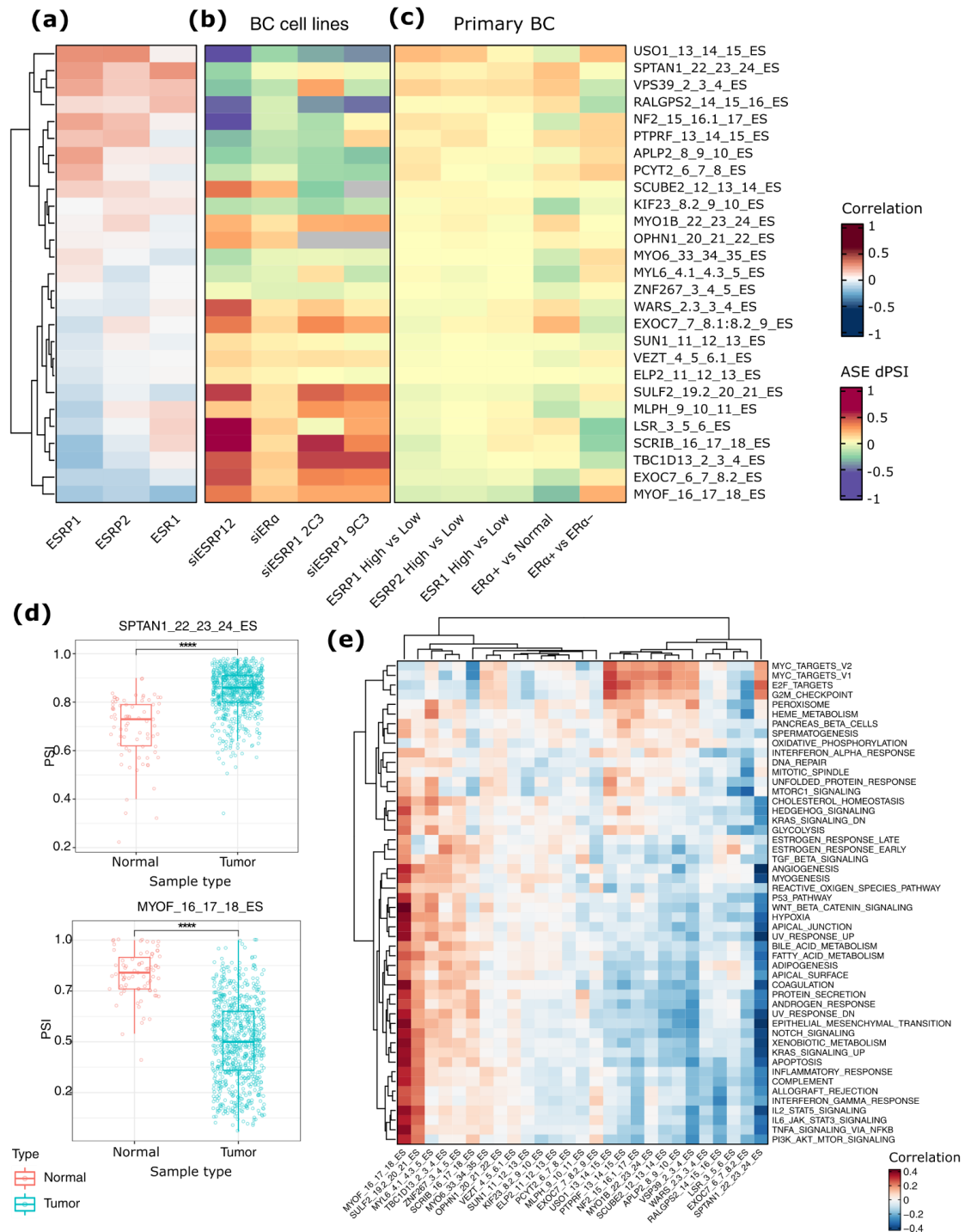
#### 2.2.4.1 Analysis of ER $\alpha$ and ESRP1/2 regulated ASEs in tumor tissues data

The set of ASEs detected as overlapped in the *ESRP1/2* and *ER $\alpha$*  silencing experiments were explored in primary BC ASE data from TCGASpliceSeq (M. C. Ryan et al. 2012) revealing 27 unique events detectable in primary BCs (**Supplementary Table 20b**). Among these events, 23 (85.2%) and 18 (66.7%) correlate significantly ( $p < 0.05$ ) with *ESRP1* and *ESRP2* gene expressions, respectively (**Figure 2.2-12a** and **Supplementary Table 20b**). Twelve ASEs were correlated with the expression of both genes. The most significantly correlated event was an ES on *USO1* gene (Spearman rho = 0.44 and 0.45 for *ESRP1* and *ESRP2*, respectively;  $p < 0.0001$ ). Finally, six ASEs were also correlated with *ESR1* expression, particularly an ES event at *SPTAN1* (Spearman rho = 0.39, 0.19, and 0.42 for *ESRP1*, *ESRP2*, and *ESR1*, respectively;  $p < 0.0001$ ), *MYOF* (Spearman rho = -0.30, -0.23, and -0.36 for *ESRP1*, *ESRP2*, and *ESR1*, respectively;  $p < 0.0001$ ) and *VPS39* genes (Spearman rho = 0.34, 0.19, and 0.23 for *ESRP1*, *ESRP2*, and *ESR1*, respectively;  $p < 0.0001$ ) (**Figure 2.2-12a**). Analysis of the 27 ASEs using *siESRP1* RNA-Seq data of endocrine-resistant BC cell lines (2C3, 9C3 cell lines) from (Gökmen-Polar et al. 2019), confirmed 14 ASE (63.6%) significantly (adj.  $p < 0.05$ ) regulated upon *ESRP1* silencing in these cells (**Figure 2.2-12b**). The regulation of these ASEs was coherent between our *siESRP1/2* silencing experiment and the one performed in 2C3 (Spearman rho = 0.61,  $p < 0.05$ ) and in 9C3 cells (Spearman rho = 0.65,  $p < 0.05$ ).

Then, the differential inclusion/exclusion levels of the 27 ASEs were evaluated in the TCGA BC data relative to the median mRNA expression levels of *ESRP1*, *ESRP2*, or *ESR1* genes in these tumors (**Figure 2.2-12c**). This analysis confirmed a coherent opposite inclusion/exclusion level between the *siESRP1/2* silencing in MCF-7 and tumors underexpressing *ESRP1* (Spearman rho = 0.83,  $p < 0.0001$ ), *ESRP2* (Spearman rho = 0.65,  $p < 0.001$ ). Furthermore, when BC PSI data were compared with PSI values computed using RNA-Seq of normal breast tissues, a total of 21 ASEs were confirmed as characterized by significantly different inclusion/exclusion levels (**Figure 2.2-12c**). The two most significantly different ASEs obtained in this analysis are reported in **Figure 2.2-12d**. In particular, the most significant ASE was an ES of the *MYOF* gene where exon 17 was significantly more excluded in BC compared to normal tissues (dPSI = -0.33,  $p < 0.0001$ ). Conversely, the exon 23 of *SPTAN1* was significantly more included in BC compared to normal samples (dPSI = -0.16,  $p < 0.0001$ ). Coherently with the analysis of clinical data in relation to *ESRP1* and *ESRP2* expression levels, most of the 27 ASEs were significantly related ( $p < 0.05$ ) to the tumor molecular subtype (22 events), the fraction of altered genome (20 events), and menopause status and diagnosis age (8 events) (**Supplementary Table 20c**).

To functionally enrich the 27 ASEs with respect to specific molecular pathways, an analysis with PEGASAS (Phillips et al. 2020) was performed considering the PSI values of these ASEs measured in TCGA BCs (**Supplementary Table 20d**). The analysis revealed two main clusters of pathways correlated with the 27 ASEs (**Figure 2.2-12e**). Specifically, a cluster composed of terms mainly to inflammation process, including interferon gamma response, TNF $\alpha$  signaling via NF $\kappa$ B, and IL6/JAK/STAT3 signaling was positively

correlated ( $r > 0.30$ ) with six ES events involving *MYOF*, *SULF2*, *MYL6*, *TBC1D13*, *ZNF267*, and *SCRIB* (Figure 2.2-12e and Supplementary Table 20d). Terms related to EMT and metabolism were also included in this cluster. Conversely, another cluster of molecular pathways related mainly to cell proliferation processes like G2M checkpoint, and E2F targets was positively related to ES events at *USO1*, *PTPRF*, *NF2*, *MYOB*, *SCUBE2*, *VPS39*, and *SPTAN1*.



**Figure 2.2-12:** Analysis of the 27 ASEs overlapping between the two experiments in tumor data. (a) heat map plot showing the correlation between *ESRP1*, *ESRP2* and *ESR1* genes expression and the expression of ASEs exons in tumor samples, whose inclusion changes (*dPSI* values) upon *ESR1* or *ESRP1/2* silencing in wild type (*MCF-7*) or resistant (*LCC2*, *LCC9*)

*BC cell lines are reported in (b). (c) heat map plot showing the inclusion levels of the 27 ASEs in primary BC samples classified according to the expression levels of ESRP1, ESRP2 and ESR1 genes. (d) boxplots representing the inclusion levels (Percent Spliced-in Index, PSI) of exon 22 of SPATIN1 and exon 16 of MYOF genes in tumor versus normal samples. (e) heat map showing the correlation scores between ASEs and the enriched molecular pathways obtained from PEGASAS analysis.*

### 2.2.5 Discussion:

In this study, the functional roles of the two splicing factors ESRP1 and ESRP2 in the control of ER $\alpha$ + BC transcriptome were explored. Moreover, evidence for the direct regulation of the expression of these two epithelial factors by ER $\alpha$  in ER $\alpha$ + BC was provided, as shown by ER $\alpha$  RNA-Seq and ChIP-Seq data analysis in both cell lines as well as in primary BCs. The effects of ESRP1/2 depletion on the ER $\alpha$ + MCF-7 BC cells transcriptome were explored, revealing a significant number of regulated events at both gene and splicing levels. Furthermore, the potential role(s) and clinical significance of these events were explored in primary BCs by analyzing the expression of both genes and its association with AS data in different BC subjects. Finally, the observed posttranscriptional events resulting from the depletion of ESRP1/2 in MCF-7 cells were compared to those resulting from silencing the upstream regulator of both genes, ER $\alpha$ , in MCF-7 cells. This resulted in the identification of common post transcriptional events, some of which could serve as prognostic markers in the context of BC.

The splicing factors ESRP1 and ESRP2 are not only overexpressed in ER $\alpha$ + BCs but also showed significant copy number alterations in these tumors. In particular, we show that the ESRP1 locus is amplified in 60% of analysed samples, suggesting that this amplification may be a mechanism underlying the overexpression of *ESRP1* gene in ER $\alpha$ + BCs. Our results are in line with (Sebestyén et al. 2016) showing that *ESRP1* among other genes located near the reported 8q24 amplification (Cancer Genome Atlas Network 2012) is overexpressed in BCs and its overexpression is frequently associated with frequent amplification of the gene in these tumors. Our results are also in line with (Jeong et al. 2017) showing that *ESRP1* is overexpressed in ovarian cancer and that this overexpression is associated with copy number amplification of the gene locus in these tumors. Jeong and coworkers also demonstrated DNA hypomethylation as another mechanism underlying the observed *ESRP1* and *ESRP2* overexpression in ovarian cancers (Jeong et al. 2017). In contrast, our results regarding *ESRP1* amplification are discordant with observations of *ESRP1* and *ESRP2* genes underexpression in colorectal cancer tissues (Deloria et al. 2016), which potentially suggests that the expression of these two factors may be cancer-type-specific.

As expected, siRNA-mediated silencing of both *ESRP1* and *ESRP2* genes in the epithelial ER $\alpha$ + MCF-7 BC cells led the cells towards a more mesenchymal phenotype, as previously reported in other cell lines (Yueqin Yang et al. 2016; Warzecha et al. 2010; Huang, Xu, and Cheng 2014). *ESRP1/2* combined silencing induced significant changes in the expression as well as the AS patterns of EMT-related genes, in line with previously reported studies (Yueqin Yang et al. 2016; Shapiro et al. 2011). Noteworthy, although ESRP1 and ESRP2 are well described for their role in controlling EMT, silencing the two factors in MCF-7 cells did not induce a strong overall gene expression change of the EMT-related genes as reported by GO enrichment analysis at gene level, in line with previous observations (Shapiro et al. 2011; Gökmen-Polar et al. 2019). In contrast, ESRP1/2 depletion strongly hampered the splicing patterns of a great number of EMT-related genes. Notably, comparing of the list of genes whose expression or AS patterns significantly changed under *ESRP1/2* silencing in our cellular model with a study by Yang and colleagues (Yueqin Yang et al. 2016), 109 coherent (100%) AS pattern changes were common to both studies (**Supplementary Figure 16a,b** and **Supplementary Table**



**21a).** Furthermore, the comparison of our list of ASEs differentially expressed upon *ESRP1/2* silencing with an EMT time-course RNA-Seq dataset by overexpressing ZEB1 in epithelial cells, by (Yueqin Yang et al. 2016), revealed 171 common ASEs, most of which (155 ASEs, 90.64%) showed coherent changes between the two studies (**Supplementary Figure 16c,d** and **Supplementary Table 21b**), further supporting our findings. In addition, RBP-binding motif enrichment analysis revealed a higher number of ASEs with a predicted binding motif for *ESRP1/2* splicing factors, suggesting the direct implication of these two factors in the observed AS pattern changes in our cellular model. Among this, 33 ASEs regulated in our dataset and harboured a predicted motif for *ESRP1* and *ESRP2* splicing factors are in line with previously reported results (Harvey et al. 2018), showing the presence of *ESRP1/2* binding motifs within 233 target exons. Moreover, the inclusion of these events was found to be induced or repressed by the two antagonistic factors *ESRP1* and *hnRNPM* during the process of EMT (Harvey et al. 2018). In line with previous results, most genes harbouring a predicted binding site for *ESRP1/2* splicing factors were primarily enriched in EMT-related processes.

In addition to EMT, *ESRP1/2* silencing in MCF-7 cells induced significant changes in the expression as well as the AS patterns of genes primarily involved in interferon signaling, metabolism, and cell cycle pathways. In particular, genes encoding for cell signaling molecules were strongly associated with a splicing regulation induced by *ESRP1/2* combined silencing. For instance, the Rac family small-GTPase 1 (*RAC1*), and *KRAS* proto-oncogene, GTPase (*KRAS*) genes had a significant AS pattern switch induced by *ESRP1/2* silencing. In the case of *RAC1* gene, the Rac1b isoform including the exon 3b was decreased by *ESRP1/2* combined silencing, while the canonical *RAC1* was induced. The Rac1b isoform is constitutively active and highly expressed in different cancers including breast, thyroid, colorectal and lung tumors (Schnelzer et al. 2000; Gonçalves et al. 2014; C. Zhou et al. 2013). Furthermore, canonical *RAC1* and Rac1b isoforms were characterized using data from pancreatic ductal adenocarcinoma and BC cell lines as antagonistically acting on the EMT process, where *RAC1* induces EMT, while Rac1b represses EMT (Zinn et al. 2019; Eiden and Ungefroren 2021). The correlation analysis of exon 4 (or conventionally exon 3b) inclusion levels with molecular pathways in  $ER\alpha+$  BCs showed an inverted correlation of high inclusion of exon 4 with each of PI3K/Akt/mTOR signaling pathway, metabolism and oxidative phosphorylation, with interferon signaling, JAK/STAT3 signaling, TGF $\beta$  and EMT pathways (**Supplementary Table 19f**), in line with previously reported results (Melzer et al. 2019, 2017), indicating that Rac1b isoform is not able to activate these pathways (Ungefroren et al. 2020). In contrast, higher inclusion level of exon 4 (and thus higher Rac1b isoform expression) positively correlated with estrogen response and Myc targets pathways (**Supplementary Table 19f**), in line with (Eiden and Ungefroren 2021) showing a clear association between higher Rac1b/*RAC1* ratio and  $ER\alpha+$  epithelial BC phenotype.

Analysis of *ESRP1/2* silencing-induced AS changes revealed a number of ASEs with potential prognostic value in the context of BC. Of interest, 11 *ESRP1/2*-regulated ES/EI events were recently identified as robust classifiers of Basal-like BCs into A and B subtypes based on their aggressiveness and drug-response (Villemin et al. 2021). Noteworthy, all of these 11 ASEs in genes (*ANXA6*, *AP1B1*, *CTNND1*, *DNM2*, *ENAH*, *FNBP1*, *MBNL1*, *SLK*, *SPAG9*, *TSC2*) were also differentially expressed in the *ESRP1* knock down study by (Yueqin

Yang et al. 2016). The top significant ASE among this list of events corresponds to the significant exon inclusion of exon 22 of the annexin A6 (*ANXA6*) gene (**Figure 2.2-13a** and **Supplementary Table 16a**). Although it is completely unknown for its splicing role in BC, this event was reported to be highly expressed in the Basal-like B as compared to Basal-like A BC patients and cell lines, and was proposed as a novel ASE to stratify these groups of patients based on their prognoses and drug-response (Villemin et al. 2021). The expression analysis of this exon inclusion event in normal, ER $\alpha$ +, and ER $\alpha$ - breast samples indicate the higher expression of the event in normal and ER $\alpha$ - breast tissues with respect to ER $\alpha$ + samples, suggesting a negative correlation of event expression and ER $\alpha$ + status (**Figure 2.2-13c**). More particularly, this analysis in ER $\alpha$ + BC subtype revealed that the event is significantly inversely correlated with the expression status of the ER $\alpha$ -regulated *ESRP1* and *ESRP2* genes (Spearman rho = -0.42, adj-p = 1.04E-33; Spearman rho = -0.32, adj-p = 3.39E-19, for *ESRP1* and *ESRP2*, respectively) (**Figure 2.2-13b** and **Supplementary Table 19d**). Furthermore, differential expression analysis of the event indicated significant inclusion differences between highly and lowly *ESRP1/2* expressing patients (dPSI = -0.075, adj-p = 2.61E-23; dPSI = -0.06, adj-p = 2.52E-14 for *ESRP1* and *ESRP2*, respectively) (**Figure 2.3-13c** and **Supplementary Table 19e**). In addition, the association analysis of this event with the OS and DFS in ER $\alpha$ + patients revealed that high expression of the event associates with a better patient OS time (HR=0.42, log-rank p-value = 0.04) (**Figure 2.2-13d** and **Supplementary Table 19f**). Conversely, a high expression of *ESRP1* gene in ER $\alpha$ + BCs is associated with worse patient's outcomes, including shorter OS, and both worse DMFS and RFS (HR=2.11, log-rank p-value=0.029; HR=2.82, log-rank p-value=0.014; HR=1.8, log-rank p-value=8.4E-5, for DMFS and RFS, respectively) (**Figure 2.2-13d**). Noteworthy, in ER $\alpha$ + BCs, the overall *ANXA6* gene expression is also associated with a better patient OS (**Figure 2.2-13d**). Furthermore, a high expression of the event in ER $\alpha$ + BC patients positively correlated with molecular pathways such as the IL2/STAT5 and IL6/JAK/STAT6 signaling pathways, with KRAS signaling, EMT, and TGFB signaling pathways, with TNF $\alpha$  signaling via NF $\kappa$ B, with PI3K/Akt/mTOR and P53 signaling pathways, in addition to androgen and estrogen signaling pathways. Conversely, a high expression of the event is inversely correlated with MYC targets, G2M checkpoints, and E2F targets hallmarks (**Supplementary Table 19g**). All together, these data suggest a potential role of the discussed event in BC and that it could serve as a potential biomarker for patients' outcomes.

Moreover, the isoform switch analysis (dIU) applied on our si*ESRP1/2* RNA-seq dataset indicated that the *ANXA6* exon 22 encodes for the Annexin protein domain (PF00191). Further annotation of this protein domain, using the DIGGER database (Louadi et al. 2021) and other literature (Garbuglia et al. 2000), indicated its involvement in the interaction with the two calcium sensing proteins, S100A1 and S100B. The functional biological role of these complexes (*ANXA6-S100A1* and *ANXA6-S100B*) is poorly annotated in literature, with some exceptions where these complexes were suggested to play a role in regulating Ca<sup>2+</sup> fluxes in skeletal cells (Arcuri et al. 2002). Noteworthy, the domain-domain interaction of *ANXA6* with S100A1 and S100B proteins mediated by the Annexin protein domain (PF00191) encoded by exon 22 is completely lost (Louadi et al. 2021) by the skipping of the exon regulated by *ESRP1/2* silencing in our dataset. In addition, *ANXA6* long (including exon 22, *ANXA6-1*) and short (excluding exon 22, *ANXA6-2*) isoforms were described for their

differential regulation of Ca<sup>2+</sup> influxes in PC12 cell lines (Podszywalow-Bartnicka et al. 2010). Importantly, among the interactions involving S100B protein is its direct physical interaction with the fibroblast growth factor 2 (FGF2) protein via its EF\_hand\_1 interaction domain (Louadi et al. 2021), as previously confirmed by (Gupta et al. 2013). Furthermore, S100B was shown to bind the basic FGF2 (bFGF2) and this binding interferes with FGF2-FGFR1 interaction, which decreases the FGF2-induced activation of the PI3K/Akt signaling pathway and consequently reduces FGF2-induced cellular proliferation of MCF-7 and MDA-MB468 BC cell lines (Gupta et al. 2013). Moreover, a recent study in triple-negative BCs showed that the *S100B* gene is relatively lowly expressed in tumor samples compared to normal samples (Yen et al. 2018). Yen and colleagues observed a significant inhibitory effect on cell migration and an increase in the expression of epithelial E-cadherin when transfected the TNBC cell lines, MDA-MB-231 and Hs578T, with recombinant human S100B, suggesting S100B as a biomarker for metastasis in BC (Yen et al. 2018).

The *ANXA6* gene was indeed investigated in multiple cancers, revealing the usefulness of its mRNA and protein expression states as a biomarker in various cancers (Qi et al. 2015). Particularly in BC, *ANXA6* is downregulated in EGFR-overexpressing and ER $\alpha$ - BC cells, and was identified to inhibit the EGFR/Ras signaling pathway (Vilá de Muga et al. 2009). The reduced expression of *ANXA6* gene in BT-549 invasive BC cell lines promoted a rapid degradation of the active EGFR, consequently attenuated the downstream signaling activity of the MAPK/ERK/PI3K/Akt pathway, and enhanced the anchorage-independent cell growth (Vilá de Muga et al. 2009). *ANXA6*-deficient cells were also more sensitive to EGFR-targeted tyrosine kinase inhibitors, lapatinib and PD153035 (Vilá de Muga et al. 2009). The reduced expression of *ANXA6* gene is also associated with better RFS but with poorer DMFS of basal-like BC patients (Vilá de Muga et al. 2009). Recently, it was reported that a prolonged exposure to lapatinib of TNBC cell lines, expressing low *ANXA6* levels, induced a significant increase in the *ANXA6* mRNA levels and led to accumulation of cholesterol in late endosomes (Widatalla et al. 2019). The inhibition of lapatinib-induced upregulation of *ANXA6* in lapatinib-resistant BC cell lines, MDA-MB468, using specific siRNAs for *ANXA6* or by withdrawal of lapatinib from these cells, localized cholesterol to plasma membrane, restored the EGFR-dependent activation of the ERK1/2 pathway, and sensitized the cells to tyrosine kinase inhibitors (Widatalla et al. 2019). Loss of *ANXA6* was associated with early onset and rapid tumor growth of xenograft TNBC tumors in mice (Whalen et al. 2019). Efforts investigating the molecular mechanisms underlying *ANXA6* contribution to TNBC cell growth and motility indicated the Ca<sup>2+</sup> activated RasGRF2 as the potential mediator of *ANXA6* tumor suppressor effects in TNBC (Whalen et al. 2019). All together, these data suggest evidence that the expression status as well as the AS patterns of *ANXA6* in ER $\alpha$ + BCs will open new strategies that may improve the prognosis of BC patients and predict their response to certain targeted therapeutic options.

ESRP1/2 combined silencing in the ER $\alpha$ + MCF-7 BC cells induced a significant increase in the expression of *PGR* gene (log<sub>2</sub>FC= 1.14, adj-p= 1.73E-12) (**Supplementary Table 13a**). Noteworthy, the observed increase in the *PGR* gene expression occurred with no observed changes in the expression level of ER $\alpha$  (log<sub>2</sub>FC = 0.21, adj-p = 0.22), suggesting that ER $\alpha$  pathway is not primarily involved in the observed ESRP1/2-mediated increase of *PGR* gene expression. *PGR* was previously characterized to attenuate tumor growth by contracting

the mitogenic signals and actions of ER $\alpha$  in BCs co-expressing both receptors (ER $\alpha$ + /PGR+) (Finlay-Schultz et al. 2017). The treatment of (ER $\alpha$ + /PGR+) patient-derived xenografts with natural or synthetic (MPA) progestins was able to antagonize mitogenic effects of estrogens, blunted ER $\alpha$ -mediated gene expression, and reduced tumor growth of these BC models (Finlay-Schultz et al. 2017). The genome-wide analysis of the PGR-mediated actions in these models indicated that only less than 25% of ER $\alpha$  bindings are affected by progestins treatment, suggestive of other possible PGR-mediated mechanisms (Finlay-Schultz et al. 2017). DNA-binding profiling of both receptors indicated two distinct groups based on the binding profiles of PGR. In one group, more than 50% of PGR binding sites were co-occupied by ER $\alpha$ , showing a propensity for both receptors to coordinately gain or lose target binding in presence of progesterone, while in a second group, PGR but not ER $\alpha$ , associated with a large fraction of RNA polymerase III-transcribed tRNA genes, independent of hormone treatment (Finlay-Schultz et al. 2017).

Interestingly, ESRP1/2 silencing also regulated the AS pattern of *PGR* gene in our cellular model (**Supplementary Figure 17**). This regulation at splicing level resulted in the differential expression and usage of PGR isoforms. PGR isoforms expression in BC cell lines and prognostic values in BC patients have been previously reported. Notably, the association of PGR-A isoform with invasion and metastasis activation was previously indicated (McFall et al. 2018). PGR-A isoform promoted metastasis in luminal BCs by controlling the expression of microRNAs involved in the crosstalk with ER $\alpha$  signaling pathway (McFall et al. 2018). In contrast, PGR-B isoform was found to counteract and reduce luminal BC aggressiveness and to inhibit the transcription of critical genes involved in proliferation such as cyclin D1 (Montalto et al. 2019). In different BC cell lines, such as T47D, the exposure to progesterone induces a rapid activation of the EGFR, c-Src, and MAPK signaling pathways which results in an increase in the phosphorylation levels of PGR at Ser345 (Pedroza, Subramani, and Lakshmanaswamy 2020). On one hand, ESRP1 is reported to be associated with worse outcomes in ER $\alpha$ + but not ER $\alpha$ - BC patients and is overexpressed in tamoxifen and fulvestrant resistant cell lines (Gökmen-Polar et al. 2019). On the other hand, PGR gene expression is lost during endocrine therapy which leads to more aggressive tumors (Branković-Magić et al. 2002) and ER $\alpha$ + /PGR- BC patients had poorer survival outcome compared to ER $\alpha$ + /PGR+ ones (Blows et al. 2010). In addition, other studies reported that a loss in PGR results in the activation and upregulation of the PI3K pathway in tamoxifen resistant cell lines (Arpino et al. 2005). Putting these together, it is reasonable to conclude that the *ER $\alpha$ /ESRP1/PGR* axis presented in this study could be a good target to further investigate the molecular mechanisms underlying the worse ESRP1 prognostic value in ER $\alpha$ + BCs, especially by deciphering its impact on PGR protein isoforms and their interactions at the promoter or enhancers of target genes.

A recently published study reported that FGF2 is capable of inducing BC growth through the ligand-independent activation and recruitment of ER $\alpha$  and PGR isoforms A, B, and 4 (PGR<sub>A</sub>, PGR<sub>B</sub>, PGR <sub>$\Delta$ 4</sub>, respectively) to regulatory sequences of target genes, such as Myc (Giulianelli et al. 2019). In this study, authors treated hormone-deprived MCF-7 and T47D cell lines with FGF2 for 5 minutes and observed a significant increase in the phosphorylation levels of both ER $\alpha$  (pSer167 and pSer118) and PGR (pSer294) (Giulianelli et al. 2019). Interestingly, this increase was accompanied by increased ERK1/2 and AKT

activation in both cell lines. The application of small molecule inhibitors specific for ERK1/2 (PD98059), PI3K (LY294002), or FGFR (PD173074 and BGJ398), blocked the FGF2-induced ER $\alpha$  and PGR phosphorylation, suggesting that the signal transduction pathways activated by FGF2 through different FGFRs are responsible for the increased ER $\alpha$  and PGR phosphorylation in hormone-deprived cells (Giulianelli et al. 2019). Similarly, the FGF2-induced cell proliferation was strongly inhibited by addition of antiestrogens (fulvestrant, ICI 182.780), or by the FGFR inhibitors (PD173074 and BGJ398) to hormone-starved cell lines (Giulianelli et al. 2019). Genetic blockade using siRNA targeting ER $\alpha$  in hormone-starved MCF-7 cell lines had similar inhibitory effects on cell proliferation as those induced by antiestrogens (Giulianelli et al. 2019). Importantly, using T47D cell lines which have a higher basal estrogen-independent expression of *PGR*, authors showed that FGF2-induced Myc expression is mediated by ER $\alpha$  binding at both promoter and enhancer regions of the gene, and by PGR binding at enhancer region (Giulianelli et al. 2019). Interestingly, authors showed that FGF2 induces an isoform-specific expression of PGR isoforms A, B and PGR $\Delta$ 4 under hormone-starved conditions, which interact with ER $\alpha$  at the promoter of target genes (Giulianelli et al. 2019).

Importantly, mass-spectrometry-based proteomics analysis of ER $\alpha$  and PGR protein interactome at the regulatory sequences of target genes revealed significant changes in the relative abundance of *ANXA6* isoforms (P08133-1, and P08133-2) upon FGF2 stimulation of hormone-starved MCF-7 and T47D cell lines (Giulianelli et al. 2019). Particularly, the *ANXA6* long isoform increased by ESRP1/2 silencing in our dataset was relatively increased upon stimulation with FGF2 (Giulianelli et al. 2019). These data suggest that silencing ESRP1/2 may induce an increase in the activation of the PI3K/AKT signaling pathway through the *ANXA6/SB100/FGF2* axis, which results in an increased expression of *PGR*. On the other hand, silencing ESRP1/2 in our cellular model regulated the AS patterns of 40 genes coding for protein partners of PGR, including *ANXA6*, that were confirmed to interact with PGR at the regulatory sequences (promoters and enhancers) of target genes such as *Myc* (Giulianelli et al. 2019). All together, these data suggest a possible implication of ESRP1/2 splicing factors in the control of PGR signaling and transcriptional activity, either directly, by regulating the expression of the *PGR* gene, or indirectly, by regulating the AS patterns of genes encoding for upstream regulators of PGR or for its partners. In addition, ESRP1/2 through the control of the AS patterns of genes involved in intracellular signaling such as those involved in growth receptor signaling pathway, notably, *ANXA6*, previously described to inhibit the Ras/ERK1/2 signaling pathway, attenuate the expression of *PGR* gene, and mediates its downstream effects.

ESRP1/2 regulated the expression of 22 ASEs exhibiting differential inclusion levels with respect to *ESRP1* mRNA expression levels in ER $\alpha$ + BCs, where lower inclusion levels of these ASEs associated with high expression of *ESRP1* mRNA levels (dPSI > 0.05, p < 0.0001). Among these, an ES event of *HNRNPA2B1* exon 12 repressed by ESRP1/2 silencing showed a lower expression in highly *ESRP1* expressing as compared to lowly expressing patients (dPSI = -0.11, Wilcoxon p-value = 5.18E-16). This ES event involved an exon within the 3'UTR of the gene which is annotated to be sensitive to NMD process (McGlinicy et al. 2010). Splicing at this 3'UTR region recruits the exon junction complex to a position located about 60 nucleotides downstream of the stop codon, creating a premature context that triggers NMD (McGlinicy et al. 2010). Furthermore,

*HNRNPA2B1* was described as upregulated in tamoxifen-resistant cell lines (Klinge et al. 2019), has an oncogenic role in BC, and its high expression levels are associated with better patient prognosis (Hu et al. 2017). However, no changes were observed in the expression of *HNRNPA2B1* gene and isoforms upon silencing *ESRP1/2* in our cellular model. Other ASEs include a multiple exons skipping event in the bromodomain PHD transcription factor (*BPTF*) gene. Both exons 5 and 6 were more included under *ESRP1/2* silencing in our cellular model and showed expression difference with respect to *ESRP1* mRNA expression levels in ER $\alpha$ + BCs, and negatively correlated with *ESRP1* gene expression in these tumors. Correlation analysis of the expression of this event with molecular hallmarks revealed that high expression level in ER $\alpha$ + BCs is positively correlated with UV response\_down, with apical junction, KRAS signalling\_up, EMT, TGFB signaling, protein secretion, coagulation, angiogenesis, Notch signaling, apoptosis, hypoxia, p53, cholesterol homeostasis, estrogen response, PI3K/Akt/mTOR pathway, and glycolysis (**Supplementary Table 18f**). In contrast, a high expression of the event inversely correlates with Myc targets, G2M checkpoints, and E2F targets. Other ASEs included MYOF exon 17 which is more included upon *ESRP1/2* silencing. Correlation analysis of the event expression levels with molecular hallmarks in ER $\alpha$ + BCs revealed that a high expression is positively correlated with wnt beta catenin signaling, coagulation, with EMT, TNFa signaling via NFkB, hypoxia, with IL2/STAT5 signaling, with apoptosis and KRAS signaling. Conversely, high expression levels are inversely correlated with E2F targets, Myc targets, and oxidative phosphorylation hallmarks. MYOF exon 17 was found to be differentially expressed between luminal and TNBC subtypes, showing an increased expression level in high-metastatic TNBCs while strongly excluded in epithelial BCs (Oh et al. 2021). Furthermore, the same exon was found to be associated with increased metastatic potential of BC cell lines, where a higher inclusion level was observed in highly-metastatic BT-549 as compared to lowly-metastatic T47D BC cell lines (Oh et al. 2021).

Similarly, *ESRP1* was found to induce the expression of several exons that are associated with patients' outcomes. In particular, an ES event involving exon 5 of the splicing factor MBNL1 was strongly repressed by *ESRP1/2* silencing in our cellular model, showed a positive correlation with *ESRP1/2* expression in ER $\alpha$ + BCs (Spearman rho=0.50, p=2.41E-48; Spearman rho=0.30, p=6.38E-17 for *ESRP1* and *ESRP2*, respectively), where exon 7 was higher expressed in patients with high *ESRP1/2* expression levels as compared to those lowly expression the gene (dPSI=0.13, p=7.18E-31; dPSI=0.10, p=2.01E-13, for *ESRP1* and *ESRP2*, respectively). Indeed, MBNL1 gene isoform representing the inclusion form of the regulated exon was identified as being the most differentially included exon in cancer, both in cell lines and patients samples (Tabaglio et al. 2018). Interestingly, analysis of the exon expression levels in ER $\alpha$ + BCs and its association with molecular hallmarks revealed that high expression of the exon is inversely correlated with IL2/STAT5 signaling, apoptosis, IL6/JAK/STAT3 signaling, TNFa signaling via NFkB, KRAS signaling, PI3K/Akt/mTOR signaling, Notch signaling, hypoxia, with EMT, p53 signaling, fatty acid metabolism, cholesterol homeostasis, adipogenesis, with androgen and estrogen response pathways (**Supplementary Table 18e**). Conversely, higher inclusion levels of the exon positively correlated with proliferation related hallmark terms such as G2M checkpoints, Myc targets and E2F targets. Efforts to identify upstream regulators of

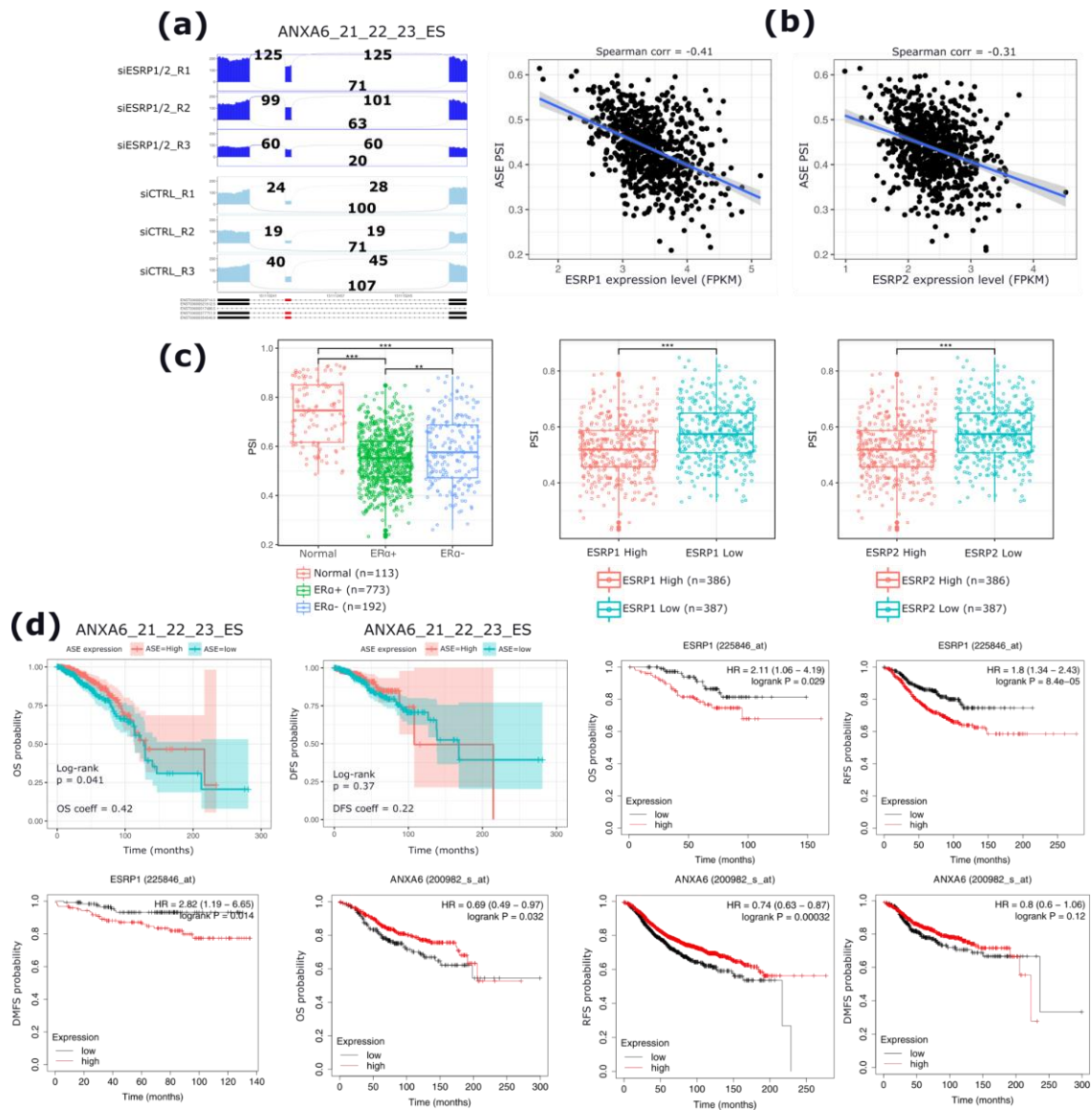
*MBNL1* exon 5 higher inclusion in cancer cell lines and patient samples revealed that the frequent downregulation of *MBNL1* in cancers is responsible for the higher inclusion levels of this exon and that *MBNL1* binding upstream of exon 5 results in its exclusion (Gates, Coonrod, and Berglund 2011). Other events included an ES of exon 6 of the *CCDC50* gene, resulting in an increased expression of the short truncated isoform and decreased expression of the full length isoform. The expression of the exon positively correlated with higher *ESRP1/2* mRNA levels in ER $\alpha$ + BCs (Spearman rho=0.35, p=5.38E-24; Spearman rho=0.35, p=4.23E-23, for *ESRP1* and *ESRP2*, respectively) and consistently showed higher inclusion levels in ER $\alpha$ + BCs highly expressing *ESRP1/2* (dPSI=0.11, p=3.46E-17, dPSI=0.10, p=2.66E-16) (**Supplementary Table 18e**). Furthermore, higher expression of the exon inversely correlated with EMT related terms, immune response and IFN signaling pathways, while positively correlated with proliferation-related hallmarks (**Supplementary Table 18f**). The exon was identified as recurrently overexpressed in multiple human solid tumors (Danan-Gotthold et al. 2015). In particular, the short *CCDC50* isoform lacking exon 6, increased in our dataset, is overexpressed in clear cell renal cell carcinoma (ccRCC), and its overexpression is associated with better survival in ccRCC patients (Sun et al. 2020). *In vivo* and *in vitro* functional experiments indicated that the overexpression of the short isoform induces proliferation, migration, invasion, and tumorigenesis of ccRCC through its interaction with the zinc finger protein *ZNF395*, while the full length isoform exerted tumor suppressor effects (Sun et al. 2020). Other reports showed that the expression of *CCDC50* short isoform is triggered by the overexpression of the splicing factor *SRSF3* by skipping of exon 6 in hepatocellular carcinoma (HCC) (Che and Fu 2020; H. Wang et al. 2019). *SRSF3* improved the stability of *CCDC50* short isoform in the cytoplasm which further increased the activation of Ras/Foxo4 signaling pathway and induced proliferation and metastasis of HCC (H. Wang et al. 2019). Conversely to ccRCC, increased mRNA levels of *CCDC50* short isoform significantly correlated with poor tumor differentiation, advanced tumor grade metastasis stage, and unfavorable prognosis (H. Wang et al. 2019). All together, these data suggest a potential role of *ESRP1* and *ESRP2* in ER $\alpha$ + BCs by controlling the AS patterns of downstream targets associated with patient outcomes.

Furthermore, we show in the present study that *ESRP1* and *ESRP2* expression in the epithelial MCF-7 BC cells is strongly dependent on hormone-independent ER $\alpha$  activity, and that their mRNA levels correlate with that of ER $\alpha$  in ER $\alpha$ + BCs. In line with these findings, the analysis of overall isoforms changes upon *ESRP1/2* silencing based on dIU analysis revealed similar results to that induced by ER $\alpha$  gene silencing. Notably, both conditions resulted in the enrichment of isoforms characterized by longer 3'UTR and 5'UTRs, by loss in expressed protein domains, more intron retention events, and high proportion of protein coding and NMD-insensitive isoforms. Moreover, both conditions resulted in a significant differential usage of alternative first and last exons. Indeed, a global protein-RNA interaction map of *ESRP1* binding based on enhanced crosslinking immunoprecipitation coupled with high throughput sequencing (eCLIP-seq) in mouse epidermis has been recently published and showed evidence for a widespread binding of *ESRP1* at 3'UTR and 5'UTR regions of target epithelial genes, further supporting its posttranscriptional functions beyond splicing regulation (Peart et al., n.d.). On the other hand, we identified 64 local ASEs commonly regulated by silencing

---

ER $\alpha$  or by that of *ESRP1* and *ESRP2* genes in our cellular model. Importantly, association analysis of the expression of the identified ASEs with respect to clinical and molecular tumor features in ER+ BCs revealed their association with molecular BC subtypes, fraction of genome altered, luminal versus non luminal subtypes, and menopause status, coherently to molecular features significantly associated with *ESRP1* expression in these tumors. Importantly, the expression levels of three ASEs including *SULF2*, *MYO6*, and *VPS39* significantly associated with *PGR* expression status in ER $\alpha$ + BCs, where higher expression of these three events was associated with higher PGR expression (**Supplementary Table 20c**). In particular, *SULF2* gene in addition to the significant exon inclusion event of exon 20 showed a significant isoform switch upon silencing *ESRP1/2*, resulting in the upregulation of the canonical full length isoform and a significant decrease in the alternative short isoform (**Supplementary Figure 17**). Although exon 20 of *SULF2* gene does not encode for a protein domain, its higher expression levels positively correlated with molecular pathways in ER $\alpha$ + BCs including IL2/STAT5 and IL6/JAK/STAT3 signaling pathways, inflammatory response and complement activation, with PI3K/Akt/mTOR, KRAS and Notch signal transduction pathways. In contrast, higher expression of exon 20 negatively correlated with estrogen response, Myc and E2F targets (**Supplementary Table 19f**). Clearly, further studies of these commonly regulated ASEs are required in order to decipher their downstream effects at the molecular level.





**Figure 2.2-13:** Top significant ASE (ANXA6 exon 22 inclusion) induced by ESRP1/2 silencing in ERα+ MCF-7 BC cell lines, its expression status in different BC subtypes and its prognostic value in ERα+ BC patients. (a) Sashimi plots representing the significant increase in the inclusion level of the exon under siESRP1/2 as compared to control condition. Total number of reads supporting either inclusion or exclusion of the exon. Isoforms representing the inclusion or exclusion form of the event are represented at the end of the Sashimi plots, with the exon 22 involved in the event highlighted in red color. (b) Correlation plots reporting the correlation between the expression of the event and that of ESRP1 and ESRP2 genes in ERα+ BCs, respectively. The correlation coefficient is indicated. (c) The expression level of the event (PSI) in Normal, ERα+ and ERα- BC samples, and relative to the expression level (High vs Low) of ESRP1 and ESRP2 genes in ERα+ BCs, respectively. (d) Associations of the expression levels of the exon inclusion event, ESRP1, and ANXA6 genes with the patient's outcomes (OS, DFS, DMFS, and RFS) in ERα+ patients, respectively.

---

## 2.3 DSCAM-AS1-driven proliferation of breast cancer cells involves regulation of alternative exon splicing and 3'UTR usage.

---

This chapter is based on the following publications:

Elhasnaoui, Jamal; Miano, Valentina; Ferrero, Giulio; Doria, Elena; Leon, Antonette E.; Fabricio, Aline S.C.; Annaratone, Laura; Castellano, Isabella; Sapino, Anna; De Bortoli, Michele. 2020. "DSCAM-AS1-Driven Proliferation of Breast Cancer Cells Involves Regulation of Alternative Exon Splicing and 3'-End Usage" *Cancers* 12, no. 6: 1453. <https://doi.org/10.3390/cancers12061453>.

---

Based on a previous publication by our group, we knew that hormone-independent activity of ER $\alpha$  in the ER $\alpha$ + MCF-7 BC cells controls the expression of not only protein coding genes but also that of noncoding RNA transcripts, among which, *DSCAM-AS1*, the most highly expressed lncRNA in our MCF-7 cells (Miano et al. 2016). Silencing ER $\alpha$  in this model strongly reduced the expression of *DSCAM-AS1* ( $\log_2FC=-0.73$ ,  $adj-p=4.78E-13$ ) (**Supplementary Figure 19** and **Supplementary Table 1a**). After having confirmed *DSCAM-AS1* expression to be luminal-specific using data of both ER $\alpha$ + BC cell lines and primary BC tissues (Miano et al. 2016), *DSCAM-AS1* was silenced in MCF-7 BC cells using two specific siRNAs, recapitulating the ER $\alpha$  silencing effects, by inducing a cell growth arrest and promoting expression of EMT markers. Therefore, we further characterized the expression of *DSCAM-AS1* in different BC microarray datasets and in-house cohorts and investigated its association with clinical and molecular features of the investigated patients. In addition, *DSCAM-AS1*-mediated proliferation of BC cells was investigated in different BC cells lines including MCF-7 cells. Furthermore, based on a previous publication, we knew that the lncRNA *DSCAM-AS1* physically interacts with the splicing factor hnRNPL in MCF-7 cells (Niknafs, Han, Ma, Speers, et al. 2016). Thus, we checked this interaction also in our MCF-7 cells investigating its functional role in these cells. To this end, a paired-end RNA-seq experiment consisting of MCF-7 cells transfected with control or *DSCAM-AS1*-targeting LNA (locked nucleic acids) was performed. The resulting RNA-seq dataset was analysed for the downstream effects of *DSCAM-AS1* knockdown at both gene and splicing levels. All the results of this chapter can be found in the aforementioned publication (Elhasnaoui et al. 2020) that is included in the appendix of this thesis.

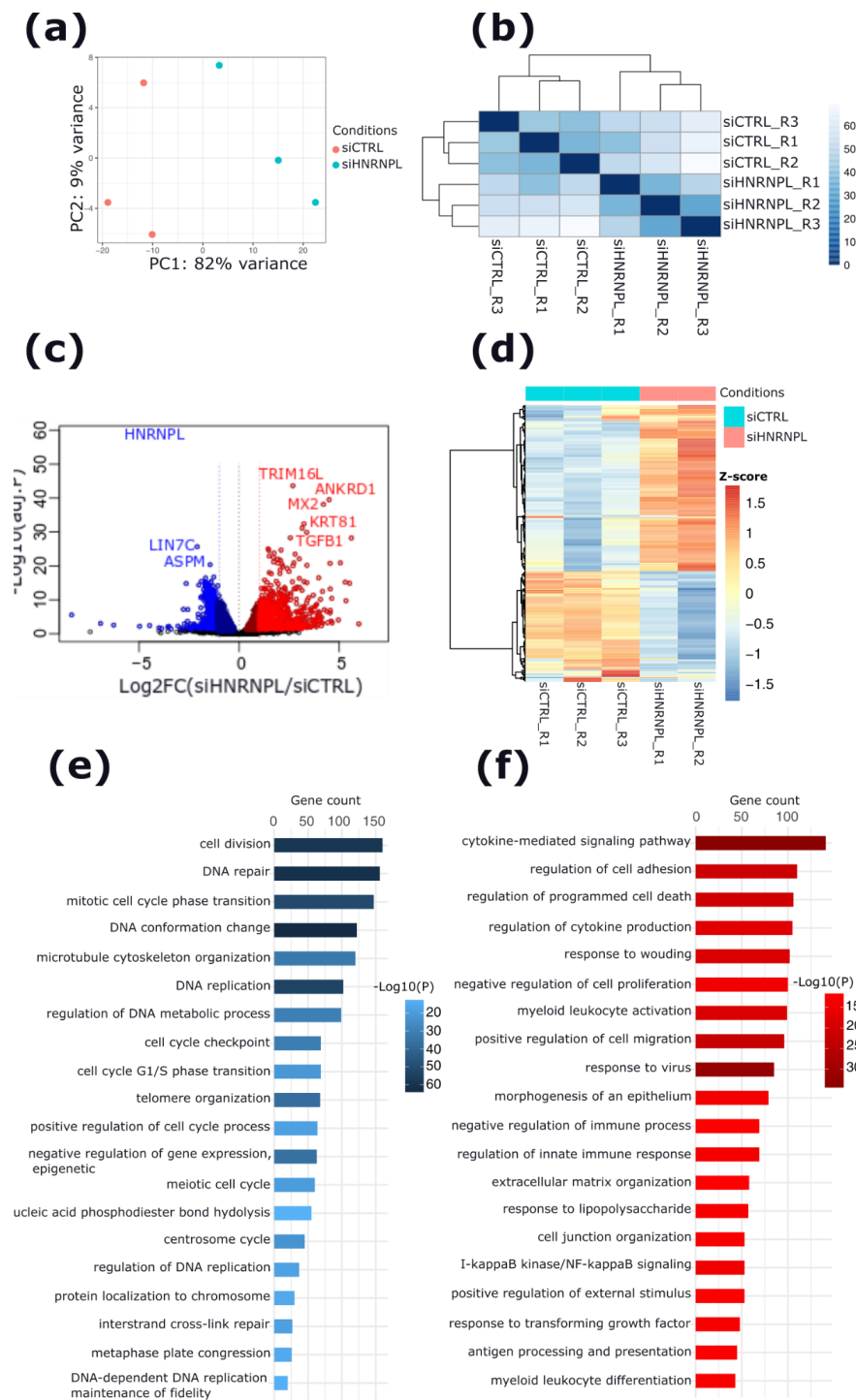
---

## 2.4 Deciphering the role of the splicing factor hnRNPL in the control of luminal epithelial BC transcriptome

The splicing factor hnRNPL is a member of the Heterogeneous nuclear ribonucleoproteins family. We and others have shown, as presented in the previous section, that hnRNPL physically interacts with the lncRNA *DSCAM-AS1* in MCF-7 BC cell lines. Therefore, to further understand the biological role of this interaction, an RNA-seq experiment consisting of transfecting MCF-7 cells with control or *HNRNPL*-targeting siRNA was performed. This dataset was analyzed for downstream effects of hnRNPL depletion at both gene and isoform levels. DE genes and isoforms and their functional pathways were identified. Isoform and splicing level changes were identified using an isoform switch analysis (dIU) and rMATS, respectively. To understand the functional role of the interaction of *DSCAM-AS1* with hnRNPL, changes common to both *HNRNPL* and *DSCAM-AS1* silencing conditions were identified and were analyzed for their downstream consequences. Therefore, the results of this chapter will be then presented into four different sections. The first section 2.4.1 is dedicated to the gene and isoform differential expression changes occurring upon *HNRNPL* silencing in MC-7 BC cells. In the second section 2.4.2 are presented the significant changes induced by hnRNPL depletion at the level of isoforms and alternative splicing (AS). The expression of identified ASEs is then evaluated in different BC tissue samples based on ER $\alpha$  positivity status and on the level of *HNRNPL* mRNA expression. The ASEs detected in BC samples are then subjected to a pathway-guided enrichment analysis to determine the molecular pathways correlated with the expression of each event. In section 2.4.3 are presented the overlapping ASEs common to *DSCAM-AS1* and *HNRNPL* silencing experiments. Finally, the findings of this chapter and their potential implication in BC progression are shortly discussed at the end of this chapter.

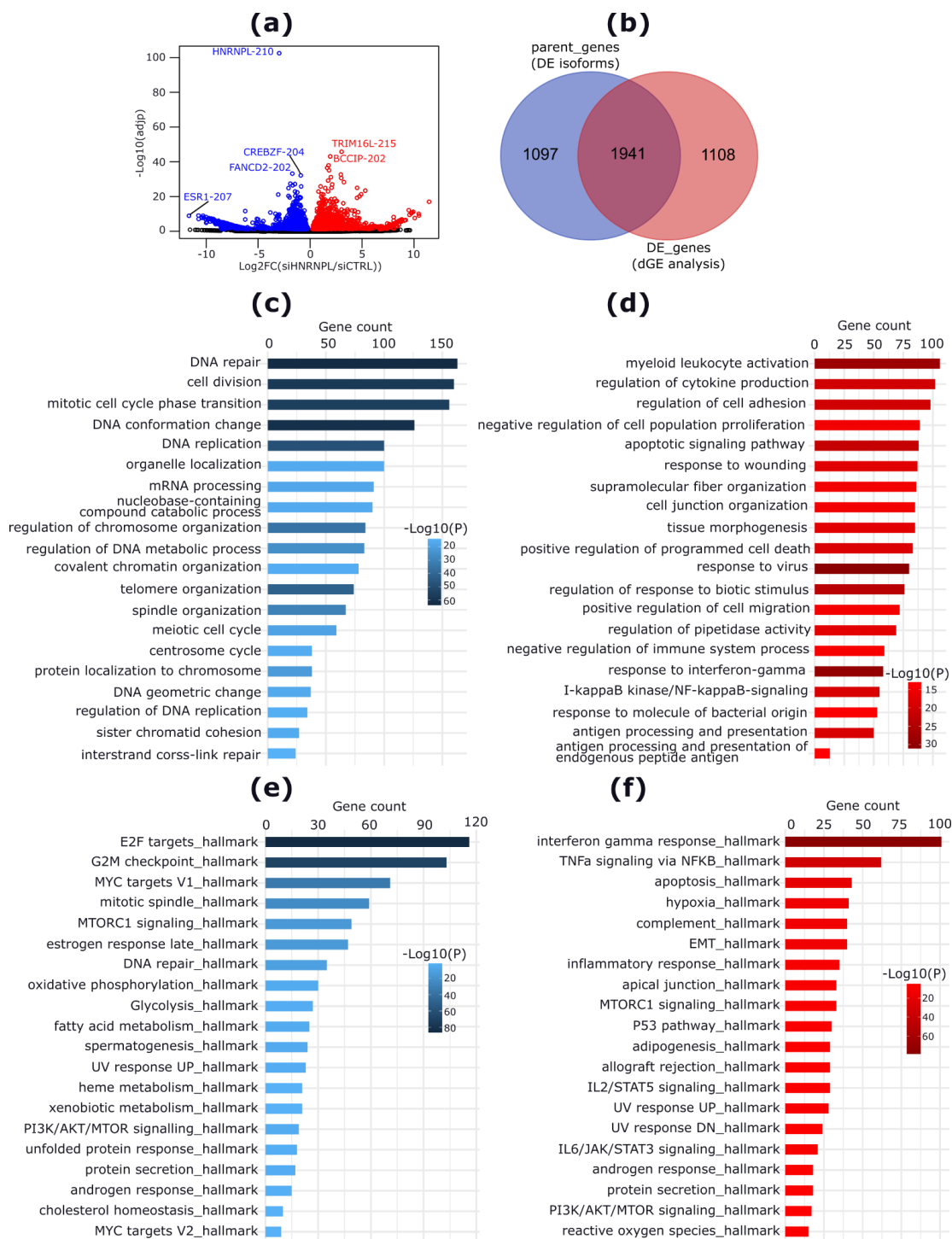
### 2.4.1 HNRNPL knock-down strongly hampers the proliferation of MCF-7 BC and induces a mesenchymal-like phenotype

To investigate the functional role of *HNRNPL* in ER $\alpha$ + BCs, we performed an RNA-seq experiment of MCF-7 cells transfected with control or *HNRNPL*-targeting siRNA (Supplementary Figure 25). The analysis of this dataset (**Figure 2.4-1a,b**) evidenced 3,050 DE genes (adj-p <0.05) upon *HNRNPL* silencing (**Figure 2.4-1a,b** and **Supplementary Table 28a**). Specifically, 1,547 genes showed a significant decrease in their expression levels, while 1,502 were upregulated (**Figure 2.4-1c,d** and **Supplementary Table 28a**). Notably, the mRNA and protein expression levels of *HNRNPL* were confirmed by qRT-PCR and western blot analysis (**Supplementary Figure 24**). Furthermore, a functional enrichment analysis of the genes DE upon *HNRNPL* silencing was performed and showed distinct biological processes related to downregulated and upregulated genes. Specifically, downregulated genes were mainly related to cell division, DNA repair, cell cycle checkpoints, and G1/S transition (**Figure 2.4-1c** and **Supplementary Table 28b**), while upregulated genes were mainly enriched in terms related to cytokine-mediated signaling pathway, programmed cell death, positive regulation of cell migration, response to wounding, morphogenesis of an epithelium, and activation of the immune response (**Figure 2.4-1d** and **Supplementary Table 28c**). In addition, a hallmarks gene sets enrichment analysis showed distinct hallmark terms enriched for downregulated and upregulated genes. In particular, repressed genes are enriched in hallmark gene sets such as E2F targets, G2M checkpoints, MYC targets V1, mitotic spindle, estrogen response late, DNA repair, MTORC1 signaling, and fatty acid metabolism. In contrast, upregulated genes were mainly enriched in hallmark gene sets such as interferon gamma response, TNF $\alpha$  signaling via NF $\kappa$ B, hypoxia, apoptosis, EMT, inflammatory response, complement, p53 pathway, IL6/JAK/STAT3 signaling, in addition to metabolism related terms such as glycolysis and cholesterol homeostasis (**Supplementary Table 28d,e**).



**Figure 2.4-1:** Gene differential expression changes induced by siRNA-mediated silencing of HNRNPL in MCF-7 cells. (a) PCA plot showing the quality of the RNA-seq replicates analysed. (b) Heat map plot representing the distance between the RNA-seq replicates. (c) volcano plot showing the  $\log_2FC$  and significance of genes identified as DE. Downregulated and upregulated genes are represented by blue and red colors, respectively. The labels of the first top DE genes are provided. (d) heat map showing the top 500 varying genes upon HNRNPL silencing. The z-score color bar represents in blue repressed ( $z\text{-score} < 0$ ) and in red induced genes ( $z\text{-score} > 0$ ), respectively. (e,f) bar plots showing the top 20 enriched GO clusters related to downregulated and upregulated genes upon HNRNPL silencing, respectively.

In addition to identifying *HNRNPL*-silencing mediated effects at gene level, isoform level changes were also identified by performing a differential isoform expression (dIE) analysis. This resulted in the identification of a total of 4,148 DE isoforms, including 2,285 downregulated and 1,863 upregulated (**Figure 2.4-2a** and **Supplementary Table 29a**). Furthermore, dIE and dGE analyses overlapped for 1,941 genes (exact hypergeometric test p-value,  $p < 0$ , representation factor=12.3) that were identified as DE by both types of analyses (**Figure 2.4-2b**). Functional enrichment analysis using gene names of the parent genes of DE isoforms revealed distinct terms differentially enriched for downregulated and upregulated isoforms. Notably, parent genes of downregulated isoforms were mainly enriched in cell cycle, cell proliferation (**Figure 2.4-2c,e** and **Supplementary Table 29b,c**), while genes of upregulated isoforms were mainly enriched in EMT process, TGFBR signaling pathway, extracellular matrix organization (**Figure 2.4-2d,f** and **Supplementary Table 29d,e**). Importantly, dIE analysis indicated isoforms that responded in opposite direction with respect to that of their parent genes. For instance, 18 isoforms were downregulated while their parent genes showed an increased expression upon *HNRNPL* silencing and 24 isoforms were identified as upregulated by dIE analysis while their parent genes were repressed by *HNRNPL* silencing (**Supplementary Table 29a**).

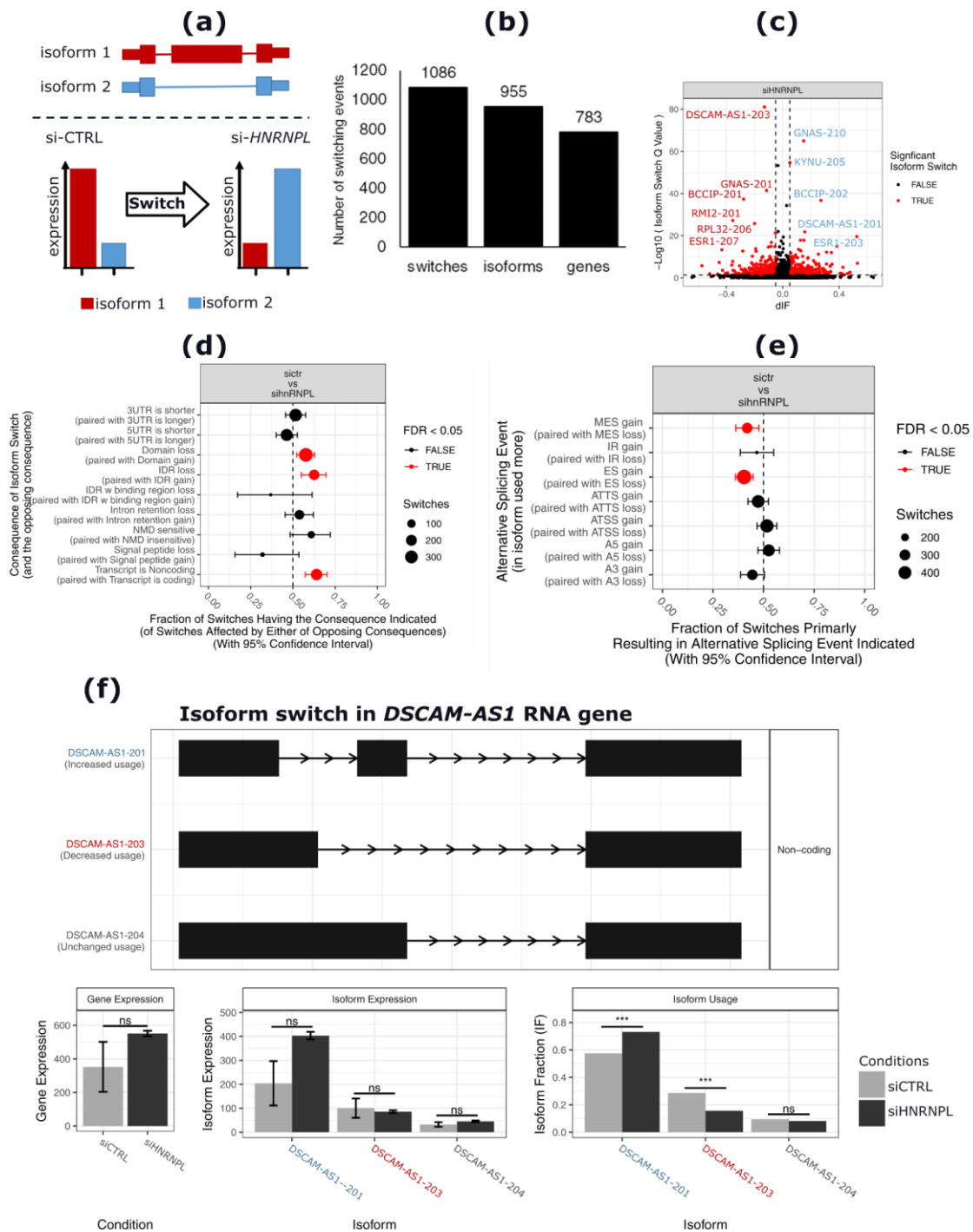


**Figure 2.4-2:** Isoform level differential expression changes induced by HNRNPL silencing. (a) Volcano plot reporting  $\text{Log}_2\text{FC}$  (x-axis) and significance (y-axis) of DE isoforms. Downregulated and upregulated gene isoforms are represented by blue and red colors, respectively. (b) Venn diagram showing the overlap between genes with at least one DE isoform and those genes reported as DE by differential gene expression analysis. (c-d) Enriched biological processes related to parent genes of downregulated (c) and upregulated (d) isoforms, respectively. (e-f) enriched Hallmark Gene Sets related to parent genes of downregulated (c) and upregulated (d) isoforms, respectively.

## 2.4.2 HNRNPL sustains cell proliferation of MCF-7 BC cells by stabilizing the expression of DSCAM-AS1 and ESR1 isoforms

In order to identify those gene isoforms responding to *HNRNPL* silencing in a different manner than the response of their parent genes, isoform differential usage (dIU) (**Figure 2.4-3a**) was applied on our RNA-seq dataset. Indeed, dIU analysis identified a total of 1,086 isoform pairs of 783 genes responding differently to *HNRNPL* silencing in MCF-7 cells (**Figure 2.4-3b,c** and **Supplementary Table 30a**). Genes harbouring switching isoform pairs included 80% protein coding and 20% noncoding genes (**Supplementary Table 30a**). Among these, *DSCAM-AS1* RNA transcripts were among the top significant switching genes, where the DSCAM-AS1-203 isoform is repressed while DSCAM-AS1-201 isoform is induced. Similarly, the *ESR1* gene ranked among the top 15 significant switching protein coding genes, with the canonical full length ESR1-207 isoform induced and the short ESR1-203 isoform is repressed (**Figure 2.4-3c** and **Supplementary Table 30a**). In case of *DSCAM-AS1*, the switching isoform pairs differ by the inclusion/exclusion of an alternative exon and an alternative 5' splice site (**Figure 2.4-3f**), while in case of *ESR1* gene the switching isoform pairs differ by the length of their open reading frames, which resulted in the repression of the full length isoform (6,466 bp, 595 amino acids) and induction of the shortest, 3'truncated isoform (624 bp, 84 amino acids) (**Supplementary Figure 26**). Functional enrichment analysis of genes harbouring an isoform switching event revealed an enrichment of biological process terms such as covalent chromatin organization, negative regulation of cell cycle, DNA replication, cell cycle checkpoints, and autophagy (**Supplementary Table 30b**). Furthermore, hallmark gene sets enrichment analysis revealed enriched terms such as estrogen response (top significant, 29 genes), apoptosis (22 genes), E2F targets (22 genes), and DNA repair (15 genes) (**Supplementary Table 30c**). On the other hand, the identified switching events resulted in the enrichment of some isoform features in the silencing condition as compared to the control condition (**Figure 2.4-3d** and **Supplementary Table 30d**) such as the number of expressed protein domains which were less in silencing condition (262 losses vs 205 gains, proportion q-value=0.009), and number of protein coding isoforms which were less expressed than protein coding isoforms in silencing condition as compared to control condition (163 noncoding vs 124, proportion q-value=0.024). In contrast, other isoform features such as the 3'UTR (190 shorter vs 173 longer, proportion q-value=0.40) and 5'UTR length (149 shorter vs 186 longer, proportion q-value=0.05), NMD-sensitivity (46 NMD-sensitive vs 39 NMD-insensitive, proportion q-value=0.51), and IR events (129 IR losses vs 108 IR gains, proportion q-value=0.19) were not significantly different among compared conditions (**Figure 2.4-3d**). The comparison of both conditions for differences in AS predicted some ASE types to be enriched upon *HNRNPL* silencing such as single ES (232 skipped vs 270 included, proportion q-value=0.01) and multiple exons skipping events (12 skipped vs 30 included, proportion q-value=0.0003) (**Figure 2.4-3e** and **Supplementary Table 27e**). The proportions of isoforms resulting from other opposing types of ASEs were not significantly different between compared conditions.



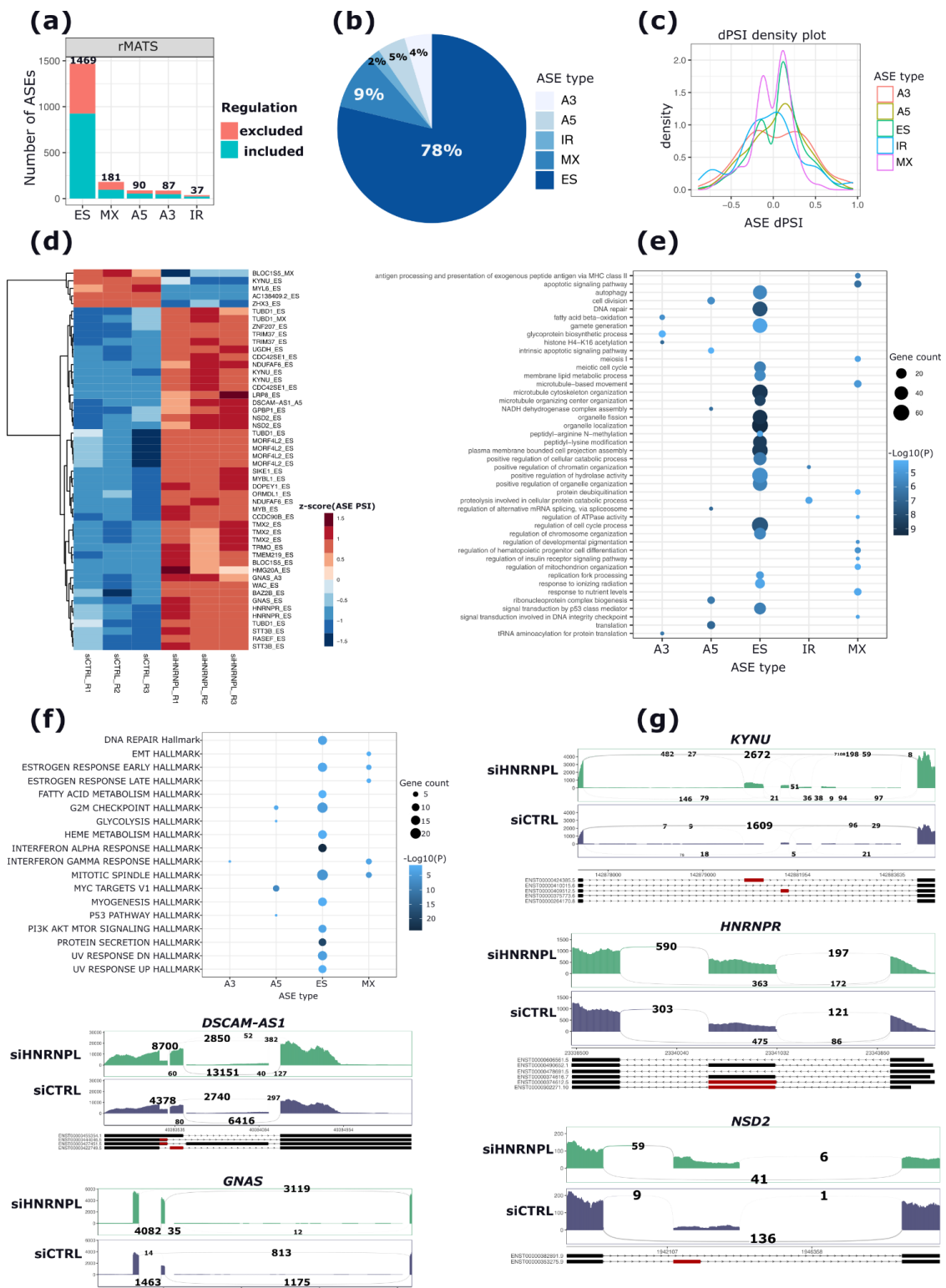


**Figure 2.4-3: Differential isoform usage (dIU) and their potential downstream consequences induced by HNRNPL silencing in MCF-7 BC cells.** (a) Schematic representation of the principle behind dIU analysis. (b) Number of significant switching events and isoforms/genes involved. (c) Volcano plot reporting differential isoform fractions (dIF) on x-axis and their significance  $-\log_{10}(q\text{-value})$  on y-axis. (d) functional consequences of identified isoform switching pairs. Fractions of isoforms with either the indicated or opposing consequence are statistically compared among conditions. Dot size is proportional to the number of switches having the indicated or opposing consequence. Significantly different isoform fractions are highlighted in red. (e) AS enrichment analysis among compared conditions indicating differences in the fraction of isoforms resulting from either the indicated or opposing ASE. Dot size is proportional to the number of switches having the indicated or opposing consequence. Significantly different isoform fractions are highlighted in red. (f) Switch plot showing the switching isoform pairs of *DSCAM-AS1* RNA gene. Gene isoforms are represented with their

exons as black blocks and introns as solid lines. Histograms report the expression level of the gene and isoforms indicated as well as their differential usage, respectively. Statistically significant differences are indicated. ns, nonsignificant; \*\*\*,  $p < 0.00001$ .

### 2.4.3 Analysis of alternative splicing changes induced by *HNRNPL* silencing reveals important functions of hnRNPL in MCF-7 BC cells

In order to identify hnRNPL-regulated exons, a differential AS analysis was performed using rMATS (Shen et al. 2014, 2012) and Whippet (Sterne-Weiler et al. 2018). Since rMATS is the most well documented and most stable algorithm, only rMATS results will be discussed. Analysis of AS changes induced by *HNRNPL* silencing in MCF-7 BC cells revealed a high number of ASEs differentially spliced among compared conditions. Notably, 1,864 ASEs ( $|\text{dPSI}| > 0.05$ ,  $\text{adj}p < 0.05$ ) in 1,464 genes classified as 1,469 (78%) ES in 1,244 genes, 181 (9%) MX in 189 genes, 90 A5 in 84 genes, 87 A3 in 83 genes, and finally 37 IR in 35 genes (**Figure 2.4-4a,b** and **Supplementary Table 31a-e**). Noteworthy, *HNRNPL* silencing induced more inclusion than exclusion ASEs (e.g. 926 EI vs 543 ES) in almost all the ASE types (**Figure 2.4-4c** and **Supplementary Table 31a**), suggesting possible defects in the splice sites recognition by the spliceosomal machinery and line with the inhibitory effect of hnRNPL on AS process (Motta-Mena, Heyd, and Lynch 2010; Heiner et al. 2010). The top 50 significant ASEs occurred at genes involved in BC cell proliferation and survival, including *DSCAM-AS1* (Elhasnaoui et al. 2020), *NSD2*, *TRIM37*, and *MYB* involved in histone modification process (Kuo et al. 2011; Bhatnagar et al. 2014; Audia and Campbell 2016), *GNAS* involved in macromolecules methylation and BC cell proliferation and migration (Jin et al. 2019), *HMG20A* involved in peptidyl-lysine modification, and the ER $\alpha$ -target gene Tubulin delta 1 (*TUBD1*) involved in mitotic cell cycle progression (Hua et al. 2008) (**Figure 2.5-4d** and **Supplementary Table 31a**). Importantly, a differential enrichment of GO biological processes for spliced genes depending on the ASE types was observed (**Figure 2.4-4e** and **Supplementary Table 31f**). ES harbouring genes were mainly enriched in cell division and chromosome maintenance, DNA repair, and microtubule cytoskeleton organization. MX harbouring genes on the other hand were enriched in terms such as apoptotic signaling pathway, antigen processing and presentation, and microtubule-based cell movement. A5 harbouring genes were mainly enriched in translation, translational elongation, histone modification, and ribonucleoprotein complex biogenesis. A3 harbouring genes were enriched in histone H4-K16 acetylation, tRNA aminoacylation for protein translation, and fatty acid beta-oxidation (**Figure 2.4-4e** and **Supplementary Table 31f**). Similarly, a Hallmark Gene Sets enrichment analysis revealed distinct terms enriched for spliced genes depending on ASE type, with ES harbouring genes predominantly involved in cell cycle related terms, estrogen response, in interferon signaling pathway and metabolism (**Figure 2.4-4f** and **Supplementary Table 31f**). The top five significant ASEs induced by *HNRNPL* silencing in our cellular model includes an A5 and ES events in *DSCAM-AS1* RNA gene that were more included upon *HNRNPL* silencing, an EI event in *GNAS*, MX inclusion event in *KYNU*, and EI in each of *HNRNPR* and *NSD2*, as reported in **Figure 2.4-4g**.



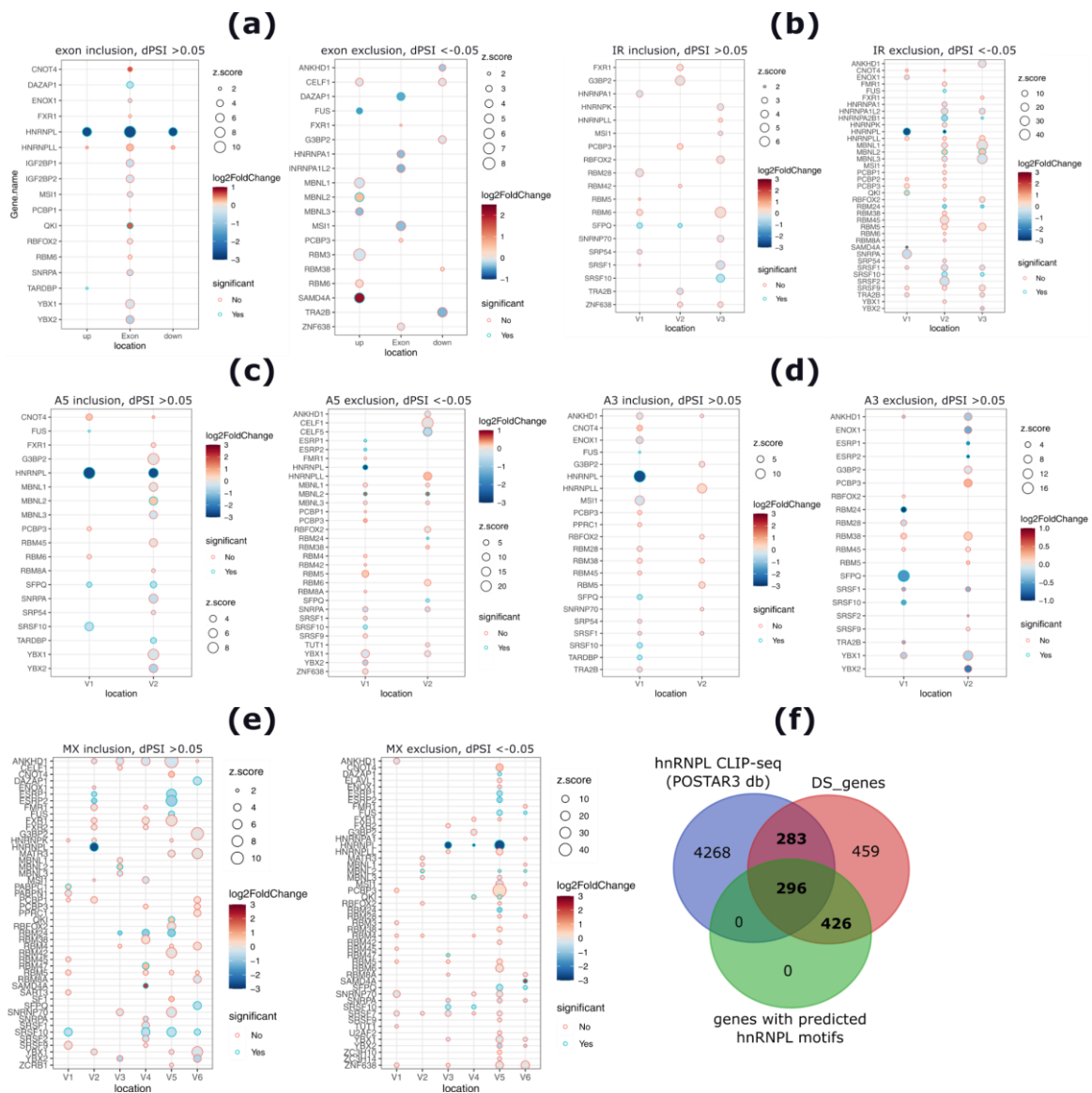
**Figure 2.4-4:** Differential AS changes induced by HNRNPL silencing in MCF-7 BC cells. (a) Stacked barplot reporting the number of significantly changing AEs upon HNRNPL silencing classified on ASE type and type of regulation (e.g. inclusion, exclusion). (b) Pie-chart showing the relative proportion of ASE types, with ES events as most (78%) represented AEs. (c) Density plot showing the distribution of the differential inclusion levels (dPSI) of different ASE types. (d) Heatmap plot showing the relative inclusion level (PSI) of the top 50 significant AEs in the RNA-seq replicate samples. (e-f) Dot plots showing the biological process and Hallmark Gene Sets terms enriched for genes harbouring a significant ASE classified by ASE type, respectively. (g) Sashimi plots of selected AEs showing a significant differential

inclusion level upon *HNRNPL* silencing in MCF-7 BC cells. Gene isoforms involved in each ASE are plotted below the plots, with exons involved in the ASE are highlighted in red. The number of normalized read counts supporting either inclusion or exclusion of the indicated exon are reported.

### 2.4.3.1 hnRNPL binding motifs are enriched around the identified ASE in our cellular model

We sought next to determine whether hnRNPL is primarily involved in the observed AS patterns changes. Thus, an RBP-binding motif enrichment analysis was performed involving the genomic regions involved in each ASE. The identified ASEs were classified based on their types and direction of regulation (e.g. inclusion/exclusion). This analysis, as expected, revealed hnRNPL binding motifs to be present in 965 (52%) ASEs in 723 (49%) genes, classified as 747 ES (198 included, 549 excluded), 118 MX (70 included, 48 excluded), 34 A5 (21 included, 13 excluded), 30 A3 (18 included, 12 excluded), and 36 (21 included, 15 excluded) IR events (**Supplementary Table 32a**).

In addition, an RBP binding motif enrichment analysis was performed with respect to a control set of expressed exons not significantly regulated by *HNRNPL* silencing (e.g.  $|\text{dPSI}| < 0.01$ ,  $\text{adj}p > 0.5$ ) revealed significant enrichment of different RBP sets enriched for different ASEs based on their direction of regulation. In particular, in case of cassette exons, hnRNPL binding motifs were predicted to be enriched only exclusively in case of exon inclusion events, with the highest enrichment score associated with intronic binding (**Figure 2.4-5a** and **Supplementary Table 32b**). In addition to hnRNPL binding motifs, 16 other RBPs, three of which (*QKI*, *DAZAPI*, *TARDBP*) are DE upon *HNRNPL* silencing, were enriched in case of exon inclusion events, showing enriched motifs most frequently in exons (**Figure 2.4-5a** and **Supplementary Table 32b**), suggesting these RBPs to enhance the inclusion of bound exons, as previously suggested (Sebestyén et al. 2016). In contrast, 19 RBPs motifs, four of which (*MBNL2*, *FUS*, *DAZAPI*, *SAMD4A*) are DE upon *HNRNPL* silencing, were predicted to be enriched in case of exon exclusion events, with motifs are most frequently enriched in upstream intronic and in exonic regions (**Figure 2.4-5a** and **Supplementary Table 32b**), consistent with a possible positional effect (Sebestyén et al. 2016). Furthermore, hnRNPL binding motifs were enriched in case of IR exclusion, A5 inclusion, A3 inclusion, and MX events (**Figure 2.4-5b-e** and **Supplementary Table 32c-f**).



**Figure 2.4-5:** Dot plots representing predicted RBP binding motifs (y-axis) to be enriched and their locations (x-axis) in the identified ASEs upon HNRNPL silencing, including (a) ES/EI, (b) IR, (c) A5, (d) A3, and (e) MX inclusion/exclusion events, respectively. Dot size represents the enrichment z-score. Color bar plots represent the log<sub>2</sub>FC of enriched RBPs upon HNRNPL silencing, and dot borders represent the significance of gene level DE status. In case of IR events, V1-V3 represent upstream, within intron, and downstream of regulated introns, respectively. In case of A5/A3 events, V1-V2 represent upstream and downstream regulated splice sites, respectively. In case of MX events, V1 to V6 represent the mutual exclusive exons (V2, V5) and their upstream (V1, V4) and downstream (V3, V6) flanking 100 nucleotides intronic sequences. (f) Venn diagram plot reporting the overlap between genes with CLIP-seq binding site and those regulated in our datasets.

To support our RBP binding motif prediction analysis, predicted hnRNPL binding motifs were overlapped with hnRNPL binding sites reported in two different CLIP-seq experiments, in the epithelial HepG2 and K562 cell lines retrieved from the POSTAR2 database (Zhu et al. 2019; Fei et al. 2017). This database in its newly released version POSTAR3 database includes 2,075 public CLIP-seq datasets from six different species (Zhu et al. 2019). On one hand, this analysis revealed 580 genes identified as differentially spliced in our dataset also harboured an hnRNPL CLIP-seq binding site, of which 296 harboured a predicted binding motif for

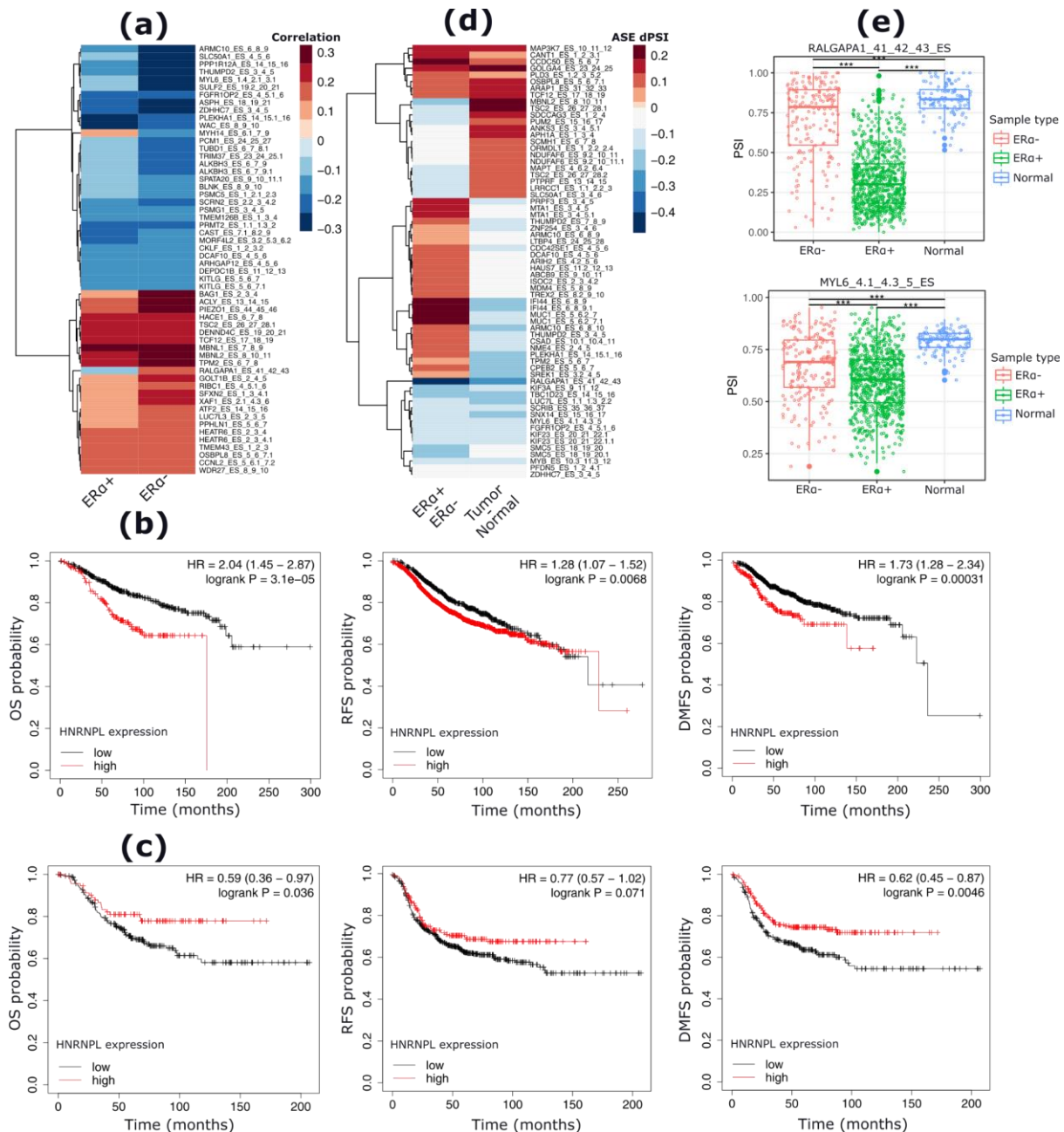
hnRNPL (**Figure 2.4-5f** and **Supplementary Table 32g**), further supporting our predictions. On the other hand, 283 differentially spliced genes showed an hnRNPL CLIP-seq binding site but did not show any predicted hnRNPL binding motif. This may be a result of the narrow sequences (event + 200 nucleotides length) scanned for RBP binding motifs in our analysis, and this set of genes could have an hnRNPL binding motif outside of the scanned sequence interval ( $ASE \pm 200$  nt). Noteworthy, two of our candidate genes, *ESRI* and *ESRPI*, that were both downregulated and differentially spliced upon *HNRNPL* silencing showed both a predicted and CLIP-seq hnRNPL binding sites. On the other hand, although no hnRNPL binding sites on *DSCAM-ASI* gene body were observed from CLIP-seq, our RBP binding motif predictions indicated three binding sites for hnRNPL on *DSCAM-ASI*, in line with previously reported results (Niknafs, Han, Ma, Speers, et al. 2016; Gawronski et al. 2018). Importantly, *DSCAM-ASI* alternative splice site donor event (A5, dPSI=0.187, adjp=0) contained predicted binding motifs for 27 splicing factors of which eight (U2AF2, HNRNPC, PTBP1, PTBP2, MATR3, RALY, RBM5, and RBM6) are bound by hnRNPL based on both hnRNPL CLIP-seq (Zhu et al. 2019) and RIP experiments (Fei et al. 2017), suggesting that the spliced *DSCAM-ASI* region upon *HNRNPL* silencing could play a role in mediating hnRNPL interactions with other factors and/or target pre-mRNAs.

#### 2.4.3.2 hnRNPL regulated events are differentially expressed among BC subtypes and are related to disease prognosis

*HNRNPL* silencing in our MCF-7 cells induced 1,864 ASEs resulting in the alteration of splicing patterns of 1,464 genes. To understand whether the expression of the identified ASEs depends on *HNRNPL* mRNA levels in ER $\alpha$ + BC samples, the expression levels (e.g. PSI) of ASEs in BC samples was retrieved from TCGASpliceSeq (M. Ryan et al. 2016) together with *HNRNPL* mRNA expression levels from TCGA BRCA database. This analysis resulted in the retrieval of 585 ASEs that were detectable in 1,207 BC samples classified as 773 ER $\alpha$ +, 193 ER $\alpha$ - and 113 normal tissues (**Supplementary Table 33a**). Among retrieved ASEs, 200 ASEs correlated (Average Spearman coefficient = -0.11,  $p < 0.05$ ) with *HNRNPL* mRNA levels in ER $\alpha$ + BC samples. In particular, 90 ASEs positively correlated with *HNRNPL* mRNA levels while 110 ASEs inversely correlated with *HNRNPL* mRNA levels in ER $\alpha$ + samples ( $n=773$ ) (**Supplementary Table 33b**). On the other hand, 155 ASEs were significantly correlated (Average Spearman coefficient =,  $p < 0.05$ ) with *HNRNPL* mRNA levels in ER $\alpha$ - samples ( $n=193$ ), including 59 positively and 96 negatively correlated, respectively (**Supplementary Table 33c**). Importantly, different subsets of ASEs correlated with *HNRNPL* gene expression in the analysed samples in an ER $\alpha$  status-dependent manner. In particular, while 68 ASEs correlated with *HNRNPL* mRNA levels in both ER $\alpha$ + and ER $\alpha$ - BC samples (**Figure 2.4-6a**), 132 ASEs exclusively correlated with *HNRNPL* mRNA levels in ER $\alpha$ + samples, and 87 ASEs exclusively correlated with the expression of the gene in ER $\alpha$ - BC samples (Supplementary Table 33b,c), suggesting that hnRNPL may function in a BC subtype-dependent manner and its function could be regulated by factors exhibiting a BC subtype-specific expression, among which, its partner, the luminal-specific lncRNA *DSCAM-ASI*.

Furthermore, this observation is in line with the different prognostic values of *HNRNPL* in BC samples depending on ER $\alpha$  gene expression status.

Notably, the analysis of the association between *HNRNPL* mRNA expression levels and patients clinical features indicates that *HNRNPL* overexpression associates with worse patients outcomes, including short overall survival (OS), worse relapse-free survival (RFS), and worse distant-metastasis-free survival (DMFS) in ER $\alpha$ + BC patients, while higher *HNRNPL* mRNA levels associate with better outcomes in ER $\alpha$ - patients, including longer OS, and better RFS and DMFS (**Figure 2.4-6b,c**). In addition, no significant differences in ASEs inclusion/exclusion levels between highly *HNRNPL* expressing and lowly *HNRNPL* expressing ER $\alpha$ + patients, classified based on the median *HNRNPL* gene expression level (**Supplementary Table 33d**). In contrast, significant differences ( $|dPSI| > 0.05$ ; Wilcoxon test p-value  $< 0.0001$ ) in the inclusion/exclusion levels of 108 ASEs (57 excluded, 51 included) were observed between ER $\alpha$ + and ER $\alpha$ - BC samples (**Supplementary Table 33d**). The most significant event between these two groups of samples was an ES event at the *RALGAP1*, showing a higher inclusion level in ER $\alpha$ - as compared to ER $\alpha$ + BC samples (dPSI = -0.48; Wilcoxon p-value = 1.49E-58). Noteworthy, this ES event showed a different correlation with *HNRNPL* expression depending on ER $\alpha$  status. Notably, The ES event at *RALGAP1* gene is positively correlated (Spearman rho = 0.16, p = 0.014) with *HNRNPL* expression in ER $\alpha$ -, while is inversely correlated (Spearman rho = -0.10, p = 0.004) with *HNRNPL* expression in ER $\alpha$ + tumors (**Figure 2.4-6a**). A similar event with inverted correlation coefficients relatively to ER $\alpha$  status is an ES event at *MYH14* gene, showing a positive correlation (Spearman rho = 0.11, p = 0.004) in ER $\alpha$ + tumors while an inverse correlation (Spearman rho = -0.15, p = 0.039) with *HNRNPL* mRNA levels in ER $\alpha$ - tumors (Figure 2.5-6a). The second top significant ASE is an ES event at the *IFT172* gene, showing a higher inclusion level in ER $\alpha$ + as compared to ER $\alpha$ - BC samples (dPSI = 0.20, Wilcoxon p = 1.14E-47). In ER $\alpha$ + BC samples, no differences in the inclusion/exclusion levels of the detectable ASEs were observed based on the expression level of ER $\alpha$  gene (e.g. *ESR1* high vs *ESR1* low), nor based on *DSCAM-AS1* gene expression levels (e.g. *DSCAM-AS1* high vs *DSCAM-AS1* low). In addition, 138 ASEs (77 excluded, 61 included) were significantly differentially expressed between tumor and normal breast tissue samples (Supplementary Table 33d), with the most significant ASE being an ES at *PRMT2* gene, showing a higher inclusion level normal as compared to tumors samples (dPSI = -0.32, Wilcoxon p = 5.48E-56) (**Supplementary Table 33d**). The ASEs with significant differences in inclusion/exclusion levels between both ER $\alpha$ + and ER $\alpha$ -, and tumors compared to normal samples are shown in **Figure 2.4-6d**, of which two example ASEs, at *RALGAP1* and *MYL6* are shown in **Figure 2.4-6e**.



**Figure 2.4-6:** (a) Heat map plot showing the Spearman correlation coefficient between *hnRNPL*-regulated ASES and *HNRNPL* mRNA levels in BC samples with respect to ERα status (e.g. ERα+ and ERα-). Only ASES significantly correlated with *HNRNPL* gene expression levels in both sample types are reported in the plot. (b,c) Kaplan-Meier plots reporting the association between *HNRNPL* gene expression levels and clinical outcomes in ERα+ (b) and ERα- (c) patients, respectively. HR, Hazard Ratio, significance, log rank test p-value. OS, Overall Survival; RFS, Relapse-Free Survival; DMFS, Distant-Metastasis-Free Survival. The analysis was performed with the 35201\_@microarray probe set. (d) difference in the inclusion levels of *hnRNPL*-regulated ASES between (i) ERα+ and ERα- and (ii) tumors versus normal breast tissues. Only significantly different ASES in both comparisons are reported. (e) Box plots reporting the inclusion levels (PSI) of two ASES in ERα-, ERα+ and normal breast tissues, respectively. \*\*\*, p < 0.0001.

In order to understand the molecular oncogenic pathways associated with the ASES detectable in tumor samples from TCGASpliceSeq database, a correlation analysis of each ASE with 50 molecular pathways in these samples was performed using the PEGASAS (Phillips et al. 2020). This analysis revealed two main

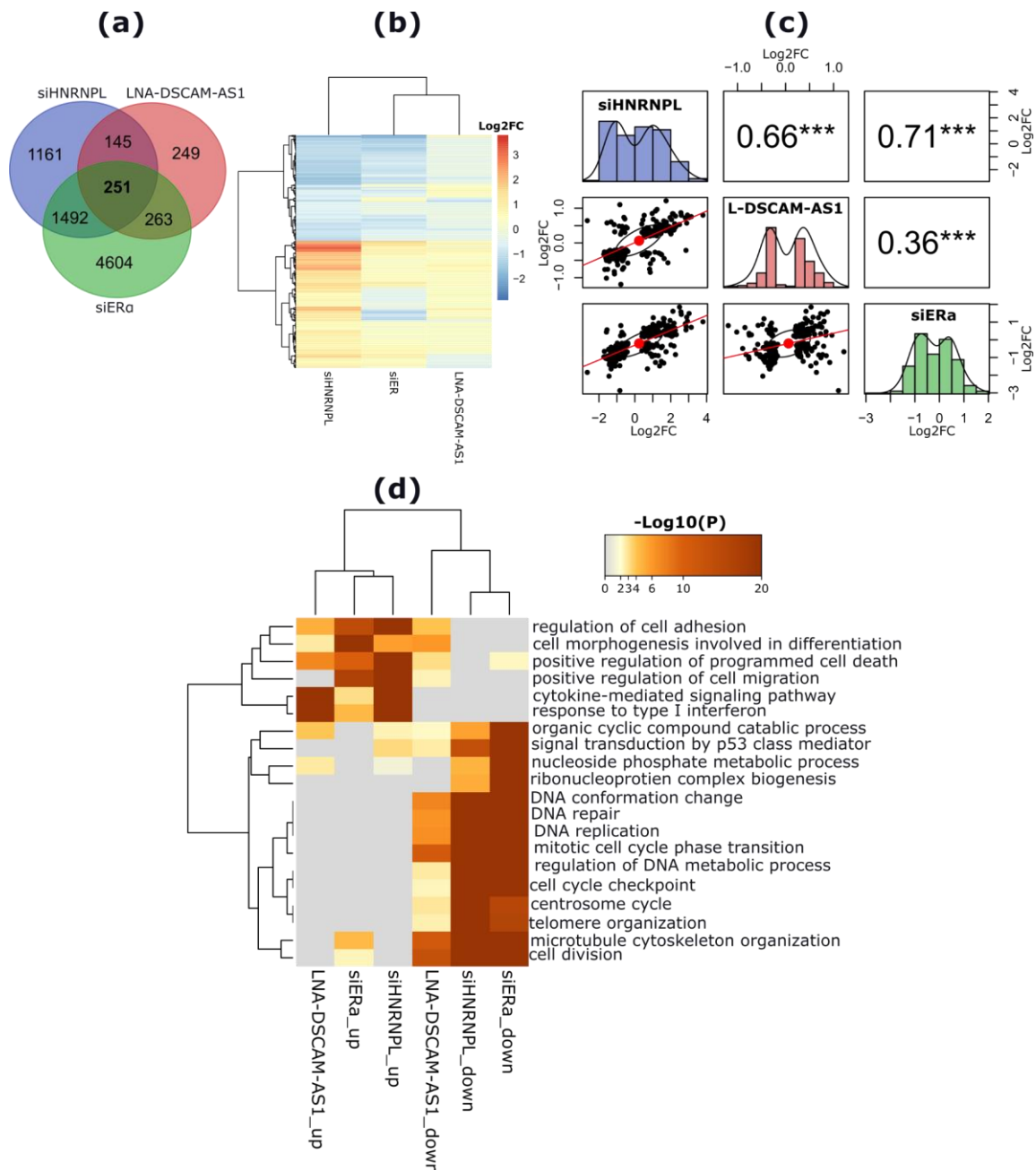


molecular pathway clusters differentially correlated with the analysed ASEs. In particular, the first main cluster contained molecular pathways such as EMT, TGFB signaling, P53 pathway, Hypoxia, estrogen response, cholesterol homeostasis, and signal transduction pathways such as KRAS and PI3K/Akt/mTOR signaling pathways. This cluster also contained metabolism-related terms such as adipogenesis, fatty acid metabolism, and oxidative phosphorylation. The second cluster contained terms related to cell cycle progression such as E2F targets, G2M checkpoint pathway, Myc targets, mitotic spindle, DNA repair (**Supplementary Table 33d**). Moreover, unsupervised hierarchical clustering of ASEs based on their PSI values and correlation coefficients revealed three distinct clusters differentially correlated with the two molecular pathways clusters. For instance, a cluster formed by 12 ES events at (*SULF2*, *CD47*, *GFR1OP2*, *TBC1D23*, *PRMT2*, *TMEM126B*, *MRPL48*, *PLEKHAI*, *PICALM*, and *TMEM126B*) inversely correlated with terms of the second cluster (e.g. cell cycle related pathways) while they positively correlated with molecular pathways of the first cluster (e.g. p53 pathway, Hypoxia, EMT, TGFB signaling, KRAS signaling, estrogen and androgen response pathways). A second cluster formed by 11 AS events at (*TPM2*, *OSBPL8*, *TCF12*, *GOLGA4*, *DENND4C*, *MBNLI*, *CCDC50*, and *MAP3K7*) positively correlated with cell cycle related pathways of the second cluster (e.g. E2F targets, Myc targets, mitotic spindle, DNA repair) and they showed an exclusive negative correlation with the molecular pathways of the first cluster containing EMT and signal transduction related pathways (**Supplementary Table 33d**).

To verify whether the hnRNPL-regulated ASEs are associated with prognosis in BC, we sought to evaluate the association between the expression of the detected ASEs and patients overall (OS) and disease-free survivals (DFS) focusing on ER $\alpha$ + subjects. Thus, clinical information including OS and DFS of 773 ER $\alpha$ + patients were retrieved from the TCGA GDC data portal (Grossman et al. 2016b). BC samples were classified into highly or lowly expressing the ASE based on the median inclusion level of the event. This analysis revealed a total of 44 ASEs including 25 and 19 ASEs significantly associated with OS and DFS times, respectively, of which 2 ASEs (CKLF\_ES\_1\_2\_4 and XAF1\_ES\_6\_7\_8) significantly associated with both (**Supplementary Table 33e**). In particular, the ES event at *MYL6* gene (*MYL6\_ES\_1.4\_2.1\_3.1*) appeared as the top significant ASE associated with longer OS time in the analysed samples (OS coefficient=22.05, log rank p-value =2.56E-4). This event ranked as the top significant event being excluded (dPSI =-0.116, adjp=0) by *HNRNPL* gene silencing in MCF-7 cells (**Supplementary Table 31a**), inversely correlated (Spearman rho=-0.31, p=1.27E-06) with *HNRNPL* gene expression levels in ER $\alpha$ - samples, and exhibited a significantly higher inclusion levels in ER $\alpha$ + compared to ER $\alpha$ - BC samples (dPSI=0.14, Wilcoxon p=2.75E-39). On the other hand, an ES event at the *RBM4* splicing factor gene was significantly associated with a worse DFS (DFS coefficient=-21.77, log rank p-value=0.013) in the analysed samples. This event is induced (dPSI=0.162, adjp=0.007) by *HNRNPL* depletion in MCF-7 cells.

#### 2.4.4 *hnRNPL*, *ERα*, and *DSCAM-AS1* are three factors synergistically acting on proliferation in *ERα*+ BCs

At the gene level, *HNRNPL* silencing strongly hampered the expression of cell cycle related genes and induced a significant decrease in the proliferative potential of MCF-7 cells (**Supplementary Figure 27**). Interestingly, *HNRNPL* silencing induced a significant decrease in *ERα* gene expression ( $\log_2FC = -0.97$ ,  $adjp = 1.38E-8$ ) (**Supplementary Figure 26**) in MCF-7 BC cells, which may explain the observed reduction in cell proliferation. The comparison of the lists of genes showing significant expression changes upon *HNRNPL* or *ERα* silencing experiments revealed a high number (1,743) of genes commonly regulated (representation factor = 1.5, exact hypergeometric test p-value,  $p < 5.86E-130$ ) in both, of which 884 downregulated and 639 upregulated in both datasets conditions (**Figure 2.4-7a-c and Supplementary Table 34a**), indicating that both *HNRNPL* and *ERα* genes are synergistically acting on the same pathway in these cells. These commonly regulated genes include the two epithelial splicing factors *ESRP1* and *ESRP2*, the two direct target genes of *ERα* in MCF-7 cells. Surprisingly, although *ERα* gene expression levels decreased upon *HNRNPL* silencing, the expression of its downstream target lncRNA, *DSCAM-AS1* tended to increase upon *HNRNPL* silencing (**Figure 2.4-3f**). On the other hand, a significant number (395) of genes were commonly regulated (representation factor = 2.5; exact hypergeometric test p-value,  $p < 2.71E-82$ ) in both *HNRNPL* and *DSCAM-AS1* silencing experiments, 338 (86%) of which showed a coherent regulation (**Figure 2.4-7a-c and Supplementary Table 34b**). As shown in **Figure 2.4-7d**, GO enrichment analysis of DE genes in the three experimental conditions indicates two separate clusters of biological processes, one of which containing cell cycle related terms such as mitotic cell cycle phase transition, DNA conformation change, DNA repair, DNA replication, and cell cycle checkpoints, mainly related to downregulated genes in the three datasets. The second cluster contained biological processes related to cell movement such as cell adhesion, cell morphogenesis involved in differentiation, positive regulation of cell migration, regulation of cell death, and cytokine-mediated signaling pathway was exclusively related to upregulated genes in the three conditions (**Figure 2.4-7d and Supplementary Table 34**). All together, these data suggest that *HNRNPL*, *ERα* and *DSCAM-AS1* are three factors synergistically acting on the transcriptional regulation of target genes. Quite surprisingly, *HNRNPL* silencing in MCF-7 cells induced an isoform-level regulation of *DSCAM-AS1* and an increase in the overall gene expression (**Figure 2.4-4g**) although *ERα* expression was decreased.



**Figure 2.4-7:** Transcriptional events commonly regulated between HNRNPL, ER $\alpha$  and DSCAM-AS1 silencing experiments. (a) Venn diagram showing the number of overlapping genes differentially expressed among the three silencing conditions. (b) Heatmap plot showing the  $\log_2FC$  in the silencing experiments of overlapping genes shown in (a). (c) Pairwise correlation plots showing the correlation between the three silencing experiments. Histograms report the distribution of  $\log_2FC$  in each of the conditions. Dot plots represent the  $\log_2FC$  in each condition. Red lines represent the correlation between each pair of experiments. Pairwise Pearson correlation coefficients are reported. \*\*\*,  $p < 0.0001$ . (d) Heatmap plot showing the significance of enriched biological processes related to the overlapping 251 genes reported in (a).

### 2.4.5 Discussion:

In this study, the role of the splicing factor hnRNPL in controlling the ER $\alpha$ <sup>+</sup> BC transcriptome was investigated. The gene and splicing level changes controlled by hnRNPL in this cancer type were investigated by the analysis of a paired-end RNA-seq experiment consisting of silencing *HNRNPL* gene with specific siRNA in MCF-7, representing the ER $\alpha$ <sup>+</sup> luminal BC subtype. Biological processes and pathways affected by the silencing of *HNRNPL* were identified. In addition, splicing level changes induced by *HNRNPL* silencing were characterized. Furthermore, in addition to characterizing the prognostic value of *HNRNPL* in BCs and its association with tumor molecular and clinical features, *HNRNPL*-regulated ASEs were investigated in BC samples and their association with clinical and molecular features was investigated.

The analysis of the RNA-seq dataset indicated that *HNRNPL* silencing in MCF-7 cells strongly hampers the expression of cell cycle-related genes. Among others, *HNRNPL* silencing reduced ER $\alpha$  gene expression, the key transcription factor mediating proliferation of these cells (Caizzi et al. 2014; Miano et al. 2016), the E2F transcription factors 1 (*E2F1*) and 2 (*E2F2*), involved in the regulation of BC cell proliferation (Stender et al. 2007). Other cell cycle related genes that are significantly downregulated by *HNRNPL* silencing are multiple minichromosome maintenance genes (*MCM2*, *MCM3*, *MCM4*, *MCM5*, *MCM6*, and *MCM7*), key factors involved in the initiation of DNA replication and DNA unwinding (Forsburg 2004), and which were recently suggested as putative markers for their prognostic value in predicting relapse-free survival of BC subtypes (Issac et al. 2019). Furthermore, *HNRNPL* silencing in these cells reduced the expression of the checkpoint kinases 1 (*CHEK1*) and 2 (*CHEK2*), crucial genes for cell cycle checkpoints and DNA-damage induced cell cycle arrest (van Jaarsveld et al. 2020). All together, these data suggest the potential role of *HNRNPL* in controlling the cell proliferation pathway in luminal BC cells, which we supported by proliferation assay in MCF-7 cells 24 and 48 hours upon the silencing (Supplementary Figure 26). On the other hand, silencing *HNRNPL* in MCF-7 cells activates the expression of genes related to cell migration in line with previously reported studies (D'Agostino, Caracciolo, and Giordano 2010).

The downregulation of ER $\alpha$  expression upon *HNRNPL* silencing could be a direct or indirect effect of hnRNPL depletion in these cells. Analysis of hnRNPL CLIP-Seq datasets from other epithelial cell lines indicates indeed a binding of hnRNPL at ER $\alpha$ , which may result in the control of the transcription and/or translation of the gene. A similar mechanism mediated by hnRNPL was reported in the case of p53 gene (Seo et al. 2017). Another possible explanation for this is that hnRNPL regulates the expression of upstream regulators of ER $\alpha$ . Notably, the analysis of the RNA-seq dataset indicates that *HNRNPL* silencing reduces the expression of multiple factors known to control the expression of ER $\alpha$ . This includes the progesterone receptor (*PGR*) gene (Harrell et al. 2006), X-box binding protein 1 (*XBPI*) (Ding et al. 2003), and the BRCA1 associated RING domain 1 (*BARD1*) genes (Dizin and Irminger-Finger 2010), all of which are upstream activators of ER $\alpha$  transcription, and which are downregulated upon *HNRNPL* silencing based on the analysis of the RNA-seq dataset.

Moreover, taking advantage of our RNA-seq dataset, and applying both an isoform-based and event-based approach, we characterized the splicing level changes induced by *HNRNPL* silencing in MCF-7 cells. GO enrichment analysis of the differentially spliced genes revealed functionally related gene sets involved in cell cycle progression, DNA repair, and G2M checkpoint pathways, in addition to RNA splicing and EMT related genes, in line with enriched terms at the gene level. Furthermore, by *in silico* analysis through RBP-binding motif enrichment, we confirmed the direct implication of hnRNPL in the observed changes at the splicing level, showing that a high number of differentially spliced genes harbours an hnRNPL binding motif, most of which were supported by CLIP-seq evidence from other epithelial cell lines. In addition, by characterizing the expression of the observed AS changes in BC tissues, we identified a set of differentially expressed ASEs and indicated their correlation with *HNRNPL* mRNA levels in the analysed samples, taking into account ER $\alpha$  expression status. The analysis of the association of *HNRNPL* gene expression with clinical outcomes of BC patients indicates that the higher gene expression is associated with better prognosis such as longer overall survival and better distant-metastasis-free and disease-relapse free survival in ER $\alpha$ + patients, while a worse prognosis (e.g. shorter OS, and worse distant-metastasis-free survival times) associated with the overexpression of the gene in ER $\alpha$ - patients. This, in combination with the observation that different subsets of ASEs correlate with *HNRNPL* mRNA levels depending on ER $\alpha$  status, suggests that hnRNPL function is affected by other factors exhibiting a subtype-specific expression, among which, its physically interacting partner, DSCAM-AS1. Indeed, LncRNAs were shown to influence the function of splicing factors including members of hnRNP family, thereby affecting their binding to target RNA transcripts, or their subcellular localization (Gonzalez et al. 2015).

### chapter 3. General conclusions

The present work consisted of deciphering the possible implications of ER $\alpha$  in the control of post transcriptional events in the luminal BC subtype. Thanks to RNA-seq, we were able to determine three possible mechanisms by which ER $\alpha$  controls post transcriptional events in these tumors. The first direct mechanism is that ER $\alpha$  controls the expression and activity of hundreds of RNA-binding proteins among which many splicing factors. ER $\alpha$  silencing in MCF-7 BC cell lines, representative of the luminal subtype, reduced the expression of epithelial-specific while induced an increase in the expression of mesenchymal-specific RNA-binding proteins and splicing factors. These gene expression changes in splicing factors resulted in a range of significant alterations at the level of isoforms upon ER $\alpha$  silencing. Furthermore, we successfully demonstrated by analysis of the RNA-Seq dataset that ER $\alpha$  controls the expression of many alternative exons that are recurrently misregulated in different cancers. Among the splicing factors significantly affected by ER $\alpha$  silencing, the two core epithelial alternative splicing regulators, ESRP1 and ESRP2 ranked among the top significantly reduced splicing factors by ER $\alpha$  silencing in MCF-7 cells. The analysis of the ESRP1/2 silencing RNA-seq experiment exemplified the functional role of ER $\alpha$  in controlling post transcriptional events by maintaining the expression of splicing factors, and resulted in significant isoform changes highly similar to those observed in the ER $\alpha$  silencing experiment. Moreover, a second mechanism demonstrating that ER $\alpha$  controls post transcriptional events is by maintaining the expression of lncRNAs, of which DSCAM-AS1 was investigated. Notably, by analysing the *DSCAM-AS1* silencing RNA-seq experiment, we identified significant changes at the level of isoforms which may be considered as a potential mechanism involved in the DSCAM-AS1-mediated proliferation in BC cells. DSCAM-AS1-mediated effects at the level of isoforms were centred around its physical interaction with hnRNPL, a key regulator of alternative splicing. We successfully confirmed this physical interaction in our MCF-7 cellular model, and *in silico* analysis by performing an RBP-binding motif enrichment analysis suggested a possible implication of hnRNPL in splicing changes observed upon DSCAM-AS1 silencing. On the other hand, *HNRNPL* silencing strongly inhibited cellular proliferation of MCF-7 cells and induced a decrease in ER $\alpha$  gene expression. Quite interestingly, *HNRNPL* silencing resulted in a significant regulation of DSCAM-AS1 at the isoform level, further supporting that their interaction might be functional. However, some technical differences between *HNRNPL* and *DSCAM-AS1* silencing experiments, in that the latter has a low sequencing depth, less suitable for alternative splicing characterization, prevented a fair comparison of the two experiments. Clearly, a total RNA-seq experiment with deeper sequencing will be an optimal solution to fairly compare the two silencing conditions.

From a computational point of view, we successfully applied two different approaches on our RNA-seq datasets to characterize isoform regulation events, (i) an isoform-based approach, represented by the IsoformSwitchAnalyzeR algorithm (Vitting-Seerup and Sandelin 2017b), and (ii) event-based approach, represented by rMATS (Shen et al. 2014) or Whippet (Sterne-Weiler et al. 2018). Based on the relative abundance of individual isoforms, the isoform-based approach enabled the general description of the overall changes considering an isoform-resolution scale. Compared to other approaches, the strength of the isoform-based approach resides in the fact that it enables an easier biological interpretation of the isoform-level changes

since the full length isoform is annotated for several features including coding potential, sensitivity to nonsense mediated decay, protein domains and other features. Therefore, the downstream consequences of a switching pair of isoforms could be easily predicted. The drawback of this approach is that it strictly depends on the quantification tool performing the quantification of isoform expression, which sometimes fails to correctly assign short RNA-seq reads to their transcript of origin when they are aligned to shared exons of alternatively spliced genes, especially when analyzing unstranded RNA-seq datasets (S. Zhao et al. 2015). On the other hand, the event-based approach is simpler to interpret from a splicing regulation perspective, but is harder to draw biological conclusions from since ASEs are analyzed as disjoint blocks. Therefore, a best alternative to tackle these approach-specific issues was to apply both approaches since they are complementary one to another. Such alternative also increases the chances for detecting complex non-binary ASEs which may take place (Vaquero-Garcia et al. 2016).

## chapter 4. Materials and Methods

### 4.1 Computational methods:

In this section are described the different computational methods used for the analysis of the different RNA-sequencing datasets for the determination of changes at both gene, isoform, and splicing levels. Furthermore, data used from external databases are also described.

#### 4.1.1 Analysis of *ESRP1*, *ESRP2*, *ESR1*, and *HNRNPL* expression in TCGA clinical data

Analysis of *ESRP1*, *ESRP2*, *ESR1*, and *HNRNPL* expression in primary tumor RNA-Seq data was performed by considering the 774 BCs from the BRCA cohort of The Cancer Genome Atlas (TCGA). The *ESRP1*, *ESRP2*, *ESR1*, and *HNRNPL* expression levels in Fragment Per Kilobase Mapped Reads were retrieved from GDC (Genomic Data Commons) data portal (Grossman et al. 2016a). Clinical data were obtained from cBioPortal (Gao et al. 2013), considering the clinical data from the dataset named “Breast Invasive Carcinoma (TCGA, PanCancer Atlas)”. The same dataset was used to obtain the copy number status information expressed as GISTIC score. The analysis was performed only on ER $\alpha$ + tumors defined based on the immunohistochemistry level of ER $\alpha$ . Samples with a positive ER $\alpha$  status based on immunohistochemistry but associated with an *ESR1* FPKM lower than 1 were excluded from the analysis. 192 BCs characterized by a negative ER $\alpha$  immunohistochemistry level and *ESR1* FPKM lower than 1 were considered as ER $\alpha$  negative tumors.

The statistical analysis of *ESRP1*, *ESRP2*, and *ESR1* expression in relation to different patient clinical data was performed considering both the expression level of these genes and by separating the samples in highly and lowly expressing patients based on the median expression values of each gene. The analysis of the gene expression levels in relation to continuous clinical data was performed using a Spearman correlation analysis (implemented in the *cor\_test* R function) while categorical variables were analyzed using a Kruskal-Wallis test (implemented in the *kruskal.test* R function). Chi-square analysis was applied to analyze categorical clinical data with respect to two groups of patients (high and low) defined based on the median expression of the analyzed genes.

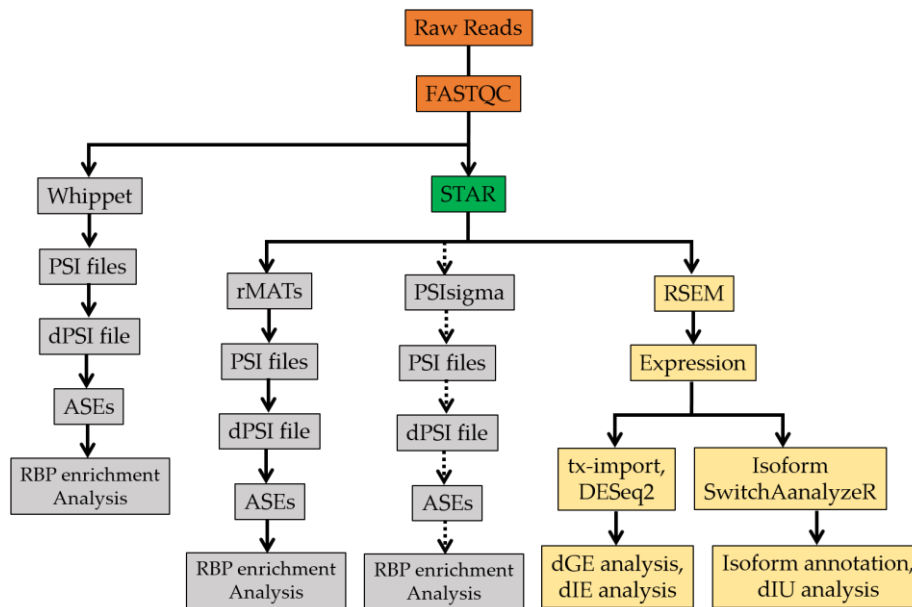
#### 4.1.2 Overlap with ER $\alpha$ ChIP-Seq data

Analysis of ER $\alpha$  chromatin binding at *ESRP1*, and *ESRP2* gene locus was performed by considering the reference of ER $\alpha$  chromatin interactions in MCF-7 cells from (Ferrero et al. 2017). In the analysis were considered data of hormone-deprived cells (E2-independent), cells maintained in hormone-enriched media (E2-constitutive), or cells treated with 17-beta estradiol (E2-treatment). ER $\alpha$  chromatin bindings in tamoxifen-sensitive or resistant cell lines (MCF-7 and BT-474) and primary BCs were obtained from GSE32222 (Ross-Innes et al. 2012). The peak genomic coordinates were converted in hg38 genome assembly using LiftOver utility of the UCSC genome browser (Kuhn, Haussler, and Kent 2013).



### 4.1.3 RNA-sequencing dataset analysis

A flow chart of the mRNA sequencing data analysis, including pre-processing and quality control of short RNA-seq reads, quantification at both gene and isoform levels, identification of differentially expressed genes and isoforms, identification of differentially regulated ASEs is summarized in (Figure 4.1-1).



**Figure 4.1-1:** Computational pipeline used for the analysis of changes in gene expression in the 4 investigated RNA-seq datasets (*siERa*, *siESRP1/2*, *LNA-DSCAM-AS1*, *siHNRNPL*), from raw reads to identification of regulated candidate genes and isoforms. Alignment based pipeline includes an alignment step with STAR (Dobin et al. 2013), quantification with RSEM (B. Li and Dewey 2011), differential expression analysis (dGE, dIE) with tx-import-DESeq2 (Soneson, Love, and Robinson 2015; Love, Huber, and Anders 2014), and identification of isoform differential usage using IsoformSwitchAnalyzeR (Vitting-Seerup and Sandelin 2017b). Local alternative splicing changes (ASEs) were detected using three independent tools including Whippet (Sterne-Weiler et al. 2018), rMATS (Shen et al. 2014), and PSIsigma (K.-T. Lin and Krainer 2019). ASEs, local Alternative Splicing Events; dGE, Differential Gene Expression; dIE, differential Isoform Expression; dIU, Differential Isoform Usage; PSI, Percent Spliced-in Index; RBP, RNA Binding Protein.

#### 4.1.3.1 Short RNA-seq reads preprocessing and alignment

The ER $\alpha$  silencing RNA-seq experiment is a polyA+, paired-end, unstranded RNA-seq dataset generated using MCF-7 cells under hormone-deprived culture conditions, as previously described (Miano et al. 2018). The raw and processed RNA-seq dataset are deposited at GSE108693. The *DSCAM-AS1* silencing RNA-seq dataset is a polyA+, paired-end, unstranded RNA-seq dataset generated using MCF-7 cells cultured in full medium, as previously described (Elhasnaoui et al. 2020). The raw and processed RNA-seq dataset were deposited at GSE150591. The *HNRNPL* and *ESRP1/2* silencing RNA-seq experiments are both total, paired-end, stranded RNA-seq datasets which were generated using MCF-7 cells under full medium culture conditions (manuscripts in preparation). Prior to quantification, raw RNA-seq reads were assessed for Phred quality scores using the FASTQC program (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), retaining only reads of 75

bass length. After quality confirmation, clean reads were aligned against the human reference genome based on the hg38 genome assembly (GRCh38.p10 assembly) and gencode v27 gene annotation model using STAR v2.5.1b (Dobin et al. 2013). STAR was run in two-pass mode allowing alignment to the transcriptome coordinates by setting the option `--quantMode` to `TranscriptomeSAM`.

#### 4.1.3.2 Gene and isoform expression quantification

The quantification of the expression at gene and isoform level was performed using RSEM (RNA-seq by Expectation Maximisation) (B. Li and Dewey 2011). RSEM was run in default setting on *ERα* and *DSCAM-AS1* RNA-seq datasets, and with the following options “`--calc-ci --paired-end --forward-prob 0`” on HNRNPL and ESRP1/2 RNA-seq datasets. RSEM allows the quantification of the expression of genes and isoforms from single-end or paired-end RNA-seq experiments (B. Li and Dewey 2011). RSEM handles ambiguous reads that map on overlapping genes or isoforms, by computing a posterior estimate and 95% credibility interval defining a maximum likelihood estimate for the abundance of each gene and isoform (B. Li and Dewey 2011).

#### 4.1.3.3 Differential gene/isoform expression analysis

The identification of differentially expressed genes and isoforms in the investigated RNA-seq datasets, comparing silencing to control conditions, was performed using the DESeq2 R package (v1.26.0) in default parameters (Love, Huber, and Anders 2014). The expression at isoform level was summarized to gene-level using the *tx-import* bioconductor package (Soneson, Love, and Robinson 2015) and the resulting count matrices were provided to DESeq2. Prior to DE analysis, lowly expressed genes and isoforms (< 10 normalized counts) were discarded from the analysis and only genes and isoforms with more than 10 normalized read counts in at least one condition (3 out of 6 samples, e.g expressed in at least one condition) were considered for further downstream analysis. Then, a gene or isoform was considered as differentially expressed if its associated BH-adjusted p-value < 0.05. All the data visualization plots including heat maps and volcano plots, were made using ggplot2 R package (v.3.2.1) (Wilkinson 2011).

#### 4.1.3.4 Gene ontology enrichment analysis

The identification of affected biological pathways due to changes at gene or isoform level in each of the analysed RNA-seq datasets was performed based on a gene ontology (GO) enrichment analysis using the Gene Annotation and Analysis Resource Metascape program (Y. Zhou et al. 2019). In particular, this analysis was performed separately for up-regulated and down-regulated genes, using the Single List Analysis option of Metascape (Y. Zhou et al. 2019). Statistically enriched GO terms related to each category of genes were obtained from the GO Biological Processes and Hallmark Get Sets branches of Metascape (Y. Zhou et al. 2019). Next, GO terms that were associated with an enrichment factor > 1.5 and an accumulative hypergeometric test adj. p-value < 0.05 were considered as significantly enriched. To reduce redundancy, GO terms showing a high number of overlapping genes and a large degree of redundancies were clustered into groups of clusters based on their degree of similarities, and each cluster was represented by the top significant

GO term (Highest gene count and smallest adj-p). The Top 20 significant clusters were selected for visualization purposes.

#### 4.1.3.5 Differential isoform usage analysis: IsoformSwitchAnalyzeR

To test for isoform switching events, the IsoformSwitchAnalyzeR Bioconductor package was applied (Vitting-Seerup and Sandelin 2017b). Briefly, the tool is fed with a transcript sequence file (gencode.transcripts.fasta) and isoforms relative abundances (in TPM units, Transcript per Million fragments Mapped) obtained from RSEM. Next, an isoform fraction (IF) metric representing the usage of each isoform in each RNA-seq replicate is calculated dividing the relative isoform abundance to that of its parent gene ( $TPM_{iso}/TPM_{gene}$ ). Prior to statistical testing of the differential usage of isoforms (dIU) between conditions, lowly expressed genes with an overall expression less than 0.5 TPM unit were discarded. In addition, lowly expressed isoforms having an  $IF < 0.01$  (e.g. not contributing to the expression of the gene) were excluded from downstream analysis. Then, For each isoform, a differential isoform usage (dIF) was calculated as  $(IF_{silencing} - IF_{control})$  using a modified DEXSeq utility implemented in IsoformSwitchAnalyzeR (Vitting-Seerup and Sandelin 2017b). Finally, isoforms with a  $|dIF| > 0.05$  and a BH-adjp  $< 0.05$  were considered as significantly switching isoforms.

##### 4.1.3.5.1 Isoform level annotation

Next, sequences of expressed isoforms in each condition were retrieved from the transcript fasta file and were annotated for different features including the presence or absence of signal peptide sequences, for associated protein domains and coding potential using signalP (Almagro Armenteros et al. 2019), Pfam (El-Gebali et al. 2019), and CPC2 (Y.-J. Kang et al. 2017) tools, respectively. The sensitivity of a given transcript to nonsense mediated decay (NMD) process is defined based on the distance between the last exon junction and the first polyadenylation site ( $< 60$  bases NMD-sensitive) (Vitting-Seerup and Sandelin 2017b). Significant switching isoform pairs were then compared for these features, defining a relative proportion of isoforms resulting from opposing consequences, including coding versus non-coding potential, protein domains gain/loss, with or without signal peptide sequences, shortening/lengthening of the open reading frame were then evaluated for the switching isoforms from the same parent gene. Accordingly, genes harbouring switching isoform pairs with opposing features were considered as switching genes with downstream functional consequences (Vitting-Seerup and Sandelin 2017b).

##### 4.1.3.6 Differential local alternative splicing analysis

Since IsoformSwitchAnalyzeR analysis is based on the relative abundance of full length isoforms, it is not capable of quantifying changes in local alternative splicing events involved in the identified switches. Therefore, we sought to apply an event-based approach capable of measuring local AS changes based on the inclusion levels (PSI) of each ASE, which is calculated considering the total number of read counts supporting either inclusion or exclusion of each ASE. The significant results from this event-based approach showing a differential inclusion level ( $|dPSI| > 0.05$ , adjp  $< 0.05$ ) between silencing and control conditions of the different

RNA-seq datasets were then compared with that of dIU analysis and concordant changes were considered for further discussion.

#### 4.1.3.6.1 Tools for alternative splicing events detection

The analysis of local alternative splicing from short read RNA-seq data was possible through the development of different approaches capable of calculating Percent Spliced-In (PSI) values for each single ASE. These approaches are classified into three main categories ranging from (i) event-based approaches such as rMATS (Shen et al. 2014) and PSIsigma (K.-T. Lin and Krainer 2019), (ii) isoform-resolution-based approaches such as SUPPA2 (Trincado et al. 2018) and IsoformSwitchAnalyzeR (Vitting-Seerup and Sandelin 2017b), and (iii) splice-graph-based approaches such as MAJIQ (Vaquero-Garcia et al. 2016) and Whippet (Sterne-Weiler et al. 2018). Several studies have compared the performance of these analytical approaches by benchmarking representative tools for their accuracy in measuring local AS changes and their precision in estimating PSI values using either simulated (Mehmood et al. 2020) or real experimental RNA-seq data (K.-T. Lin and Krainer 2019). These and other studies have shown the outperformance of each of rMATS (Shen et al. 2014), PSIsigma (K.-T. Lin and Krainer 2019), and Whippet (Sterne-Weiler et al. 2018) in accurately estimating PSI values of ASEs. Therefore, we decided to use these three methods for the analysis of AS changes occurring in our investigated RNA-seq datasets.

##### 4.1.3.6.1.1 rMATS

rMATS is an event-based approach for AS changes quantification from replicate RNA-seq experiments and that is more suitable for detecting simple binary ASEs based on the idea that two or more isoforms have in common one or more alternative exons (Shen et al. 2014). The PSI value of an ASE (e.g. exon skipping) is calculated by rMATS based on the total number of read counts supporting the exon inclusion isoform; that splice from the upstream flanking exon into the alternative exon and then into the downstream flanking exon; plus the exon exclusion isoform that splice from the upstream flanking exon directly into the downstream flanking exon (Shen et al. 2014). The estimation of PSI value from RNA-seq replicates is influenced by several factors including the library size which influences the sequencing coverage of each ASE, and by variability among replicates of the same group which could result from biological or technical biases such as RNA integrity differences among samples. To cope with this, rMATS applies a hierarchical framework to simultaneously account for estimation uncertainty in individual replicates and variation among groups of replicates by normalizing the PSI value on the effective length represented by the number of unique isoform-specific read positions of the exon inclusion and exon exclusion isoforms (Shen et al. 2014). rMATS reports five types of ASEs including Exon Skipping (ES), Mutually Exclusive Exons (MX), Alternative 5' (A5) and 3' (A3) splice sites, and Intron retention (IR) events.

##### 4.1.3.6.1.2 PSIsigma

PSIsigm is another event-based approach but is based on a modified PSI equation which takes into account the expression levels of both upstream and downstream constitutive exons flanking an alternative exon (K.-T. Lin

and Krainer 2019). In brief, the calculation of PSI values by PSIsigma considers the splice junction reads of all isoforms in the region between two constitutive exons, enabling PSIsigma to report a PSI value for each individual exon or splice site calculated based on the splice junction reads of all isoforms that share the same alternative exon or splice site (K.-T. Lin and Krainer 2019). PSIsigma reports seven types of ASEs including single (SES) and multiple (MES) exon skipping events, MX, IR, A5, A3, in addition to alternative first (AF) and Last exon (AL) events. Since the algorithm of this tool was not stable and changes over time, we decided to use PSIsigma as a secondary check of the significant ASEs reported by the most stable, well-documented algorithm, rMATS.

#### **4.1.3.6.1.3 Whippet**

Whippet is an algorithm dedicated for the measurement of splicing changes of any complexity, comprising both simple binary and more complex, non-binary, ASEs (Sterne-Weiler et al. 2018). Whippet takes as input a gene annotation model and creates an index which models transcriptome structure by constructing directed contiguous splice graphs, where nodes are non-overlapping exonic regions of the transcriptome, and edges correspond to splice junctions or adjacent exonic regions. In this way, Whippet represents single isoforms as paths connecting different nodes inside a given splice graph (Sterne-Weiler et al. 2018). Theoretical incoming and outgoing connections of each node in the contiguous splice graph are defined based on the 5' and 3' splice sites, and transcription start or end sites. In addition, Whippet allows efficient RNA-seq alignment by recording for 5' and 3' splice sites in the contiguous splice graph an upstream and downstream k-mer sequence, respectively, so as to allow quantification of all possible splice junctions and to discover novel splice junctions that do not appear in the provided gene annotation model (Sterne-Weiler et al. 2018). After the alignment of RNA-seq short reads against the contiguous splice graph, Whippet defines ASEs based on the set of edges connecting or skipping each specific node included in the contiguous splice graph. The inclusion levels (PSI) is calculated for each node as a fraction of the number of RNA-seq short reads spanning on the paths connecting the node and the total number of reads connecting and skipping the node (Sterne-Weiler et al. 2018). To cope with ambiguous multi-mapping paths that could have common edges, Whippet implies an expectation maximization algorithm to determine a maximum likelihood estimation of the proportional abundance of each ambiguous path. In this way, the PSI value of a node is then calculated as the sum of proportional abundance of the paths containing the node (Sterne-Weiler et al. 2018). In addition to the ASE types reported by rMATS and PSIsigma, Whippet is capable of calculating the relative usage of alternative transcription and polyadenylation sites based on the relative abundance of paths supporting each of them.

#### **4.1.3.6.2 RBP binding motif enrichment analysis**

To identify the set of RNA-binding proteins potentially involved in the control of alternative splicing changes occurring in each of our datasets, an RBP-binding motif enrichment analysis was performed. First, RBP binding motifs for 105 different splicing factors were retrieved from a study by (Sebestyén et al. 2016) generating Position Weight Matrices (PWMs) for each splicing factor, based on a collection of motifs from the RNACompete study (Ray et al. 2017). In addition, for the binding motifs of a number of RBPs in other species,

confirmed in a previous study (Sebestyén et al. 2016) to be conserved between humans and the other species, were used in this analysis. This includes the splicing factor RBM47 (chicken), SF1 (*Drosophila*), SRP4 (*Drosophila*), TRA2 (*Drosophila*), and PCBP3 (mouse). Next, sequences of genomic regions involved in the ASEs extended 200 nucleotides on both sides (with the exception of MX events which were extended 100 nucleotides only) were scanned for the presence of RBP binding motifs using the MoSEA (Motif Scan and Enrichment Analysis) package (Sebestyén et al. 2016) and FIMO (Find Individual Motif Occurrences) (Grant, Bailey, and Noble 2011). For alternative polyadenylation sites, the APA nodes reported by Whippet were sorted by coordinates and classified as proximal or distal APA sites, both of which were extended 100 nucleotides on both sides from the APA site position. RBP binding motifs with an associated BH-adjp < 0.0001 were considered as significant. Next, the RBP-binding motif enrichment analysis was performed for each ASE type, region, and direction of regulation (e.g. inclusion (dPSI>0), exclusion (dPSI<0)) by comparing the frequency (observed average) of each RBP-binding motif in each regulated ASE with that observed (expected average) in a control set of expressed exons, not regulated by the silencing of our investigated genes. To define a robust z-score representing the enrichment of each RBP binding motif, MoSEA was run in a bootstrapping mode (500 iterations) each time selecting a pool of 100 randomly selected sequences of the same size from equivalent regions in non-regulated exons ( $|\text{dPSI}| < 0.01$  and  $\text{adjp} > 0.5$ ). Finally, an enrichment robust z-score was then calculated for each RBP-binding motif, region and direction of regulation, as a  $((\text{observed average} - \text{expected average}) / \text{expected SD})$  of the observed frequency in the regulated ASEs set with the mean and standard deviation of the control ASE sets. An RBP was considered as enriched if associated with a (z-score > 1.96) (Sebestyén et al. 2016). The obtained enrichment z-scores per binding motif, region, and ASE type were then visualized using the ggplot2 Bioconductor R package (Wilkinson 2011).

#### 4.1.3.6.3 Alternative splicing events annotation

To predict the downstream functional consequences of alternative splicing changes at the level of protein isoforms, we took advantage of the recently published database, DIGGER (Louadi et al. 2021), which holds different kinds of information including protein-protein interactions, domain-domain interactions, and residue-level interactions allowing a better understanding of isoform-specific effects on protein-protein interactions, and further enriching our knowledge of mechanistic effects of alternative splicing. In brief, DIGGER database holds interaction information about 24,969 reference proteins, 410,961 interactions that were curated from the BioGrid database (Oughtred et al. 2019), and high-confidence domain-domain interactions from 3did (Mosca et al. 2014) and DOMAINE (Raghavachari et al. 2008) databases. Cassette exons showing significant differential inclusion levels between the silencing and control conditions of the different RNA-seq datasets were analysed using the DIGGER database for their possible implication in specific protein-protein interactions. The effects of these observed ASEs were evaluated by comparing changes in the interactions mediated by the inclusion and exclusion isoforms of each ASEs.

#### 4.1.3.6.4 Analysis of alternative splicing events expression in BC tissues

To verify the importance of each ASE identified as significantly changing by the silencing in each RNA-seq dataset, breast cancer splicing data was retrieved from the TCGASpliceSeq database (M. Ryan et al. 2016). ASEs expression data represented by PSI values was detectable in 1072 BC samples including 774 ER $\alpha$ +, 192 ER $\alpha$ - BC samples, in addition to 119 normal breast tissues. As SpliceSeqTCGA database was made using hg19 human genome assembly, the coordinates of ASEs resulting from each RNA-seq dataset were converted from hg38 to hg19 coordinates using liftOver utility of the UCSC genome browser (Kuhn, Haussler, and Kent 2013). Next, ASEs were then overlapped with the list of exons having PSI values in the TCGASpliceSeq (M. Ryan et al. 2016) database. Each ASE was then defined by an identifier including gene name, position of the flanking upstream exon, the alternative exon, the downstream flanking exon, and the type of the event. For example, MYOF\_16\_17\_18\_ES mains an exon skipping event at MYOF gene involving the alternative the exon number 17. Of note, as the number of annotated exons per isoforms increased based on the hg38 genome assembly, it is possible that an exon could have different position indices between the hg19 and the hg38 genome assemblies. Differential inclusion/exclusion analysis of ASEs between groups of samples was calculated based on the ASE median inclusion (PSI) level in each group. A Wilcoxon test was performed to determine the statistical significance of ASEs inclusion differences (dPSI) between groups of samples using R software. The correlation between ASE inclusion level and the expression of the four genes (*ESR1*, *ESRP1*, *ESRP2*, *HNRNPL*) was performed using Spearman rank rho test implemented in the `cor.test()` function of R software. ASEs associated with a p-value <0.05 were considered as statistically significant.

#### 4.1.3.6.5 Pathway-guided enrichment analysis of alternative splicing events using PEGASAS

In order to understand the possible molecular pathways in which each ASE is involved, a pathway-guided enrichment analysis of the identified ASEs was performed using the PEGASAS algorithm (Phillips et al. 2020). The algorithm takes as input a txt file containing the inclusion level of each ASE in each BC sample and calculates the correlation between the inclusion level (PSI) of each ASE with molecular oncogenic pathways. In brief, this correlation analysis by PEGASAS involves two main steps. The first step consists of characterizing the activity of molecular pathways based on the expression level of their genes in TCGA breast cancer samples. The second step consists of assessing the correlation between each pathway-exon pairs across BC samples. Pathway-exon pairs with a Pearson correlation coefficient > 0.2 were considered as significantly correlated (Phillips et al. 2020). This analysis resulted in the identification of the correlation between the identified ASEs in each of the investigated RNA-seq datasets and 50 molecular oncogenic pathways.

#### 4.1.3.6.6 Association with BC patients overall and disease-free survival

To verify the prognostic value of AS patterns changing by the silencing of target genes in each RNA-seq dataset, the association between ASE inclusion and overall and disease-free survivals in ER $\alpha$  + patients was investigated. ER $\alpha$  + BC patients were classified into highly expressing and lowly expressing the ASE based on

the median expression level of the event across all the samples. Clinical data of BC patients including overall and disease-free survival information was retrieved from the TCGA DGC data portal (Grossman et al. 2016b). The survival analysis was performed using the survival (Therneau and Grambsch 2000) and survminer R packages.



---

## References:

- Achinger-Kawecka, Joanna, Fatima Valdes-Mora, Phuc-Loi Luu, Katherine A. Giles, C. Elizabeth Caldon, Wenjia Qu, Shalima Nair, et al. 2020. “Epigenetic Reprogramming at Estrogen-Receptor Binding Sites Alters 3D Chromatin Landscape in Endocrine-Resistant Breast Cancer.” *Nature Communications* 11 (1): 320.
- Adlerz, Linda, Marie Beckman, Sofia Holback, Roya Tehranian, Veronica Cortés Toro, and Kerstin Iverfeldt. 2003. “Accumulation of the Amyloid Precursor-like Protein APLP2 and Reduction of APLP1 in Retinoic Acid-Differentiated Human Neuroblastoma Cells upon Curcumin-Induced Neurite Retraction.” *Molecular Brain Research*. <https://doi.org/10.1016/j.molbrainres.2003.08.014>.
- Ajj, Hussein, Amand Chesnel, Sophie Pinel, François Plenat, Stéphane Flament, and Helene Dumond. 2013. “An Alkylphenol Mix Promotes Seminoma Derived Cell Proliferation through an ERalpha36-Mediated Mechanism.” *PLoS One* 8 (4): e61758.
- Alamancos, Gael P., Amadís Pagès, Juan L. Trincado, Nicolás Bellora, and Eduardo Eyras. 2015. “Leveraging Transcript Quantification for Fast Computation of Alternative Splicing Profiles.” *RNA* 21 (9): 1521–31.
- Al-Dhaheri, Mariam H., and Brian G. Rowan. 2007. “Protein Kinase A Exhibits Selective Modulation of Estradiol-Dependent Transcription in Breast Cancer Cells That Is Associated with Decreased Ligand Binding, Altered Estrogen Receptor Alpha Promoter Interaction, and Changes in Receptor Phosphorylation.” *Molecular Endocrinology* 21 (2): 439–56.
- Almagro Armenteros, José Juan, Konstantinos D. Tsirigos, Casper Kaae Sønderby, Thomas Nordahl Petersen, Ole Winther, Søren Brunak, Gunnar von Heijne, and Henrik Nielsen. 2019. “SignalP 5.0 Improves Signal Peptide Predictions Using Deep Neural Networks.” *Nature Biotechnology* 37 (4): 420–23.
- Al Saleh, Sanaa, Fahd Al Mulla, and Yunus A. Luqmani. 2011. “Estrogen Receptor Silencing Induces Epithelial to Mesenchymal Transition in Human Breast Cancer Cells.” *PLoS One* 6 (6): e20610.
- Ameres, Stefan L., and Phillip D. Zamore. 2013. “Diversifying microRNA Sequence and Function.” *Nature Reviews Molecular Cell Biology*. <https://doi.org/10.1038/nrm3611>.
- Andreassi, Catia, Hamish Crerar, and Antonella Riccio. 2018. “Post-Transcriptional Processing of mRNA in Neurons: The Vestiges of the RNA World Drive Transcriptome Diversity.” *Frontiers in Molecular Neuroscience* 11 (August): 304.
- An, Yu, Christine Y. Chen, Bryan Moyer, Piotr Rotkiewicz, Marc-André Elsliger, Adam Godzik, Ian A. Wilson, and William E. Balch. 2009. “Structural and Functional Analysis of the Globular Head Domain of p115 Provides Insight into Membrane Tethering.” *Journal of Molecular Biology* 391 (1): 26–41.
- Arab, Khelifa, Emil Karaulanov, Michael Musheev, Philipp Trnka, Andrea Schäfer, Ingrid Grummt, and Christof Niehrs. 2019. “GADD45A Binds R-Loops and Recruits TET1 to CpG Island Promoters.” *Nature Genetics* 51 (2): 217–23.
- Arafat, Hwya, Melissa Lazar, Khalifa Salem, Galina Chipitsyna, Qiaoke Gong, Te-Cheng Pan, Rui-Zhu Zhang, Charles J. Yeo, and Mon-Li Chu. 2011. “Tumor-Specific Expression and Alternative Splicing of the COL6A3 Gene in Pancreatic Cancer.” *Surgery* 150 (2): 306–15.

- Arcuri, C., I. Giambanco, R. Bianchi, and R. Donato. 2002. "Annexin V, Annexin VI, S100A1 and S100B in Developing and Adult Avian Skeletal Muscles." *Neuroscience* 109 (2): 371–88.
- Ariel, Federico, Leandro Lucero, Aurelie Christ, Maria Florencia Mammarella, Teddy Jegu, Alaguraj Veluchamy, Kiruthiga Mariappan, et al. 2020. "R-Loop Mediated Trans Action of the APOLO Long Noncoding RNA." *Molecular Cell* 77 (5): 1055–65.e4.
- Arpino, Grazia, Heidi Weiss, Adrian V. Lee, Rachel Schiff, Sabino De Placido, C. Kent Osborne, and Richard M. Elledge. 2005. "Estrogen Receptor-Positive, Progesterone Receptor-Negative Breast Cancer: Association with Growth Factor Receptor Expression and Tamoxifen Resistance." *Journal of the National Cancer Institute* 97 (17): 1254–61.
- Ast, Gil. 2004. "How Did Alternative Splicing Evolve?" *Nature Reviews Genetics*. <https://doi.org/10.1038/nrg1451>.
- Auboeuf, Didier, Dennis H. Dowhan, Martin Dutertre, Natalia Martin, Susan M. Berget, and Bert W. O'Malley. 2005. "A Subset of Nuclear Receptor Coregulators Act as Coupling Proteins during Synthesis and Maturation of RNA Transcripts." *Molecular and Cellular Biology* 25 (13): 5307–16.
- Auboeuf, Didier, Arnd Hönig, Susan M. Berget, and Bert W. O'Malley. 2002. "Coordinate Regulation of Transcription and Splicing by Steroid Receptor Coregulators." *Science* 298 (5592): 416–19.
- Audia, James E., and Robert M. Campbell. 2016. "Histone Modifications and Cancer." *Cold Spring Harbor Perspectives in Biology* 8 (4): a019521.
- Bailey, Timothy L., James Johnson, Charles E. Grant, and William S. Noble. 2015. "The MEME Suite." *Nucleic Acids Research* 43 (W1): W39–49.
- Bai, Xuefeng, Shanshan Shi, Bo Ai, Yong Jiang, Yuejuan Liu, Xiaole Han, Mingcong Xu, et al. 2019. "ENdb: A Manually Curated Database of Experimentally Supported Enhancers for Human and Mouse." *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkz973>.
- Baralle, Francisco E., and Jimena Giudice. 2017. "Alternative Splicing as a Regulator of Development and Tissue Identity." *Nature Reviews. Molecular Cell Biology* 18 (7): 437–51.
- Barash, Yoseph, John A. Calarco, Weijun Gao, Qun Pan, Xinchun Wang, Ofer Shai, Benjamin J. Blencowe, and Brendan J. Frey. 2010. "Deciphering the Splicing Code." *Nature* 465 (7294): 53–59.
- Barton, Matthias, Edward J. Filardo, Stephen J. Lolait, Peter Thomas, Marcello Maggiolini, and Eric R. Prossnitz. 2018. "Twenty Years of the G Protein-Coupled Estrogen Receptor GPER: Historical and Personal Perspectives." *The Journal of Steroid Biochemistry and Molecular Biology*. <https://doi.org/10.1016/j.jsbmb.2017.03.021>.
- Batra, Ranjan, Konstantinos Charizanis, Mini Manchanda, Apoorva Mohan, Moyi Li, Dustin J. Finn, Marianne Goodwin, et al. 2014. "Loss of MBNL Leads to Disruption of Developmentally Regulated Alternative Polyadenylation in RNA-Mediated Disease." *Molecular Cell* 56 (2): 311–22.
- Bennesch, Marcela A., and Didier Picard. 2015. "Minireview: Tipping the Balance: Ligand-Independent Activation of Steroid Receptors." *Molecular Endocrinology* 29 (3): 349–63.

- Bernstein, Bradley E., Alexander Meissner, and Eric S. Lander. 2007. "The Mammalian Epigenome." *Cell*. <https://doi.org/10.1016/j.cell.2007.01.033>.
- Bhatnagar, Sanchita, Claude Gazin, Lynn Chamberlain, Jianhong Ou, Xiaochun Zhu, Jogender S. Tushir, Ching-Man Virbasius, et al. 2014. "TRIM37 Is a New Histone H2A Ubiquitin Ligase and Breast Cancer Oncoprotein." *Nature* 516 (7529): 116–20.
- Bhat-Nakshatri, Poornima, Eun-Kyung Song, Nikail R. Collins, Vladimir N. Uversky, A. Keith Dunker, Bert W. O'Malley, Tim R. Geistlinger, Jason S. Carroll, Myles Brown, and Harikrishna Nakshatri. 2013. "Interplay between Estrogen Receptor and AKT in Estradiol-Induced Alternative Splicing." *BMC Medical Genomics* 6 (June): 21.
- Bhat-Nakshatri, Poornima, Guohua Wang, Hitesh Appaiah, Nikhil Luktuke, Jason S. Carroll, Tim R. Geistlinger, Myles Brown, Sunil Badve, Yunlong Liu, and Harikrishna Nakshatri. 2008. "AKT Alters Genome-Wide Estrogen Receptor Alpha Binding and Impacts Estrogen Signaling in Breast Cancer." *Molecular and Cellular Biology* 28 (24): 7487–7503.
- Biamonti, Giuseppe, Morena Catillo, Daniela Pignataro, Alessandra Montecucco, and Claudia Ghigna. 2014. "The Alternative Splicing Side of Cancer." *Seminars in Cell & Developmental Biology*. <https://doi.org/10.1016/j.semcdb.2014.03.016>.
- Bittencourt, Danielle, Martin Dutertre, Gabriel Sanchez, Jérôme Barbier, Lise Gratadou, and Didier Auboeuf. 2008. "Cotranscriptional Splicing Potentiates the mRNA Production from a Subset of Estradiol-Stimulated Genes." *Molecular and Cellular Biology* 28 (18): 5811–24.
- Black, D. L. 2000. "Protein Diversity from Alternative Splicing: A Challenge for Bioinformatics and Post-Genome Biology." *Cell* 103 (3): 367–70.
- Black, Douglas L. 2003. "Mechanisms of Alternative Pre-Messenger RNA Splicing." *Annual Review of Biochemistry* 72 (February): 291–336.
- Blanco, Enrique, Mar González-Ramírez, Anna Alcaine-Colet, Sergi Aranda, and Luciano Di Croce. 2020. "The Bivalent Genome: Characterization, Structure, and Regulation." *Trends in Genetics: TIG* 36 (2): 118–31.
- Blows, Fiona M., Kristy E. Driver, Marjanka K. Schmidt, Annegien Broeks, Flora E. van Leeuwen, Jelle Wesseling, Maggie C. Cheang, et al. 2010. "Subtyping of Breast Cancer by Immunohistochemistry to Investigate a Relationship between Subtype and Short and Long Term Survival: A Collaborative Analysis of Data for 10,159 Cases from 12 Studies." *PLoS Medicine*. <https://doi.org/10.1371/journal.pmed.1000279>.
- Bocchinfuso, Wayne P., and Kenneth S. Korach. 1997. "Journal of Mammary Gland Biology and Neoplasia." <https://doi.org/10.1023/a:1026339111278>.
- Bombonati, Alessandro, and Dennis C. Sgroi. 2011. "The Molecular Pathology of Breast Cancer Progression." *The Journal of Pathology*. <https://doi.org/10.1002/path.2808>.
- Bonetti, Alessandro, Federico Agostini, Ana Maria Suzuki, Kosuke Hashimoto, Giovanni Pascarella, Juliette Gimenez, Leonie Roos, et al. 2020. "RADICL-Seq Identifies General and Cell Type-specific Principles of

- Genome-Wide RNA-Chromatin Interactions.” *Nature Communications*. <https://doi.org/10.1038/s41467-020-14337-6>.
- Bonnal, Sophie C., Irene López-Oreja, and Juan Valcárcel. 2020. “Roles and Mechanisms of Alternative Splicing in Cancer — Implications for Care.” *Nature Reviews Clinical Oncology*. <https://doi.org/10.1038/s41571-020-0350-x>.
- Bouris, Panagiotis, Spyros S. Skandalis, Zoi Piperigkou, Nikos Afratis, Konstantina Karamanou, Alexios J. Aletras, Aristidis Moustakas, Achilleas D. Theocharis, and Nikos K. Karamanos. 2015. “Estrogen Receptor Alpha Mediates Epithelial to Mesenchymal Transition, Expression of Specific Matrix Effectors and Functional Properties of Breast Cancer Cells.” *Matrix Biology: Journal of the International Society for Matrix Biology* 43 (April): 42–60.
- Branković-Magić, M., R. Janković, Z. Nešković-Konstantinović, and D. Nikolić-Vukosavljević. 2002. “Progesterone Receptor Status of Breast Cancer Metastases.” *Journal of Cancer Research and Clinical Oncology*. <https://doi.org/10.1007/s00432-001-0299-9>.
- Brannan, Kristopher W., Wenhao Jin, Stephanie C. Huelga, Charles A. S. Banks, Joshua M. Gilmore, Laurence Florens, Michael P. Washburn, et al. 2016. “SONAR Discovers RNA-Binding Proteins from Analysis of Large-Scale Protein-Protein Interactomes.” *Molecular Cell* 64 (2): 282–93.
- Bray, Nicolas L., Harold Pimentel, Páll Melsted, and Lior Pachter. 2016. “Near-Optimal Probabilistic RNA-Seq Quantification.” *Nature Biotechnology* 34 (5): 525–27.
- Brisken, C., and B. O’Malley. 2010. “Hormone Action in the Mammary Gland.” *Cold Spring Harbor Perspectives in Biology*. <https://doi.org/10.1101/cshperspect.a003178>.
- Brooks, Angela N., Peter S. Choi, Luc de Waal, Tanaz Sharifnia, Marcin Imielinski, Gordon Saksena, Chandra Sekhar Pedamallu, et al. 2014. “A Pan-Cancer Analysis of Transcriptome Changes Associated with Somatic Mutations in U2AF1 Reveals Commonly Altered Splicing Events.” *PloS One* 9 (1): e87361.
- Brow, David A. 2002. “Allosteric Cascade of Spliceosome Activation.” *Annual Review of Genetics* 36 (June): 333–60.
- Brown, Rhonda L., Lauren M. Reinke, Marin S. Damerow, Denise Perez, Lewis A. Chodosh, Jing Yang, and Chonghui Cheng. 2011. “CD44 Splice Isoform Switching in Human and Mouse Epithelium Is Essential for Epithelial-Mesenchymal Transition and Breast Cancer Progression.” *Journal of Clinical Investigation*. <https://doi.org/10.1172/jci44540>.
- Busch, Anke, and Klemens J. Hertel. 2012. “Evolution of SR Protein and hnRNP Splicing Regulatory Factors.” *Wiley Interdisciplinary Reviews. RNA* 3 (1): 1–12.
- Cai, Zhaokui, Changchang Cao, Lei Ji, Rong Ye, Di Wang, Cong Xia, Sui Wang, et al. 2020. “RIC-Seq for Global in Situ Profiling of RNA-RNA Spatial Interactions.” *Nature* 582 (7812): 432–37.
- Caizzi, Livia, Giulio Ferrero, Santina Cutrupi, Francesca Cordero, Cecilia Ballaré, Valentina Miano, Stefania Reineri, et al. 2014. “Genome-Wide Activity of Unliganded Estrogen Receptor- $\alpha$  in Breast Cancer Cells.” *Proceedings of the National Academy of Sciences of the United States of America* 111 (13): 4892–97.

- Campbell, R. A., P. Bhat-Nakshatri, N. M. Patel, D. Constantinidou, S. Ali, and H. Nakshatri. 2001. "Phosphatidylinositol 3-kinase/AKT-Mediated Activation of Estrogen Receptor Alpha: A New Model for Anti-Estrogen Resistance." *The Journal of Biological Chemistry* 276 (13): 9817–24.
- Cancer Genome Atlas Network. 2012. "Comprehensive Molecular Portraits of Human Breast Tumours." *Nature* 490 (7418): 61–70.
- Cao, Jingwei, Zhanbin Tang, and Zhiqiang Su. 2020. "Long Non-Coding RNA LINC01426 Facilitates Glioblastoma Progression via Sponging miR-345-3p and Upregulation of VAMP8." *Cancer Cell International*. <https://doi.org/10.1186/s12935-020-01416-3>.
- Carlevaro-Fita, Joana, Anisa Rahim, Roderic Guigó, Leah A. Vardy, and Rory Johnson. 2016. "Cytoplasmic Long Noncoding RNAs Are Frequently Bound to and Degraded at Ribosomes in Human Cells." *RNA* 22 (6): 867–82.
- Carninci, Piero, Albin Sandelin, Boris Lenhard, Shintaro Katayama, Kazuro Shimokawa, Jasmina Ponjavic, Colin A. M. Semple, et al. 2006. "Genome-Wide Analysis of Mammalian Promoter Architecture and Evolution." *Nature Genetics* 38 (6): 626–35.
- Carroll, Jason S., and Myles Brown. 2006. "Estrogen Receptor Target Gene: An Evolving Concept." *Molecular Endocrinology* 20 (8): 1707–14.
- Carroll, Lara S., and Mario R. Capecchi. 2015. "Hoxc8 Initiates an Ectopic Mammary Program by Regulating Fgf10 and Tbx3 Expression and Wnt/ $\beta$ -Catenin Signaling." *Development* 142 (23): 4056–67.
- Castello, Alfredo, Bernd Fischer, Katrin Eichelbaum, Rastislav Horos, Benedikt M. Beckmann, Claudia Strein, Norman E. Davey, et al. 2012. "Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins." *Cell* 149 (6): 1393–1406.
- Castiglioni, F., E. Tagliabue, M. Campiglio, S. M. Pupa, A. Balsari, and S. Ménard. 2006. "Role of Exon-16-Deleted HER2 in Breast Carcinomas." *Endocrine-Related Cancer*. <https://doi.org/10.1677/erc.1.01047>.
- Castle, John C., Chaolin Zhang, Jyoti K. Shah, Amit V. Kulkarni, Auinash Kalsotra, Thomas A. Cooper, and Jason M. Johnson. 2008. "Expression of 24,426 Human Alternative Splicing Events and Predicted Cis Regulation in 48 Tissues and Cell Lines." *Nature Genetics* 40 (12): 1416–25.
- Cavalcanti, Fernanda N., Thais F. G. Lucas, Maria Fatima M. Lazari, and Catarina S. Porto. 2015. "Estrogen Receptor ESR1 Mediates Activation of ERK1/2, CREB, and ELK1 in the Corpus of the Epididymis." *Journal of Molecular Endocrinology* 54 (3): 339–49.
- Cech, Thomas R., and Joan A. Steitz. 2014. "The Noncoding RNA Revolution-Trashing Old Rules to Forge New Ones." *Cell* 157 (1): 77–94.
- Celik, Leyla, Julie Davey Dalsgaard Lund, and Birgit Schiøtt. 2007. "Conformational Dynamics of the Estrogen Receptor  $\alpha$ : Molecular Dynamics Simulations of the Influence of Binding Site Structure on Protein Dynamics<sup>†</sup>." *Biochemistry*. <https://doi.org/10.1021/bi061656t>.
- Celotto, A. M., and B. R. Graveley. 2001. "Alternative Splicing of the *Drosophila* Dscam Pre-mRNA Is Both Temporally and Spatially Regulated." *Genetics* 159 (2): 599–608.

- Chantalat, Elodie, Frédéric Boudou, Henrik Laurell, Gaëlle Palierne, René Houtman, Diana Melchers, Philippe Rochaix, et al. 2016. "The AF-1-Deficient Estrogen Receptor ER $\alpha$ 46 Isoform Is Frequently Expressed in Human Breast Tumors." *Breast Cancer Research: BCR* 18 (1): 123.
- Chaudhury, Arindam, Shebna Cheema, Joseph M. Fachini, Natee Kongchan, Guojun Lu, Lukas M. Simon, Tao Wang, et al. 2016. "CELF1 Is a Central Node in Post-Transcriptional Regulatory Programmes Underlying EMT." *Nature Communications* 7 (November): 13362.
- Chen, D., T. Riedl, E. Washbrook, P. E. Pace, R. C. Coombes, J. M. Egly, and S. Ali. 2000. "Activation of Estrogen Receptor Alpha by S118 Phosphorylation Involves a Ligand-Dependent Interaction with TFIID and Participation of CDK7." *Molecular Cell* 6 (1): 127–37.
- Chen, Duoqiao, Taylor M. Parker, Poornima Bhat-Nakshatri, Xiaona Chu, Yunlong Liu, Yue Wang, and Harikrishna Nakshatri. 2021. "Nonlinear Relationship between Chromatin Accessibility and Estradiol-Regulated Gene Expression." *Oncogene* 40 (7): 1332–46.
- Cheng, Pai-Chiao, Hsiang-Kai Chang, and Shu-Hui Chen. 2010. "Quantitative Nanoproteomics for Protein Complexes (QNanoPX) Related to Estrogen Transcriptional Action." *Molecular & Cellular Proteomics: MCP* 9 (2): 209–24.
- Chen, Lijian, Aruo Nan, Nan Zhang, Yangyang Jia, Xin Li, Yihui Ling, Jiabin Dai, et al. 2019. "Circular RNA 100146 Functions as an Oncogene through Direct Binding to miR-361-3p and miR-615-5p in Non-Small Cell Lung Cancer." *Molecular Cancer*. <https://doi.org/10.1186/s12943-019-0943-0>.
- Chen, Lisha, Hongbing Yao, Kai Wang, and Xiangfeng Liu. 2017. "Long Non-Coding RNA MALAT1 Regulates ZEB1 Expression by Sponging miR-143-3p and Promotes Hepatocellular Carcinoma Progression." *Journal of Cellular Biochemistry*. <https://doi.org/10.1002/jcb.26158>.
- Chen, Meng, Guoliang Lyu, Miao Han, Hongbo Nie, Ting Shen, Wei Chen, Yichi Niu, et al. 2018. "3' UTR Lengthening as a Novel Mechanism in Regulating Cellular Senescence." *Genome Research*, February. <https://doi.org/10.1101/gr.224451.117>.
- Chen, Zhike, Yate-Ching Yuan, Yuanzhong Wang, Zheng Liu, Hei Jason Chan, and Shiuan Chen. 2015. "Down-Regulation of Programmed Cell Death 4 (PDCD4) Is Associated with Aromatase Inhibitor Resistance and a Poor Prognosis in Estrogen Receptor-Positive Breast Cancer." *Breast Cancer Research and Treatment*. <https://doi.org/10.1007/s10549-015-3446-8>.
- Cheung, Edwin, and W. Lee Kraus. 2010. "Genomic Analyses of Hormone Signaling and Gene Regulation." *Annual Review of Physiology* 72: 191–218.
- Che, Yingying, and Lin Fu. 2020. "Aberrant Expression and Regulatory Network of Splicing Factor-SRSF3 in Tumors." *Journal of Cancer* 11 (12): 3502–11.
- Cicatiello, Luigi, Margherita Mutarelli, Oli M. V. Grober, Ornella Paris, Lorenzo Ferraro, Maria Ravo, Roberta Tarallo, et al. 2010. "Estrogen Receptor Alpha Controls a Gene Network in Luminal-like Breast Cancer Cells Comprising Multiple Transcription Factors and microRNAs." *The American Journal of Pathology* 176 (5): 2113–30.

- Colognori, David, Hongjae Sunwoo, Andrea J. Kriz, Chen-Yu Wang, and Jeannie T. Lee. 2019. "Xist Deletional Analysis Reveals an Interdependency between Xist RNA and Polycomb Complexes for Spreading along the Inactive X." *Molecular Cell* 74 (1): 101–17.e10.
- Conn, Simon J., Katherine A. Pillman, John Toubia, Vanessa M. Conn, Marika Salmanidis, Caroline A. Phillips, Suraya Roslan, Andreas W. Schreiber, Philip A. Gregory, and Gregory J. Goodall. 2015. "The RNA Binding Protein Quaking Regulates Formation of circRNAs." *Cell* 160 (6): 1125–34.
- Consortium, International Human Genome Sequencing, and International Human Genome Sequencing Consortium. 2001. "Initial Sequencing and Analysis of the Human Genome." *Nature*. <https://doi.org/10.1038/35057062>.
- Cook, Kate B., Hilal Kazan, Khalid Zuberi, Quaid Morris, and Timothy R. Hughes. 2011. "RBPDB: A Database of RNA-Binding Specificities." *Nucleic Acids Research* 39 (Database issue): D301–8.
- Coons, Laurel A., Sylvia C. Hewitt, Adam B. Burkholder, Donald P. McDonnell, and Kenneth S. Korach. 2017. "DNA Sequence Constraints Define Functionally Active Steroid Nuclear Receptor Binding Sites in Chromatin." *Endocrinology* 158 (10): 3212–34.
- Curtis, S. W., T. Washburn, C. Sewall, R. DiAugustine, J. Lindzey, J. F. Couse, and K. S. Korach. 1996. "Physiological Coupling of Growth Factor and Steroid Receptor Signaling Pathways: Estrogen Receptor Knockout Mice Lack Estrogen-like Response to Epidermal Growth Factor." *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.93.22.12626>.
- Dago, Dougba Noel, Claudio Scafoglio, Antonio Rinaldi, Domenico Memoli, Giorgio Giurato, Giovanni Nassa, Maria Ravo, Francesca Rizzo, Roberta Tarallo, and Alessandro Weisz. 2015. "Estrogen Receptor Beta Impacts Hormone-Induced Alternative mRNA Splicing in Breast Cancer Cells." *BMC Genomics* 16 (May): 367.
- D'Agostino, Luca, Valentina Caracciolo, and Antonio Giordano. 2010. "NSP 5a3a's Link to Nuclear-Cytoplasmic Proteins B23 and hnRNP-L between Normal and Aberrant Breast Cell Lines." *Cell Cycle* 9 (6): 1131–42.
- Danan-Gotthold, Miri, Regina Golan-Gerstl, Eli Eisenberg, Keren Meir, Rotem Karni, and Erez Y. Levanon. 2015. "Identification of Recurrent Regulated Alternative Splicing Events across Human Solid Tumors." *Nucleic Acids Research* 43 (10): 5130–44.
- Dassi, Erik. 2017. "Handshakes and Fights: The Regulatory Interplay of RNA-Binding Proteins." *Frontiers in Molecular Biosciences* 4. <https://doi.org/10.3389/fmolb.2017.00067>.
- Davis, Carrie A., Benjamin C. Hitz, Cricket A. Sloan, Esther T. Chan, Jean M. Davidson, Idan Gabdank, Jason A. Hilton, et al. 2018. "The Encyclopedia of DNA Elements (ENCODE): Data Portal Update." *Nucleic Acids Research* 46 (D1): D794–801.
- Deloria, Abigail J., Doris Höflmayer, Philip Kienzl, Justyna Łopatecka, Sandra Sampl, Martin Klimpfinger, Tamara Braunschmid, et al. 2016. "Epithelial Splicing Regulatory Protein 1 and 2 Paralogues Correlate with Splice Signatures and Favorable Outcome in Human Colorectal Cancer." *Oncotarget* 7 (45): 73800–816.
- Desterro, Joana, Pedro Bak-Gordon, and Maria Carmo-Fonseca. 2020. "Targeting mRNA Processing as an Anticancer Strategy." *Nature Reviews. Drug Discovery* 19 (2): 112–29.

- Ding, Lihua, Jinghua Yan, Jianhua Zhu, Hongjun Zhong, Qiujun Lu, Zonghua Wang, Cuifen Huang, and Qinong Ye. 2003. "Ligand-Independent Activation of Estrogen Receptor Alpha by XBP-1." *Nucleic Acids Research* 31 (18): 5266–74.
- Dizin, Eva, and Irmgard Irminger-Finger. 2010. "Negative Feedback Loop of BRCA1-BARD1 Ubiquitin Ligase on Estrogen Receptor Alpha Stability and Activity Antagonized by Cancer-Associated Isoform of BARD1." *The International Journal of Biochemistry & Cell Biology* 42 (5): 693–700.
- Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. "STAR: Ultrafast Universal RNA-Seq Aligner." *Bioinformatics* 29 (1): 15–21.
- Dominguez, Daniel, Peter Freese, Maria S. Alexis, Amanda Su, Myles Hochman, Tsultrim Palden, Cassandra Bazile, et al. 2018. "Sequence, Structure, and Context Preferences of Human RNA Binding Proteins." *Molecular Cell* 70 (5): 854–67.e9.
- Dominguez, Daniel, Yi-Hsuan Tsai, Robert Weatheritt, Yang Wang, Benjamin J. Blencowe, and Zefeng Wang. 2016. "An Extensive Program of Periodic Alternative Splicing Linked to Cell Cycle Progression." *eLife* 5 (March). <https://doi.org/10.7554/eLife.10288>.
- Dowhan, Dennis H., Eugene P. Hong, Didier Auboeuf, Andrew P. Dennis, Michelle M. Wilson, Susan M. Berget, and Bert W. O'Malley. 2005. "Steroid Hormone Receptor Coactivation and Alternative RNA Splicing by U2AF65-Related Proteins CAPERalpha and CAPERbeta." *Molecular Cell* 17 (3): 429–39.
- Dreyfuss, Gideon, V. Narry Kim, and Naoyuki Kataoka. 2002. "Messenger-RNA-Binding Proteins and the Messages They Carry." *Nature Reviews. Molecular Cell Biology* 3 (3): 195–205.
- Dutertre, Martin, and Carolyn L. Smith. 2003. "Ligand-Independent Interactions of p160/steroid Receptor Coactivators and CREB-Binding Protein (CBP) with Estrogen Receptor-Alpha: Regulation by Phosphorylation Sites in the A/B Region Depends on Other Receptor Domains." *Molecular Endocrinology* 17 (7): 1296–1314.
- Dzięgielewska, Żaneta, and Małgorzata Gajewska. 2019. "Stromal-Epithelial Interactions during Mammary Gland Development." *Stromal Cells - Structure, Function, and Therapeutic Implications*. <https://doi.org/10.5772/intechopen.80405>.
- Echeverria, Pablo C., and Didier Picard. 2010. "Molecular Chaperones, Essential Partners of Steroid Hormone Receptors for Activity and Mobility." *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*. <https://doi.org/10.1016/j.bbamcr.2009.11.012>.
- Eiden, Caroline, and Hendrik Ungefroren. 2021. "The Ratio of RAC1B to RAC1 Expression in Breast Cancer Cell Lines as a Determinant of Epithelial/Mesenchymal Differentiation and Migratory Potential." *Cells* 10 (2). <https://doi.org/10.3390/cells10020351>.
- El-Gebali, Sara, Jaina Mistry, Alex Bateman, Sean R. Eddy, Aurélien Luciani, Simon C. Potter, Matloob Qureshi, et al. 2019. "The Pfam Protein Families Database in 2019." *Nucleic Acids Research* 47 (D1): D427–32.



- Elhasnaoui, Jamal, Valentina Miano, Giulio Ferrero, Elena Doria, Antonette E. Leon, Aline S. C. Fabricio, Laura Annaratone, Isabella Castellano, Anna Sapino, and Michele De Bortoli. 2020. "DSCAM-AS1-Driven Proliferation of Breast Cancer Cells Involves Regulation of Alternative Exon Splicing and 3'-End Usage." *Cancers*. <https://doi.org/10.3390/cancers12061453>.
- Elsharawy, Khlood A., Omar J. Mohammed, Mohammed A. Aleskandarany, Ayman Hyder, Hekmat L. El-Gammal, Mohamed I. Abou-Dobara, Andrew R. Green, Leslie W. Dalton, and Emad A. Rakha. 2020. "The Nucleolar-Related Protein Dyskerin Pseudouridine Synthase 1 (DKC1) Predicts Poor Prognosis in Breast Cancer." *British Journal of Cancer* 123 (10): 1543–52.
- ENCODE Project Consortium. 2012. "An Integrated Encyclopedia of DNA Elements in the Human Genome." *Nature* 489 (7414): 57–74.
- Enmark, E., M. Peltö-Huikko, K. Grandien, S. Lagercrantz, J. Lagercrantz, G. Fried, M. Nordenskjöld, and J. A. Gustafsson. 1997. "Human Estrogen Receptor Beta-Gene Structure, Chromosomal Localization, and Expression Pattern." *The Journal of Clinical Endocrinology and Metabolism* 82 (12): 4258–65.
- Erkelenz, S., W. F. Mueller, M. S. Evans, A. Busch, K. Schoneweis, K. J. Hertel, and H. Schaal. 2013. "Position-Dependent Splicing Activation and Repression by SR and hnRNP Proteins Rely on Common Mechanisms." *RNA*. <https://doi.org/10.1261/rna.037044.112>.
- Eswaran, Jeyanthi, Anelia Horvath, Sucheta Godbole, Sirigiri Divijendra Reddy, Prakriti Mudvari, Kazufumi Ohshiro, Dinesh Cyanam, et al. 2013. "RNA Sequencing of Cancer Reveals Novel Splicing Alterations." *Scientific Reports* 3: 1689.
- Fang, Zheng, Junfang Zhao, Weihong Xie, Qiang Sun, Haibin Wang, and Bin Qiao. 2017. "LncRNA UCA1 Promotes Proliferation and Cisplatin Resistance of Oral Squamous Cell Carcinoma by Suppressing miR-184 Expression." *Cancer Medicine* 6 (12): 2897–2908.
- Fan, Meiyun, Pearly S. Yan, Cori Hartman-Frey, Lei Chen, Henry Paik, Samuel L. Oyer, Jonathan D. Salisbury, et al. 2006. "Diverse Gene Expression and DNA Methylation Profiles Correlate with Differential Adaptation of Breast Cancer Cells to the Antiestrogens Tamoxifen and Fulvestrant." *Cancer Research* 66 (24): 11954–66.
- Fan, Songqing, Yikun Li, Ping Yue, Fadlo R. Khuri, and Shi-Yong Sun. 2010. "The eIF4E/eIF4G Interaction Inhibitor 4EGI-1 Augments TRAIL-Mediated Apoptosis through c-FLIP Down-Regulation and DR5 Induction Independent of Inhibition of Cap-Dependent Protein Translation." *Neoplasia* 12 (4): 346–56.
- Fan, Zhenjiang, Soyeon Kim, Yulong Bai, Brenda Diergaarde, and Hyun Jung Park. 2020. "3'-UTR Shortening Contributes to Subtype-Specific Cancer Growth by Breaking Stable ceRNA Crosstalk of Housekeeping Genes." *Frontiers in Bioengineering and Biotechnology* 8 (April): 334.
- Fatica, Alessandro, and Irene Bozzoni. 2014. "Long Non-Coding RNAs: New Players in Cell Differentiation and Development." *Nature Reviews. Genetics* 15 (1): 7–21.
- Fei, Teng, Yiwen Chen, Tengfei Xiao, Wei Li, Laura Cato, Peng Zhang, Maura B. Cotter, et al. 2017. "Genome-Wide CRISPR Screen Identifies HNRNPL as a Prostate Cancer Dependency Regulating RNA

- Splicing.” *Proceedings of the National Academy of Sciences of the United States of America* 114 (26): E5207–15.
- Feng, Huijuan, Zhiyi Qin, and Xuegong Zhang. 2013. “Opportunities and Methods for Studying Alternative Splicing in Cancer with RNA-Seq.” *Cancer Letters*. <https://doi.org/10.1016/j.canlet.2012.11.010>.
- Feng, Y., D. Manka, K-U Wagner, and S. A. Khan. 2007. “Estrogen Receptor- Expression in the Mammary Epithelium Is Required for Ductal and Alveolar Morphogenesis in Mice.” *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.0706933104>.
- Ferlay, Jacques, Isabelle Soerjomataram, Rajesh Dikshit, Sultan Eser, Colin Mathers, Marise Rebelo, Donald Maxwell Parkin, David Forman, and Freddie Bray. 2015. “Cancer Incidence and Mortality Worldwide: Sources, Methods and Major Patterns in GLOBOCAN 2012.” *International Journal of Cancer. Journal International Du Cancer* 136 (5): E359–86.
- Ferrer-Bonsoms, Juan A., Ignacio Cassol, Pablo Fernández-Acín, Carlos Castilla, Fernando Carazo, and Angel Rubio. 2020. “ISOGO: Functional Annotation of Protein-Coding Splice Variants.” *Scientific Reports*. <https://doi.org/10.1038/s41598-020-57974-z>.
- Ferrero, Giulio, Valentina Miano, Marco Beccuti, Gianfranco Balbo, Michele De Bortoli, and Francesca Cordero. 2017. “Dissecting the Genomic Activity of a Transcriptional Regulator by the Integrative Analysis of Omics Data.” *Scientific Reports* 7 (1): 8564.
- Ferro, Paola, Alessandra Forlani, Marco Muselli, and Ulrich Pfeffer. 2003. “Alternative Splicing of the Human Estrogen Receptor  $\alpha$  Primary Transcript: Mechanisms of Exon Skipping.” *International Journal of Molecular Medicine*. <https://doi.org/10.3892/ijmm.12.3.355>.
- Filardo, Edward J., and Peter Thomas. 2012. “Minireview: G Protein-Coupled Estrogen Receptor-1, GPER-1: Its Mechanism of Action and Role in Female Reproductive Cancer, Renal and Vascular Physiology.” *Endocrinology* 153 (7): 2953–62.
- Finlay-Schultz, Jessica, Austin E. Gillen, Heather M. Brechbuhl, Joshua J. Ivie, Shawna B. Matthews, Britta M. Jacobsen, David L. Bentley, Peter Kabos, and Carol A. Sartorius. 2017. “Breast Cancer Suppression by Progesterone Receptors Is Mediated by Their Modulation of Estrogen Receptors and RNA Polymerase III.” *Cancer Research* 77 (18): 4934–46.
- Finn, Robert D., Alex Bateman, Jody Clements, Penelope Coghill, Ruth Y. Eberhardt, Sean R. Eddy, Andreas Heger, et al. 2014. “Pfam: The Protein Families Database.” *Nucleic Acids Research* 42 (Database issue): D222–30.
- Fish, Lisa, Nora Pencheva, Hani Goodarzi, Hien Tran, Mitsukuni Yoshida, and Sohail F. Tavazoie. 2016. “Muscleblind-like 1 Suppresses Breast Cancer Metastatic Colonization and Stabilizes Metastasis Suppressor Transcripts.” *Genes & Development* 30 (4): 386–98.
- Fiszbein, Ana, Keegan S. Krick, Bridget E. Begg, and Christopher B. Burge. 2019. “Exon-Mediated Activation of Transcription Starts.” *Cell* 179 (7): 1551–65.e17.
- Flouriot, G., H. Brand, S. Denger, R. Metivier, M. Kos, G. Reid, V. Sonntag-Buck, and F. Gannon. 2000. “Identification of a New Isoform of the Human Estrogen Receptor-Alpha (hER-Alpha) That Is Encoded by

- Distinct Transcripts and That Is Able to Repress hER-Alpha Activation Function 1.” *The EMBO Journal* 19 (17): 4688–4700.
- Forsburg, Susan L. 2004. “Eukaryotic MCM Proteins: Beyond Replication Initiation.” *Microbiology and Molecular Biology Reviews: MMBR* 68 (1): 109–31.
- Fortunati, N., S. Bertino, L. Costantino, M. De Bortoli, A. Compagnone, A. Bandino, M. G. Catalano, and G. Boccuzzi. 2010. “Valproic Acid Restores ER Alpha and Antiestrogen Sensitivity to ER Alpha-Negative Breast Cancer Cells.” *Molecular and Cellular Endocrinology* 314 (1): 17–22.
- Fougner, Christian, Helga Bergholtz, Jens Henrik Norum, and Therese Sørli. 2020. “Re-Definition of Claudin-Low as a Breast Cancer Phenotype.” *Nature Communications* 11 (1): 1787.
- Fox, Emily M., Josefa Andrade, and Margaret A. Shupnik. 2009. “Novel Actions of Estrogen to Promote Proliferation: Integration of Cytoplasmic and Nuclear Pathways.” *Steroids* 74 (7): 622–27.
- Frankiw, Luke, David Baltimore, and Guideng Li. 2019. “Alternative mRNA Splicing in Cancer Immunotherapy.” *Nature Reviews. Immunology* 19 (11): 675–87.
- Froussios, Kimon, Kira Mourão, Gordon Simpson, Geoff Barton, and Nicholas Schurch. 2019. “Relative Abundance of Transcripts ( ): Identifying Differential Isoform Abundance from RNA-Seq.” *F1000Research* 8 (February): 213.
- Fuentes, Nathalie, and Patricia Silveyra. 2019. “Estrogen Receptor Signaling Mechanisms.” *Advances in Protein Chemistry and Structural Biology* 116 (February): 135–70.
- Furth, Priscilla A. 2014. “STAT Signaling in Different Breast Cancer Sub-Types.” *Molecular and Cellular Endocrinology* 382 (1): 612–15.
- Fu, Xiang-Dong, and Manuel Ares Jr. 2014. “Context-Dependent Control of Alternative Splicing by RNA-Binding Proteins.” *Nature Reviews. Genetics* 15 (10): 689–701.
- Fu, Xing, Ming Tian, Jia Gu, Teng Cheng, Ding Ma, Ling Feng, and Xing Xin. 2017. “SF3B1 Mutation Is a Poor Prognostic Indicator in Luminal B and Progesterone Receptor-Negative Breast Cancer Patients.” *Oncotarget*. <https://doi.org/10.18632/oncotarget.22983>.
- Gabut, Mathieu, Payman Samavarchi-Tehrani, Xinchun Wang, Valentina Slobodeniuc, Dave O’Hanlon, Hoon-Ki Sung, Manuel Alvarez, et al. 2011. “An Alternative Splicing Switch Regulates Embryonic Stem Cell Pluripotency and Reprogramming.” *Cell* 147 (1): 132–46.
- Gallego-Paez, L. M., M. C. Bordone, A. C. Leote, N. Saraiva-Agostinho, M. Ascensão-Ferreira, and N. L. Barbosa-Morais. 2017. “Alternative Splicing: The Pledge, the Turn, and the Prestige : The Key Role of Alternative Splicing in Human Biological Systems.” *Human Genetics* 136 (9): 1015–42.
- Gao, J., B. A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S. O. Sumer, Y. Sun, et al. 2013. “Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal.” *Science Signaling*. <https://doi.org/10.1126/scisignal.2004088>.
- Garbuglia, M., M. Verzini, A. Hofmann, R. Huber, and R. Donato. 2000. “S100A1 and S100B Interactions with Annexins.” *Biochimica et Biophysica Acta* 1498 (2-3): 192–206.

- Garrido-Martín, Diego, Emilio Palumbo, Roderic Guigó, and Alessandra Breschi. 2018. “Ggsashimi: Sashimi Plot Revised for Browser- and Annotation-Independent Splicing Visualization.” *PLoS Computational Biology* 14 (8): e1006360.
- Gates, Devika P., Leslie A. Coonrod, and J. Andrew Berglund. 2011. “Autoregulated Splicing of Muscleblind-like 1 (MBNL1) Pre-mRNA.” *The Journal of Biological Chemistry* 286 (39): 34224–33.
- Gaudreau, Marie-Claude, Damien Grapton, Anne Helness, Charles Vadnais, Jennifer Fraszczak, Peiman Shooshtarizadeh, Brian Wilhelm, François Robert, Florian Heyd, and Tarik Möröy. 2016. “Heterogeneous Nuclear Ribonucleoprotein L Is Required for the Survival and Functional Integrity of Murine Hematopoietic Stem Cells.” *Scientific Reports* 6 (June): 27379.
- Gawronski, Alexander R., Michael Uhl, Yajia Zhang, Yen-Yi Lin, Yashar S. Niknafs, Varune R. Ramnarine, Rohit Malik, et al. 2018. “MechRNA: Prediction of lncRNA Mechanisms from RNA–RNA and RNA–protein Interactions.” *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/bty208>.
- Geisler, Sarah, and Jeff Collier. 2013. “RNA in Unexpected Places: Long Non-Coding RNA Functions in Diverse Cellular Contexts.” *Nature Reviews Molecular Cell Biology*. <https://doi.org/10.1038/nrm3679>.
- Gerstberger, Stefanie, Markus Hafner, and Thomas Tuschl. 2014. “A Census of Human RNA-Binding Proteins.” *Nature Reviews. Genetics* 15 (12): 829–45.
- Gilfillan, Siv, Elisa Fiorito, and Antoni Hurtado. 2012. “Functional Genomic Methods to Study Estrogen Receptor Activity.” *Journal of Mammary Gland Biology and Neoplasia*. <https://doi.org/10.1007/s10911-012-9254-4>.
- Gil, Noa, and Igor Ulitsky. 2020. “Regulation of Gene Expression by Cis-Acting Long Non-Coding RNAs.” *Nature Reviews. Genetics* 21 (2): 102–17.
- Giulianelli, Sebastián, Marina Riggio, Tomas Guillardoy, Cecilia Pérez Piñero, María A. Gorostiaga, Gonzalo Sequeira, Gabriela Pataccini, et al. 2019. “FGF2 Induces Breast Cancer Growth through Ligand-Independent Activation and Recruitment of ER $\alpha$  and PRB $\Delta$ 4 Isoform to MYC Regulatory Sequences.” *International Journal of Cancer. Journal International Du Cancer* 145 (7): 1874–88.
- Gökmen-Polar, Yesim, Yaseswini Neelamraju, Chirayu P. Goswami, Yuan Gu, Xiaoping Gu, Gouthami Nallamothu, Edyta Vieth, Sarath C. Janga, Michael Ryan, and Sunil S. Badve. 2019. “Splicing Factor ESRP 1 Controls ER -positive Breast Cancer by Altering Metabolic Pathways.” *EMBO Reports*. <https://doi.org/10.15252/embr.201846078>.
- Gökmen-Polar, Yesim, Yaseswini Neelamraju, Chirayu P. Goswami, Yuan Gu, Xiaoping Gu, Gouthami Nallamothu, Edyta Vieth, Sarath C. Janga, Michael Ryan, and Sunil S. Badve. 2019. “Splicing Factor Controls ER-Positive Breast Cancer by Altering Metabolic Pathways.” *EMBO Reports* 20 (2). <https://doi.org/10.15252/embr.201846078>.
- Golan-Gerstl, Regina, Michal Cohen, Asaf Shilo, Sung-Suk Suh, Arianna Bakàcs, Luigi Coppola, and Rotem Karni. 2011. “Splicing Factor hnRNP A2/B1 Regulates Tumor Suppressor Gene Splicing and Is an Oncogenic Driver in Glioblastoma.” *Cancer Research* 71 (13): 4464–72.

- Gonçalves, Vânia, Andreia F. A. Henriques, Joana F. S. Pereira, Ana Neves Costa, Mary Pat Moyer, Luís Ferreira Moita, Margarida Gama-Carvalho, Paulo Matos, and Peter Jordan. 2014. “Phosphorylation of SRSF1 by SRPK1 Regulates Alternative Splicing of Tumor-Related Rac1b in Colorectal Cells.” *RNA* 20 (4): 474–82.
- Gonzalez, Inma, Roberto Munita, Eneritz Agirre, Travis A. Dittmer, Katia Gysling, Tom Misteli, and Reini F. Lucio. 2015. “A lncRNA Regulates Alternative Splicing via Establishment of a Splicing-Specific Chromatin Signature.” *Nature Structural & Molecular Biology* 22 (5): 370–76.
- González, L., A. Zambrano, I. Lazaro-Trueba, E. Lopéz, J. J. A. González, J. Martín-Pérez, and A. Aranda. 2009. “Activation of the Unliganded Estrogen Receptor by Prolactin in Breast Cancer Cells.” *Oncogene* 28 (10): 1298–1308.
- Gosden, J. R., P. G. Middleton, and D. Rout. 1986. “Localization of the Human Oestrogen Receptor Gene to Chromosome 6q24---q27 by in Situ Hybridization.” *Cytogenetics and Cell Genetics* 43 (3-4): 218–20.
- Grant, Charles E., Timothy L. Bailey, and William Stafford Noble. 2011. “FIMO: Scanning for Occurrences of a given Motif.” *Bioinformatics* 27 (7): 1017–18.
- Green, Andrew R., Mohammed A. Aleskandarany, Devika Agarwal, Somaia Elsheikh, Christopher C. Nolan, Maria Diez-Rodriguez, R. Douglas Macmillan, et al. 2016. “MYC Functions Are Specific in Biological Subtypes of Breast Cancer and Confers Resistance to Endocrine Therapy in Luminal Tumours.” *British Journal of Cancer* 114 (8): 917–28.
- Green, Kelly A., and Jason S. Carroll. 2007. “Oestrogen-Receptor-Mediated Transcription and the Influence of Co-Factors and Chromatin State.” *Nature Reviews. Cancer* 7 (9): 713–22.
- Grisouard, Jean, Senad Medunjanin, Alexander Hermani, Ashish Shukla, and Doris Mayer. 2007. “Glycogen Synthase Kinase-3 Protects Estrogen Receptor  $\alpha$  from Proteasomal Degradation and Is Required for Full Transcriptional Activity of the Receptor.” *Molecular Endocrinology*. <https://doi.org/10.1210/me.2007-0129>.
- Grossi, Elena, Ivan Raimondi, Enrique Goñi, Jovanna González, Francesco P. Marchese, Vicente Chapaprieta, José I. Martín-Subero, Shuling Guo, and Maite Huarte. 2020. “A lncRNA-SWI/SNF Complex Crosstalk Controls Transcriptional Activation at Specific Promoter Regions.” *Nature Communications*. <https://doi.org/10.1038/s41467-020-14623-3>.
- Grossman, Robert L., Allison P. Heath, Vincent Ferretti, Harold E. Varmus, Douglas R. Lowy, Warren A. Kibbe, and Louis M. Staudt. 2016a. “Toward a Shared Vision for Cancer Genomic Data.” *New England Journal of Medicine*. <https://doi.org/10.1056/nejmp1607591>.
- . 2016b. “Toward a Shared Vision for Cancer Genomic Data.” *The New England Journal of Medicine* 375 (12): 1109–12.
- Guerrieri, Ania Naila, Federico Zacchini, Carmine Onofrillo, Sara Di Viggiano, Marianna Penzo, Alessio Ansuini, Iliaria Gandin, et al. 2020. “DKC1 Overexpression Induces a More Aggressive Cellular Behavior and Increases Intrinsic Ribosomal Activity in Immortalized Mammary Gland Cells.” *Cancers* 12 (12). <https://doi.org/10.3390/cancers12123512>.

- Guo, Chun-Jie, Xu-Kai Ma, Yu-Hang Xing, Chuan-Chuan Zheng, Yi-Feng Xu, Lin Shan, Jun Zhang, et al. 2020. “Distinct Processing of lncRNAs Contributes to Non-Conserved Functions in Stem Cells.” *Cell*. <https://doi.org/10.1016/j.cell.2020.03.006>.
- Gupta, Arun A., Ruey-Hwang Chou, Hongchun Li, Lee-Wei Yang, and Chin Yu. 2013. “Structural Insights into the Interaction of Human S100B and Basic Fibroblast Growth Factor (FGF2): Effects on FGFR1 Receptor Signaling.” *Biochimica et Biophysica Acta* 1834 (12): 2606–19.
- Gu, Siwen, Chengyu Chu, Wanna Chen, Hong Ren, Yun Cao, Xiaoyan Li, Jing He, et al. 2019. “Prognostic Value of Epithelial-Mesenchymal Transition Related Genes: SLUG and QKI in Breast Cancer Patients.” *International Journal of Clinical and Experimental Pathology* 12 (6): 2009–21.
- Guttilla, Irene K., Brian D. Adams, and Bruce A. White. 2012. “ER $\alpha$ , microRNAs, and the Epithelial-Mesenchymal Transition in Breast Cancer.” *Trends in Endocrinology and Metabolism: TEM* 23 (2): 73–82.
- Gu, Yu, Tianxiang Chen, Elena López, Weizhu Wu, Xiangdong Wang, Jiang Cao, and Lisong Teng. 2014. “The Therapeutic Target of Estrogen Receptor- $\alpha$ 36 in Estrogen-Dependent Tumors.” *Journal of Translational Medicine* 12 (January): 16.
- Hamid, Fursham M., and Eugene V. Makeyev. 2017. “A Mechanism Underlying Position-Specific Regulation of Alternative Splicing.” *Nucleic Acids Research* 45 (21): 12455–68.
- Hanahan, Douglas, and Robert A. Weinberg. 2011. “Hallmarks of Cancer: The next Generation.” *Cell* 144 (5): 646–74.
- Han, Hong, Manuel Irimia, P. Joel Ross, Hoon-Ki Sung, Babak Alipanahi, Laurent David, Azadeh Golipour, et al. 2013. “MBNL Proteins Repress ES-Cell-Specific Alternative Splicing and Reprogramming.” *Nature* 498 (7453): 241–45.
- Harrell, Joshua Chuck, Wendy W. Dye, D. Craig Allred, Paul Jedlicka, Nicole S. Spoelstra, Carol A. Sartorius, and Kathryn B. Horwitz. 2006. “Estrogen Receptor Positive Breast Cancer Metastasis: Altered Hormonal Sensitivity and Tumor Aggressiveness in Lymphatic Vessels and Lymph Nodes.” *Cancer Research* 66 (18): 9308–15.
- Hartman, Philip E. 1959. “The Biological Replication of Macromolecules. Symposia of the Society for Experimental Biology, No. XII.” *The Quarterly Review of Biology*. <https://doi.org/10.1086/402623>.
- Harvey, Samuel E., Yilin Xu, Xiaodan Lin, Xin D. Gao, Yushan Qiu, Jaegyeon Ahn, Xinshu Xiao, and Chonghui Cheng. 2018. “Coregulation of Alternative Splicing by hnRNPM and ESRP1 during EMT.” *RNA* 24 (10): 1326–38.
- Hattori, Daisuke, Yi Chen, Benjamin J. Matthews, Lukasz Salwinski, Chiara Sabatti, Wesley B. Grueber, and S. Lawrence Zipursky. 2009. “Robust Discrimination between Self and Non-Self Neurites Requires Thousands of Dscam1 Isoforms.” *Nature* 461 (7264): 644–48.
- Hayakawa, Akira, Masao Saitoh, and Keiji Miyazawa. 2016. “Dual Roles for Epithelial Splicing Regulatory Proteins 1 (ESRP1) and 2 (ESRP2) in Cancer Progression.” *Advances in Experimental Medicine and Biology*. [https://doi.org/10.1007/5584\\_2016\\_50](https://doi.org/10.1007/5584_2016_50).

- Hegele, Anna, Atanas Kamburov, Arndt Grossmann, Chrysovalantis Sourlis, Sylvia Wowro, Mareike Weimann, Cindy L. Will, Vlad Pena, Reinhard Lührmann, and Ulrich Stelzl. 2012. “Dynamic Protein-Protein Interaction Wiring of the Human Spliceosome.” *Molecular Cell*. <https://doi.org/10.1016/j.molcel.2011.12.034>.
- Heiner, Monika, Jingyi Hui, Silke Schreiner, Lee-Hsueh Hung, and Albrecht Bindereif. 2010. “HnRNP L-Mediated Regulation of Mammalian Alternative Splicing by Interference with Splice Site Recognition.” *RNA Biology* 7 (1): 56–64.
- Heo, Yoonyoung, Hye-Jin Yoon, Hanseo Ko, Soonmin Jang, and Hyung Ho Lee. 2020. “Crystal Structures of Uso1 Membrane Tether Reveal an Alternative Conformation in the Globular Head Domain.” *Scientific Reports* 10 (1): 9544.
- He, Rong-Zhang, Di-Xian Luo, and Yin-Yuan Mo. 2019. “Emerging Roles of lncRNAs in the Post-Transcriptional Regulation in Cancer.” *Genes & Diseases*. <https://doi.org/10.1016/j.gendis.2019.01.003>.
- Hewitt, Sylvia C., Sara A. Grimm, San-Pin Wu, Francesco J. DeMayo, and Kenneth S. Korach. 2020. “Estrogen Receptor  $\alpha$  (ER $\alpha$ )-Binding Super-Enhancers Drive Key Mediators That Control Uterine Estrogen Responses in Mice.” *The Journal of Biological Chemistry* 295 (25): 8387–8400.
- Hewitt, Sylvia C., and Kenneth S. Korach. 2018. “Estrogen Receptors: New Directions in the New Millennium.” *Endocrine Reviews* 39 (5): 664–75.
- Hewitt, Sylvia C., Leping Li, Sara A. Grimm, Yu Chen, Liwen Liu, Yin Li, Pierre R. Bushel, David Fargo, and Kenneth S. Korach. 2012. “Research Resource: Whole-Genome Estrogen Receptor  $\alpha$  Binding in Mouse Uterine Tissue Revealed by ChIP-Seq.” *Molecular Endocrinology* 26 (5): 887–98.
- Hir, H. Le, and H. Le Hir. 2001. “The Exon-Exon Junction Complex Provides a Binding Platform for Factors Involved in mRNA Export and Nonsense-Mediated mRNA Decay.” *The EMBO Journal*. <https://doi.org/10.1093/emboj/20.17.4987>.
- Holdt, Lesca M., Steve Hoffmann, Kristina Sass, David Langenberger, Markus Scholz, Knut Krohn, Knut Finstermeier, et al. 2013. “Alu Elements in ANRIL Non-Coding RNA at Chromosome 9p21 Modulate Atherogenic Cell Functions through Trans-Regulation of Gene Networks.” *PLoS Genetics* 9 (7): e1003588.
- Holm, C., M. Kok, R. Michalides, R. Fles, R. H. T. Koornstra, J. Wesseling, M. Hauptmann, et al. 2009. “Phosphorylation of the Oestrogen Receptor  $\alpha$  at Serine 305 and Prediction of Tamoxifen Resistance in Breast Cancer.” *The Journal of Pathology*. <https://doi.org/10.1002/path.2455>.
- Horiguchi, K., K. Sakamoto, D. Koinuma, K. Semba, A. Inoue, S. Inoue, H. Fujii, et al. 2012. “TGF- $\beta$  Drives Epithelial-Mesenchymal Transition through  $\delta$ EF1-Mediated Downregulation of ESRP.” *Oncogene* 31 (26): 3190–3201.
- Huang, Huilin, Yilin Xu, and Chonghui Cheng. 2014. “Detection of Alternative Splicing During Epithelial-Mesenchymal Transition.” *Journal of Visualized Experiments*. <https://doi.org/10.3791/51845>.
- Hua, Sujun, Caleb B. Kallen, Ruby Dhar, Maria T. Baquero, Christopher E. Mason, Beth A. Russell, Parantu K. Shah, et al. 2008. “Genomic Analysis of Estrogen Cascade Reveals Histone Variant H2A.Z Associated with Breast Cancer Progression.” *Molecular Systems Biology*. <https://doi.org/10.1038/msb.2008.25>.

- Huelga, Stephanie C., Anthony Q. Vu, Justin D. Arnold, Tiffany Y. Liang, Patrick P. Liu, Bernice Y. Yan, John Paul Donohue, et al. 2012. "Integrative Genome-Wide Analysis Reveals Cooperative Regulation of Alternative Splicing by hnRNP Proteins." *Cell Reports* 1 (2): 167–78.
- Hung, L-H, M. Heiner, J. Hui, S. Schreiner, V. Benes, and A. Bindereif. 2007. "Diverse Roles of hnRNP L in Mammalian mRNA Processing: A Combined Microarray and RNAi Analysis." *RNA*. <https://doi.org/10.1261/ma.725208>.
- Hurtado, Antoni, Kelly A. Holmes, Caryn S. Ross-Innes, Dominic Schmidt, and Jason S. Carroll. 2011. "FOXA1 Is a Key Determinant of Estrogen Receptor Function and Endocrine Response." *Nature Genetics* 43 (1): 27–33.
- Huttlin, Edward L., Raphael J. Bruckner, Joao A. Paulo, Joe R. Cannon, Lily Ting, Kurt Baltier, Greg Colby, et al. 2017. "Architecture of the Human Interactome Defines Protein Communities and Disease Networks." *Nature* 545 (7655): 505–9.
- Huttlin, Edward L., Lily Ting, Raphael J. Bruckner, Fana Gebreab, Melanie P. Gygi, John Szpyt, Stanley Tam, et al. 2015. "The BioPlex Network: A Systematic Exploration of the Human Interactome." *Cell*. <https://doi.org/10.1016/j.cell.2015.06.043>.
- Hu, Ying, Zihan Sun, Jinmu Deng, Baoquan Hu, Wenting Yan, Hongyi Wei, and Jun Jiang. 2017. "Splicing Factor hnRNPA2B1 Contributes to Tumorigenic Potential of Breast Cancer Cells through STAT3 and ERK1/2 Signaling Pathway." *Tumour Biology: The Journal of the International Society for Oncodevelopmental Biology and Medicine* 39 (3): 1010428317694318.
- Ince, B. A., M. M. Montano, and B. S. Katzenellenbogen. 1994. "Activation of Transcriptionally Inactive Human Estrogen Receptors by Cyclic Adenosine 3',5'-Monophosphate and Ligands Including Antiestrogens." *Molecular Endocrinology* 8 (10): 1397–1406.
- Inman, Jamie L., Claire Robertson, Joni D. Mott, and Mina J. Bissell. 2015. "Mammary Gland Development: Cell Fate Specification, Stem Cells and the Microenvironment." *Development* 142 (6): 1028–42.
- Inoue, Kazushi, and Elizabeth A. Fry. 2015. "Aberrant Splicing of Estrogen Receptor, HER2, and CD44 Genes in Breast Cancer." *Genetics & Epigenetics* 7 (December): 19–32.
- Ishii, Hiroki, Masao Saitoh, Kei Sakamoto, Tetsuo Kondo, Ryohei Katoh, Shota Tanaka, Mitsuyoshi Motizuki, Keisuke Masuyama, and Keiji Miyazawa. 2014. "Epithelial Splicing Regulatory Proteins 1 (ESRP1) and 2 (ESRP2) Suppress Cancer Cell Motility via Different Mechanisms." *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.m114.589432>.
- Issac, Marianne Samir Makboul, Einas Yousef, Muhammad Ramzan Tahir, and Louis A. Gaboury. 2019. "MCM2, MCM4, and MCM6 in Breast Cancer: Clinical Utility in Diagnosis and Prognosis." *Neoplasia* 21 (10): 1015–35.
- Jaarsveld, Marijn T. M. van, Difan Deng, Diana Ordoñez-Rueda, Malte Paulsen, Erik A. C. Wiemer, and Zhike Zi. 2020. "Cell-Type-Specific Role of CHK2 in Mediating DNA Damage-Induced G2 Cell Cycle Arrest." *Oncogenesis* 9 (3): 35.



- Jackson, Claire, David Browell, Hannah Gautrey, and Alison Tyson-Capper. 2013. "Clinical Significance of HER-2 Splice Variants in Breast Cancer Progression and Drug Resistance." *International Journal of Cell Biology*. <https://doi.org/10.1155/2013/973584>.
- Jain, Abhinav K., Yuanxin Xi, Ryan McCarthy, Kendra Allton, Kadir C. Akdemir, Lalit R. Patel, Bruce Aronow, et al. 2016. "LncPRESS1 Is a p53-Regulated LncRNA That Safeguards Pluripotency by Disrupting SIRT6-Mediated De-Acetylation of Histone H3K56." *Molecular Cell*. <https://doi.org/10.1016/j.molcel.2016.10.039>.
- Jensen, E. V., T. Suzuki, T. Kawashima, W. E. Stumpf, P. W. Jungblut, and E. R. DeSombre. 1968. "A Two-Step Mechanism for the Interaction of Estradiol with Rat Uterus." *Proceedings of the National Academy of Sciences of the United States of America* 59 (2): 632–38.
- Jeong, H. M., J. Han, S. H. Lee, H-J Park, H. J. Lee, J-S Choi, Y. M. Lee, Y-L Choi, Y. K. Shin, and M. J. Kwon. 2017. "ESRP1 Is Overexpressed in Ovarian Cancer and Promotes Switching from Mesenchymal to Epithelial Phenotype in Ovarian Cancer Cells." *Oncogenesis* 6 (11): e391.
- Jiang, Lihua, Meng Wang, Shin Lin, Ruiqi Jian, Xiao Li, Joanne Chan, Guanlan Dong, et al. 2020. "A Quantitative Proteome Map of the Human Body." *Cell* 183 (1): 269–83.e19.
- Jia, Rong, Si Zhang, Miaomiao Liu, Yan Zhang, Yu Liu, Mingwen Fan, and Jihua Guo. 2016. "HnRNP L Is Important for the Expression of Oncogene SRSF3 and Oncogenic Potential of Oral Squamous Cell Carcinoma Cells." *Scientific Reports* 6 (November): 35976.
- Ji, Degang, Guangrui Hu, Xuanhe Zhang, Tianhua Yu, and Jinghui Yang. 2019. "Long Non-Coding RNA DSCAM-AS1 Accelerates the Progression of Hepatocellular Carcinoma via Sponging miR-338-3p." *American Journal of Translational Research* 11 (7): 4290–4302.
- Jin, X., L. Zhu, Z. Cui, J. Tang, M. Xie, and G. Ren. 2019. "Elevated Expression of GNAS Promotes Breast Cancer Cell Proliferation and Migration via the PI3K/AKT/Snail1/E-Cadherin Axis." *Clinical & Translational Oncology: Official Publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico* 21 (9): 1207–19.
- Ji, Q., L. Zhang, X. Liu, L. Zhou, W. Wang, Z. Han, H. Sui, et al. 2014. "Long Non-Coding RNA MALAT1 Promotes Tumour Growth and Metastasis in Colorectal Cancer through Binding to SFPQ and Releasing Oncogene PTBP2 from SFPQ/PTBP2 Complex." *British Journal of Cancer* 111 (4): 736–48.
- Joel, P. B., J. Smith, T. W. Sturgill, T. L. Fisher, J. Blenis, and D. A. Lannigan. 1998. "pp90rsk1 Regulates Estrogen Receptor-Mediated Transcription through Phosphorylation of Ser-167." *Molecular and Cellular Biology* 18 (4): 1978–84.
- Kang, Christopher Y., Jonathan Wang, Dierdre Axell-House, Pranay Soni, Mon-Li Chu, Galina Chipitsyna, Konrad Sarosiek, et al. 2014. "Clinical Significance of Serum COL6A3 in Pancreatic Ductal Adenocarcinoma." *Journal of Gastrointestinal Surgery: Official Journal of the Society for Surgery of the Alimentary Tract* 18 (1): 7–15.

- Kang, Yu-Jian, De-Chang Yang, Lei Kong, Mei Hou, Yu-Qi Meng, Liping Wei, and Ge Gao. 2017. "CPC2: A Fast and Accurate Coding Potential Calculator Based on Sequence Intrinsic Features." *Nucleic Acids Research* 45 (W1): W12–16.
- Kanitz, Alexander, Foivos Gypas, Andreas J. Gruber, Andreas R. Gruber, Georges Martin, and Mihaela Zavolan. 2015. "Comparative Assessment of Methods for the Computational Inference of Transcript Isoform Abundance from RNA-Seq Data." *Genome Biology* 16 (July): 150.
- Kato, S., H. Endoh, Y. Masuhiro, T. Kitamoto, S. Uchiyama, H. Sasaki, S. Masushige, et al. 1995. "Activation of the Estrogen Receptor Through Phosphorylation by Mitogen-Activated Protein Kinase." *Science*. <https://doi.org/10.1126/science.270.5241.1491>.
- Kelemen, Olga, Paolo Convertini, Zhaiyi Zhang, Yuan Wen, Manli Shen, Marina Falaleeva, and Stefan Stamm. 2013. "Function of Alternative Splicing." *Gene* 514 (1): 1–30.
- Kim, Daehwan, Geo Pertea, Cole Trapnell, Harold Pimentel, Ryan Kelley, and Steven L. Salzberg. 2013. "TopHat2: Accurate Alignment of Transcriptomes in the Presence of Insertions, Deletions and Gene Fusions." *Genome Biology* 14 (4): R36.
- Kim, Eunhee, Janine O. Ilagan, Yang Liang, Gerrit M. Daubner, Stanley C-W Lee, Aravind Ramakrishnan, Yue Li, et al. 2015. "SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition." *Cancer Cell* 27 (5): 617–30.
- Kim, Min-Sik, Sneha M. Pinto, Derese Getnet, Raja Sekhar Nirujogi, Srikanth S. Manda, Raghothama Chaerkady, Anil K. Madugundu, et al. 2014. "A Draft Map of the Human Proteome." *Nature* 509 (7502): 575–81.
- Kirkegaard, Tove, Caroline J. Witton, Liane M. McGlynn, Sian M. Tovey, Barbara Dunne, Alison Lyon, and John M. S. Bartlett. 2005. "AKT Activation Predicts Outcome in Breast Cancer Patients Treated with Tamoxifen." *The Journal of Pathology* 207 (2): 139–46.
- Kishor, Aparna, Zhiyun Ge, and J. Robert Hogg. 2019. "hnRNP L-dependent Protection of Normal mRNAs from NMD Subverts Quality Control in B Cell Lymphoma." *The EMBO Journal*. <https://doi.org/10.15252/emboj.201899128>.
- Kittaneh, Muaiad, Alberto J. Montero, and Stefan Glück. 2013. "Molecular Profiling for Breast Cancer: A Comprehensive Review." *Biomarkers in Cancer*. <https://doi.org/10.4137/bic.s9455>.
- Klerk, Eleonora de, and Peter A. C. 't Hoen. 2015. "Alternative mRNA Transcription, Processing, and Translation: Insights from RNA Sequencing." *Trends in Genetics: TIG* 31 (3): 128–39.
- Klinge, Carolyn M., Kellianne M. Piell, Christine Schaner Tooley, and Eric C. Rouchka. 2019. "HNRNPA2/B1 Is Upregulated in Endocrine-Resistant LCC9 Breast Cancer Cells and Alters the miRNA Transcriptome When Overexpressed in MCF-7 Cells." *Scientific Reports* 9 (1): 9430.
- Klinge, Carolyn M., Krista A. Riggs, Nalinie S. Wickramasinghe, Celia G. Emberts, David B. McConda, Parul N. Barry, and Joan E. Magnusen. 2010. "Estrogen Receptor Alpha 46 Is Reduced in Tamoxifen Resistant Breast Cancer Cells and Re-Expression Inhibits Cell Proliferation and Estrogen Receptor Alpha 66-Regulated

- Target Gene Transcription.” *Molecular and Cellular Endocrinology*. <https://doi.org/10.1016/j.mce.2010.03.013>.
- Klotz, Diane M., Sylvia Curtis Hewitt, Paolo Ciana, Michele Raviscioni, Jonathan K. Lindzey, Julie Foley, Adriana Maggi, Richard P. DiAugustine, and Kenneth S. Korach. 2002. “Requirement of Estrogen Receptor-Alpha in Insulin-like Growth Factor-1 (IGF-1)-Induced Uterine Responses and in Vivo Evidence for IGF-1/estrogen Receptor Cross-Talk.” *The Journal of Biological Chemistry* 277 (10): 8531–37.
- Koedoot, Esmee, Marcel Smid, John A. Foekens, John W. M. Martens, Sylvia E. Le Dévédec, and Bob van de Water. 2019. “Co-Regulated Gene Expression of Splicing Factors as Drivers of Cancer Progression.” *Scientific Reports* 9 (1): 5484.
- Kong, Jianlu, Wenjie Sun, Chen Li, Ledong Wan, Shuo Wang, Yihua Wu, Enping Xu, Honghe Zhang, and Maode Lai. 2016. “Long Non-Coding RNA LINC01133 Inhibits Epithelial-Mesenchymal Transition and Metastasis in Colorectal Cancer by Interacting with SRSF6.” *Cancer Letters* 380 (2): 476–84.
- Kong, Yahui, Chih-Heng Hsieh, and Laura C. Alonso. 2018. “A lncRNA at the CDKN2A/B Locus With Roles in Cancer and Metabolic Disease.” *Frontiers in Endocrinology* 9 (July): 405.
- Kornblihtt, Alberto R., Ignacio E. Schor, Mariano Alló, Gwendal Dujardin, Ezequiel Petrillo, and Manuel J. Muñoz. 2013. “Alternative Splicing: A Pivotal Step between Eukaryotic Transcription and Translation.” *Nature Reviews. Molecular Cell Biology* 14 (3): 153–65.
- Koumangoye, Rainelli B., Gladys N. Nangami, Pamela D. Thompson, Vincent K. Agboto, Josiah Ochieng, and Amos M. Sakwe. 2013. “Reduced Annexin A6 Expression Promotes the Degradation of Activated Epidermal Growth Factor Receptor and Sensitizes Invasive Breast Cancer Cells to EGFR-Targeted Tyrosine Kinase Inhibitors.” *Molecular Cancer* 12 (1): 167.
- Kuhn, R. M., D. Haussler, and W. J. Kent. 2013. “The UCSC Genome Browser and Associated Tools.” *Briefings in Bioinformatics*. <https://doi.org/10.1093/bib/bbs038>.
- Kuiper, G. G., E. Enmark, M. Peltö-Huikko, S. Nilsson, and J. A. Gustafsson. 1996. “Cloning of a Novel Receptor Expressed in Rat Prostate and Ovary.” *Proceedings of the National Academy of Sciences of the United States of America* 93 (12): 5925–30.
- Kumar, Raj, Mikhail N. Zakharov, Shagufta H. Khan, Rika Miki, Hyeran Jang, Gianluca Toraldo, Rajan Singh, Shalender Bhasin, and Ravi Jasuja. 2011. “The Dynamic Structure of the Estrogen Receptor.” *Journal of Amino Acids* 2011 (July): 812540.
- Kuo, Alex J., Peggie Cheung, Kaifu Chen, Barry M. Zee, Mitomu Kioi, Josh Lauring, Yuanxin Xi, et al. 2011. “NSD2 Links Dimethylation of Histone H3 at Lysine 36 to Oncogenic Programming.” *Molecular Cell* 44 (4): 609–20.
- Kwek, Kon Yew, Shona Murphy, Andre Furger, Benjamin Thomas, William O’Gorman, Hiroshi Kimura, Nick J. Proudfoot, and Alexandre Akoulitchev. 2002. “U1 snRNA Associates with TFIIF and Regulates Transcriptional Initiation.” *Nature Structural Biology* 9 (11): 800–805.

- Lagarde, Julien, Barbara Uszczyńska-Ratajczak, Silvia Carbonell, Sílvia Pérez-Lluch, Amaya Abad, Carrie Davis, Thomas R. Gingeras, et al. 2017. “High-Throughput Annotation of Full-Length Long Noncoding RNAs with Capture Long-Read Sequencing.” *Nature Genetics*. <https://doi.org/10.1038/ng.3988>.
- Lal, S., A. Allan, D. Markovic, R. Walker, J. Macartney, N. Europe-Finner, A. Tyson-Capper, and D. K. Grammatopoulos. 2013. “Estrogen Alters the Splicing of Type 1 Corticotropin-Releasing Hormone Receptor in Breast Cancer Cells.” *Science Signaling*. <https://doi.org/10.1126/scisignal.2003926>.
- Lambert, Samuel A., Arttu Jolma, Laura F. Campitelli, Pratyush K. Das, Yimeng Yin, Mihai Albu, Xiaoting Chen, Jussi Taipale, Timothy R. Hughes, and Matthew T. Weirauch. 2018. “The Human Transcription Factors.” *Cell* 175 (2): 598–99.
- Langmead, Ben, and Steven L. Salzberg. 2012. “Fast Gapped-Read Alignment with Bowtie 2.” *Nature Methods* 9 (4): 357–59.
- Lannigan, Deborah A. 2003. “Estrogen Receptor Phosphorylation.” *Steroids* 68 (1): 1–9.
- Lapuk, Anna, Henry Marr, Lakshmi Jakkula, Helder Pedro, Sanchita Bhattacharya, Elizabeth Purdom, Zhi Hu, et al. 2010. “Exon-Level Microarray Analyses Identify Alternative Splicing Programs in Breast Cancer.” *Molecular Cancer Research: MCR* 8 (7): 961–74.
- Lari, Sara A., and Henry M. Kuerer. 2011. “Biological Markers in DCIS and Risk of Breast Recurrence: A Systematic Review.” *Journal of Cancer* 2 (May): 232–61.
- Larsen, Jill E., Vaishnavi Nathan, Jihan K. Osborne, Rebecca K. Farrow, Dhruva Deb, James P. Sullivan, Patrick D. Dospoy, et al. 2016. “ZEB1 Drives Epithelial-to-Mesenchymal Transition in Lung Cancer.” *The Journal of Clinical Investigation* 126 (9): 3219–35.
- Lauber, J., G. Plessel, S. Prehn, C. L. Will, P. Fabrizio, K. Gröning, W. S. Lane, and R. Lührmann. 1997. “The Human U4/U6 snRNP Contains 60 and 90kD Proteins That Are Structurally Homologous to the Yeast Splicing Factors Prp4p and Prp3p.” *RNA* 3 (8): 926–41.
- Leggere, Janelle C., Yuhki Saito, Robert B. Darnell, Marc Tessier-Lavigne, Harald J. Junge, and Zhe Chen. 2016. “NOVA Regulates Dcc Alternative Splicing during Neuronal Migration and Axon Guidance in the Spinal Cord.” *eLife* 5 (May). <https://doi.org/10.7554/eLife.14264>.
- Le Romancer, Muriel, Coralie Poulard, Pascale Cohen, Stéphanie Sentis, Jack-Michel Renoir, and Laura Corbo. 2011. “Cracking the Estrogen Receptor’s Posttranslational Code in Breast Tumors.” *Endocrine Reviews* 32 (5): 597–622.
- Le Tonquèze, Olivier, Bernhard Gschloessl, Vincent Legagneux, Luc Paillard, and Yann Audic. 2016. “Identification of CELF1 RNA Targets by CLIP-Seq in Human HeLa Cells.” *Genomics Data* 8 (June): 97–103.
- Levy, Shawn E., and Richard M. Myers. 2016. “Advancements in Next-Generation Sequencing.” *Annual Review of Genomics and Human Genetics* 17 (August): 95–115.
- Lian, Haifeng, Aili Wang, Yuanyuan Shen, Qian Wang, Zhenru Zhou, Ranran Zhang, Kun Li, Chengxia Liu, and Hongtao Jia. 2020. “Identification of Novel Alternative Splicing Isoform Biomarkers and Their Association with Overall Survival in Colorectal Cancer.” *BMC Gastroenterology* 20 (1): 171.

- Liao, J., and N. Xie. 2019. "Long Noncoding RNA DSCAM-AS1 Functions as an Oncogene in Non-Small Cell Lung Cancer by Targeting BCL11A." *European Review for Medical and Pharmacological Sciences* 23 (3): 1087–92.
- Li, Bo, and Colin N. Dewey. 2011. "RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome." *BMC Bioinformatics*. <https://doi.org/10.1186/1471-2105-12-323>.
- Li, Bo, Hai Sun, and Jiayu Zhang. 2020. "LncRNA DSCAM-AS1 Promotes Colorectal Cancer Progression by Acting as a Molecular Sponge of miR-384 to Modulate AKT3 Expression." *Aging* 12 (10): 9781–92.
- Li, Chunxia, Mitsuo Kato, Lily Shiue, John E. Shively, Manuel Ares, and Ren-Jang Lin. 2006. "Cell Type and Culture Condition-Dependent Alternative Splicing in Human Breast Cancer Cells Revealed by Splicing-Sensitive Microarrays." *Cancer Research*. <https://doi.org/10.1158/0008-5472.can-05-2593>.
- Li, Heng, and Richard Durbin. 2009. "Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform." *Bioinformatics* 25 (14): 1754–60.
- Li, Ji, Peter S. Choi, Christine L. Chaffer, Katherine Labella, Justin H. Hwang, Andrew O. Giacomelli, Jong Wook Kim, et al. 2018. "An Alternative Splicing Switch in FLNB Promotes the Mesenchymal Cell State in Human Breast Cancer." *eLife*. <https://doi.org/10.7554/elife.37184>.
- Linder, Patrick, and Eckhard Jankowsky. 2011. "From Unwinding to Clamping - the DEAD Box RNA Helicase Family." *Nature Reviews. Molecular Cell Biology* 12 (8): 505–16.
- Lin, Kuan-Ting, and Adrian R. Krainer. 2019. "PSI-Sigma: A Comprehensive Splicing-Detection Method for Short-Read and Long-Read RNA-Seq Analysis." *Bioinformatics* 35 (23): 5048–54.
- Lin, Shengrong, and Xiang-Dong Fu. 2007. "SR Proteins and Related Factors in Alternative Splicing." *Advances in Experimental Medicine and Biology*. [https://doi.org/10.1007/978-0-387-77374-2\\_7](https://doi.org/10.1007/978-0-387-77374-2_7).
- Li, Shengli, Zhixiang Hu, Yingjun Zhao, Shenglin Huang, and Xianghuo He. 2019. "Transcriptome-Wide Analysis Reveals the Landscape of Aberrant Alternative Splicing Events in Liver Cancer." *Hepatology* 69 (1): 359–75.
- Little, John T., and Melissa S. Jurica. 2008. "Splicing Factor SPF30 Bridges an Interaction between the Pre-spliceosome Protein U2AF35 and Tri-Small Nuclear Ribonucleoprotein Protein hPrp3." *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.m707984200>.
- Liu, Gangiang, John Mattick, and Ryan J. Taft. 2013. "A Meta-Analysis of the Genomic and Transcriptomic Composition of Complex Life." *Cell Cycle*. <https://doi.org/10.4161/cc.25134>.
- Liu, Junyu, Chuanhai Zhang, Boyang Zhang, Yao Sheng, Wentao Xu, Yunbo Luo, Xiaoyun He, and Kunlun Huang. 2020. "Comprehensive Analysis of the Characteristics and Differences in Adult and Newborn Brown Adipose Tissue (BAT): Newborn BAT Is a More Active/Dynamic BAT." *Cells* 9 (1). <https://doi.org/10.3390/cells9010201>.
- Liu, S. 2006. "The Network of Protein-Protein Interactions within the Human U4/U6.U5 Tri-snRNP." *RNA*. <https://doi.org/10.1261/ma.55406>.
- Liu, Serena, and Cole Trapnell. 2016. "Single-Cell Transcriptome Sequencing: Recent Advances and Remaining Challenges." *F1000Research*. <https://doi.org/10.12688/f1000research.7223.1>.

- Liu, Wei, Li Li, Hua Ye, Huan Tao, and Huaqin He. 2018. "Role of COL6A3 in Colorectal Cancer." *Oncology Reports* 39 (6): 2527–36.
- Liu, Xin, Ling Wang, Kehao Zhao, Paul R. Thompson, Yousang Hwang, Ronen Marmorstein, and Philip A. Cole. 2008. "The Structural Basis of Protein Acetylation by the p300/CBP Transcriptional Coactivator." *Nature*. <https://doi.org/10.1038/nature06546>.
- Liu, Yansheng, Mar González-Porta, Sergio Santos, Alvis Brazma, John C. Marioni, Ruedi Aebersold, Ashok R. Venkitaraman, and Vihandha O. Wickramasinghe. 2017. "Impact of Alternative Splicing on the Human Proteome." *Cell Reports* 20 (5): 1229–41.
- Liu, Yunze, Xin Liu, Changwei Lin, Xianhong Jia, Hongmei Zhu, Jun Song, and Yi Zhang. 2021. "Noncoding RNAs Regulate Alternative Splicing in Cancer." *Journal of Experimental & Clinical Cancer Research: CR* 40 (1): 11.
- Liu, Zheng, Qianjun Chen, and Swei Sunny Hann. 2019. "The Functions and Oncogenic Roles of CCAT1 in Human Cancer." *Biomedicine & Pharmacotherapy = Biomedecine & Pharmacotherapie* 115 (July): 108943.
- Li, Yang I., Bryce van de Geijn, Anil Raj, David A. Knowles, Allegra A. Petti, David Golan, Yoav Gilad, and Jonathan K. Pritchard. 2016. "RNA Splicing Is a Primary Link between Genetic Variation and Disease." *Science* 352 (6285): 600–604.
- Li, Yan, Ji-Ping Wang, Richard J. Santen, Tae-Hyun Kim, Hoyong Park, Ping Fan, and Wei Yue. 2010. "Estrogen Stimulation of Cell Migration Involves Multiple Signaling Pathway Interactions." *Endocrinology* 151 (11): 5146–56.
- Louadi, Zakaria, Kevin Yuan, Alexander Gress, Olga Tsoy, Olga V. Kalinina, Jan Baumbach, Tim Kacprowski, and Markus List. 2021. "DIGGER: Exploring the Functional Role of Alternative Splicing in Protein Interactions." *Nucleic Acids Research* 49 (D1): D309–18.
- Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15 (12): 550.
- Lubelsky, Yoav, and Igor Ulitsky. 2018. "Sequences Enriched in Alu Repeats Drive Nuclear Localization of Long RNAs in Human Cells." *Nature*. <https://doi.org/10.1038/nature25757>.
- Luo, M-J, and R. Reed. 1999. "Splicing Is Required for Rapid and Efficient mRNA Export in Metazoans." *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.96.26.14937>.
- Luo, Weibo, and Gregg L. Semenza. 2012. "Emerging Roles of PKM2 in Cell Metabolism and Cancer Progression." *Trends in Endocrinology and Metabolism: TEM* 23 (11): 560–66.
- Lu, Zhi-Xiang, Qin Huang, Juwon Park, Shihao Shen, Lan Lin, Collin J. Tokheim, Michael D. Henry, and Yi Xing. 2015. "Transcriptome-Wide Landscape of Pre-mRNA Alternative Splicing Associated with Metastatic Colonization." *Molecular Cancer Research*. <https://doi.org/10.1158/1541-7786.mcr-14-0366>.
- Lydon, J. P., F. J. DeMayo, C. R. Funk, S. K. Mani, A. R. Hughes, C. A. Montgomery Jr, G. Shyamala, O. M. Conneely, and B. W. O'Malley. 1995. "Mice Lacking Progesterone Receptor Exhibit Pleiotropic Reproductive Abnormalities." *Genes & Development* 9 (18): 2266–78.

- Maguire, Sarah L., Andri Leonidou, Patty Wai, Caterina Marchiò, Charlotte Ky Ng, Anna Sapino, Anne-Vincent Salomon, Jorge S. Reis-Filho, Britta Weigelt, and Rachael C. Natrajan. 2015. "SF3B1 Mutations Constitute a Novel Therapeutic Target in Breast Cancer." *The Journal of Pathology* 235 (4): 571–80.
- Makki, Jaafar. 2015. "Diversity of Breast Carcinoma: Histological Subtypes and Clinical Relevance." *Clinical Medicine Insights. Pathology* 8 (December): 23–31.
- Mallepell, Sonia, Andrée Krust, Pierre Chambon, and Cathrin Brisken. 2006. "Paracrine Signaling through the Epithelial Estrogen Receptor Alpha Is Required for Proliferation and Morphogenesis in the Mammary Gland." *Proceedings of the National Academy of Sciences of the United States of America* 103 (7): 2196–2201.
- Mao, Shuangshuang, Yuan Li, Zhiliang Lu, Yun Che, Jianbing Huang, Yuanyuan Lei, Yalong Wang, et al. 2019. "PHD Finger Protein 5A Promoted Lung Adenocarcinoma Progression via Alternative Splicing." *Cancer Medicine* 8 (5): 2429–41.
- Mardis, Elaine R. 2011. "A Decade's Perspective on DNA Sequencing Technology." *Nature*. <https://doi.org/10.1038/nature09796>.
- Marín-Béjar, Oskar, Aina M. Mas, Jovanna González, Dannys Martínez, Alejandro Athie, Xabier Morales, Mikel Galduroz, et al. 2017. "The Human lncRNA LINC-PINT Inhibits Tumor Cell Invasion through a Highly Conserved Sequence Element." *Genome Biology* 18 (1): 202.
- Masuda, Akio, Henriette Skovgaard Andersen, Thomas Koed Doktor, Takaaki Okamoto, Mikako Ito, Brage Storstein Andresen, and Kinji Ohno. 2012. "CUGBP1 and MBNL1 Preferentially Bind to 3' UTRs and Facilitate mRNA Decay." *Scientific Reports*. <https://doi.org/10.1038/srep00209>.
- Masuihi, Yoshikazu, Yoshihiro Mezaki, Matomo Sakari, Ken-Ichi Takeyama, Tasuku Yoshida, Kunio Inoue, Junn Yanagisawa, Shigemasa Hanazawa, Bert W. O'malley, and Shigeaki Kato. 2005. "Splicing Potentiation by Growth Factor Signals via Estrogen Receptor Phosphorylation." *Proceedings of the National Academy of Sciences of the United States of America* 102 (23): 8126–31.
- Matsumoto, K., K. M. Wassarman, and A. P. Wolffe. 1998. "Nuclear History of a Pre-mRNA Determines the Translational Activity of Cytoplasmic mRNA." *The EMBO Journal* 17 (7): 2107–21.
- McFall, Thomas, Brooke McKnight, Rayna Rosati, Seongho Kim, Yanfang Huang, Nerissa Viola-Villegas, and Manohar Ratnam. 2018. "Progesterone Receptor A Promotes Invasiveness and Metastasis of Luminal Breast Cancer by Suppressing Regulation of Critical microRNAs by Estrogen." *The Journal of Biological Chemistry* 293 (4): 1163–77.
- McGlinchy, Nicholas J., Lit-Yeen Tan, Nicodeme Paul, Mihaela Zavolan, Kathryn S. Lilley, and Christopher W. J. Smith. 2010. "Expression Proteomics of UPF1 Knockdown in HeLa Cells Reveals Autoregulation of hnRNP A2/B1 Mediated by Alternative Splicing Resulting in Nonsense-Mediated mRNA Decay." *BMC Genomics* 11 (October): 565.
- Mehmood, Arfa, Asta Laiho, Mikko S. Venäläinen, Aidan J. McGlinchey, Ning Wang, and Laura L. Elo. 2020. "Systematic Evaluation of Differential Splicing Tools for RNA-Seq Studies." *Briefings in Bioinformatics* 21 (6): 2052–65.

- Mei, Shenglin, Qian Qin, Qiu Wu, Hanfei Sun, Rongbin Zheng, Chongzhi Zang, Muyuan Zhu, et al. 2017. "Cistrome Data Browser: A Data Portal for ChIP-Seq and Chromatin Accessibility Data in Human and Mouse." *Nucleic Acids Research* 45 (D1): D658–62.
- Melé, Marta, Kaia Mattioli, William Mallard, David M. Shechner, Chiara Gerhardinger, and John L. Rinn. 2017. "Chromatin Environment, Transcriptional Regulation, and Splicing Distinguish lincRNAs and mRNAs." *Genome Research*. <https://doi.org/10.1101/gr.214205.116>.
- Melzer, Catharina, Ralf Hass, Hendrik Lehnert, and Hendrik Ungefroren. 2019. "RAC1B: A Rho GTPase with Versatile Functions in Malignant Transformation and Tumor Progression." *Cells* 8 (1). <https://doi.org/10.3390/cells8010021>.
- Melzer, Catharina, Ralf Hass, Juliane von der Ohe, Hendrik Lehnert, and Hendrik Ungefroren. 2017. "The Role of TGF- $\beta$  and Its Crosstalk with RAC1/RAC1b Signaling in Breast and Pancreas Carcinoma." *Cell Communication and Signaling: CCS* 15 (1): 19.
- Mercatante, Danielle R., Carl D. Bortner, John A. Cidlowski, and Ryszard Kole. 2001. "Modification of Alternative Splicing of Bcl-X Pre-mRNA in Prostate and Breast Cancer Cells." *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.m009256200>.
- Merino, Gabriela Alejandra, and Elmer Andrés Fernández. 2020. "Differential Splicing Analysis Based on Isoforms Expression with NBSplice." *Journal of Biomedical Informatics* 103 (March): 103378.
- Merkin, Jason, Caitlin Russell, Ping Chen, and Christopher B. Burge. 2012. "Evolutionary Dynamics of Gene and Isoform Regulation in Mammalian Tissues." *Science* 338 (6114): 1593–99.
- Métivier, Raphaël, Graziella Penot, Michael R. Hübner, George Reid, Heike Brand, Martin Kos, and Frank Gannon. 2003. "Estrogen Receptor-Alpha Directs Ordered, Cyclical, and Combinatorial Recruitment of Cofactors on a Natural Target Promoter." *Cell* 115 (6): 751–63.
- Miano, Valentina, Giulio Ferrero, Stefania Reineri, Livia Caizzi, Laura Annaratone, Laura Ricci, Santina Cutrupi, Isabella Castellano, Francesca Cordero, and Michele De Bortoli. 2016. "Luminal Long Non-Coding RNAs Regulated by Estrogen Receptor Alpha in a Ligand-Independent Manner Show Functional Roles in Breast Cancer." *Oncotarget* 7 (3): 3201–16.
- Miano, Valentina, Giulio Ferrero, Valentina Rosti, Eleonora Manitta, Jamal Elhasnaoui, Giulia Basile, and Michele De Bortoli. 2018. "Luminal lncRNAs Regulation by ER $\alpha$ -Controlled Enhancers in a Ligand-Independent Manner in Breast Cancer Cells." *International Journal of Molecular Sciences* 19 (2). <https://doi.org/10.3390/ijms19020593>.
- Mizutani, A., D. Koinuma, H. Seimiya, and K. Miyazono. 2016. "The Arkadia-ESRP2 Axis Suppresses Tumor Progression: Analyses in Clear-Cell Renal Cell Carcinoma." *Oncogene* 35 (27): 3514–23.
- Modrek, Barmak, and Christopher Lee. 2002. "A Genomic View of Alternative Splicing." *Nature Genetics*. <https://doi.org/10.1038/ng0102-13>.
- Mohammed, Hisham, I. Alasdair Russell, Rory Stark, Oscar M. Rueda, Theresa E. Hickey, Gerard A. Tarulli, Aurelien A. Serandour, et al. 2015. "Progesterone Receptor Modulates ER $\alpha$  Action in Breast Cancer." *Nature*. <https://doi.org/10.1038/nature14583>.



- Molina, Luis, Carlos D. Figueroa, Kanti D. Bhoola, and Pamela Ehrenfeld. 2017. "GPER-1/GPR30 a Novel Estrogen Receptor Sited in the Cell Membrane: Therapeutic Coupling to Breast Cancer." *Expert Opinion on Therapeutic Targets* 21 (8): 755–66.
- Mondal, Tanmoy, Santhilal Subhash, Roshan Vaid, Stefan Enroth, Sireesha Uday, Björn Reinius, Sanhita Mitra, et al. 2015. "MEG3 Long Noncoding RNA Regulates the TGF- $\beta$  Pathway Genes through Formation of RNA-DNA Triplex Structures." *Nature Communications* 6 (July): 7743.
- Montalto, Francesca Ida, Francesca Giordano, Chiara Chiodo, Stefania Marsico, Loredana Mauro, Diego Sisci, Saveria Aquila, et al. 2019. "Progesterone Receptor B Signaling Reduces Breast Cancer Cell Aggressiveness: Role of Cyclin-D1/Cdk4 Mediating Paxillin Phosphorylation." *Cancers* 11 (8). <https://doi.org/10.3390/cancers11081201>.
- Morris, Kevin V., and John S. Mattick. 2014. "The Rise of Regulatory RNA." *Nature Reviews Genetics*. <https://doi.org/10.1038/nrg3722>.
- Mosca, Roberto, Arnaud Céol, Amelie Stein, Roger Olivella, and Patrick Aloy. 2014. "3did: A Catalog of Domain-Based Interactions of Known Three-Dimensional Structure." *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkt887>.
- Moss, Alan C., Peter P. Doran, and Padraic Macmathuna. 2007. "In Silico Promoter Analysis Can Predict Genes of Functional Relevance in Cell Proliferation: Validation in a Colon Cancer Model." *Translational Oncogenomics* 2 (February): 1–16.
- Motta-Mena, Laura B., Florian Heyd, and Kristen W. Lynch. 2010. "Context-Dependent Regulatory Mechanism of the Splicing Factor hnRNP L." *Molecular Cell* 37 (2): 223–34.
- Moudgil, V. K. 2013. *Steroid Hormone Receptors: Basic and Clinical Aspects*. Springer Science & Business Media.
- Mumbach, Maxwell R., Jeffrey M. Granja, Ryan A. Flynn, Caitlin M. Roake, Ansuman T. Satpathy, Adam J. Rubin, Yanyan Qi, et al. 2019. "HiChIRP Reveals RNA-Associated Chromosome Conformation." *Nature Methods* 16 (6): 489–92.
- Munkley, Jennifer, Ling Li, S. R. Gokul Krishnan, Gerald Hysenaj, Emma Scott, Caroline Dalgliesh, Htoo Zarni Oo, et al. 2019. "Androgen-Regulated Transcription of Drives Alternative Splicing Patterns in Prostate Cancer." *eLife* 8 (September). <https://doi.org/10.7554/eLife.47678>.
- Murphy, L. C., S. V. Seekallu, and P. H. Watson. 2011. "Clinical Significance of Estrogen Receptor Phosphorylation." *Endocrine Related Cancer*. <https://doi.org/10.1677/erc-10-0070>.
- Nagaoka, K., K. Fujii, H. Zhang, K. Usuda, G. Watanabe, M. Ivshina, and J. D. Richter. 2016. "CPEB1 Mediates Epithelial-to-Mesenchyme Transition and Breast Cancer Metastasis." *Oncogene* 35 (22): 2893–2901.
- Nassa, Giovanni, Roberta Tarallo, Pietro H. Guzzi, Lorenzo Ferraro, Francesca Cirillo, Maria Ravo, Ernesto Nola, et al. 2011. "Comparative Analysis of Nuclear Estrogen Receptor Alpha and Beta Interactomes in Breast Cancer Cells." *Molecular bioSystems* 7 (3): 667–76.
- Neumann, Alexander. 2019. *Leveraging RNA-Sequencing Data to Obtain Insights about MRNA Splicing: From Daily Rhythms to Secretory Adaptations and Cryptic Splice Sites*.

- Nielsen, Finn Cilius, Thomas van Overeem Hansen, and Claus Storgaard Sørensen. 2016. “Hereditary Breast and Ovarian Cancer: New Genes in Confined Pathways.” *Nature Reviews Cancer*. <https://doi.org/10.1038/nrc.2016.72>.
- Niknafs, Yashar S., Sumin Han, Teng Ma, Corey Speers, Chao Zhang, Kari Wilder-Romans, Matthew K. Iyer, et al. 2016. “The lncRNA Landscape of Breast Cancer Reveals a Role for DSCAM-AS1 in Breast Cancer Progression.” *Nature Communications* 7 (September): 12791.
- Niknafs, Yashar S., Sumin Han, Teng Ma, Chao Zhang, Matthew K. Iyer, Rohit Malik, Anton Poliakov, et al. 2016. “Abstract B19: Interrogation of the Landscape of Long Noncoding RNAs in Breast Cancer to Identify an ER-Regulated Predictor of Tamoxifen Resistance.” *Mechanisms of Noncoding RNAs in Tumorigenesis*. <https://doi.org/10.1158/1538-7445.nonma15-b19>.
- Ni, Ting, Yanqin Yang, Dina Hafez, Wenjing Yang, Kurtis Kiesewetter, Yoshi Wakabayashi, Uwe Ohler, Weiqun Peng, and Jun Zhu. 2013. “Distinct Polyadenylation Landscapes of Diverse Human Tissues Revealed by a Modified PA-Seq Strategy.” *BMC Genomics* 14 (September): 615.
- Noh, Ji Heon, Kyoung Mi Kim, Kotb Abdelmohsen, Je-Hyun Yoon, Amaresh C. Panda, Rachel Munk, Jiyoung Kim, et al. 2016. “HuR and GRSF1 Modulate the Nuclear Export and Mitochondrial Localization of the lncRNA RMRP.” *Genes & Development* 30 (10): 1224–39.
- Nojima, Takayuki, Michael Tellier, Jonathan Foxwell, Claudia Ribeiro de Almeida, Sue Mei Tan-Wong, Somdutta Dhir, Gwendal Dujardin, Ashish Dhir, Shona Murphy, and Nick J. Proudfoot. 2018. “Deregulated Expression of Mammalian lncRNA through Loss of SPT6 Induces R-Loop Formation, Replication Stress, and Cellular Senescence.” *Molecular Cell* 72 (6): 970–84.e7.
- Norris, Adam D., and John A. Calarco. 2012. “Emerging Roles of Alternative Pre-mRNA Splicing Regulation in Neuronal Development and Function.” *Frontiers in Neuroscience* 6 (August): 122.
- Oh, Jagyeong, Davide Pradella, Changwei Shao, Hairi Li, Namjeong Choi, Jiyeon Ha, Sonia Ruggiero, et al. 2021. “Widespread Alternative Splicing Changes in Metastatic Breast Cancer Cells.” *Cells* 10 (4). <https://doi.org/10.3390/cells10040858>.
- Ohno, S. 1972. “So Much ‘Junk’ DNA in Our Genome.” *Brookhaven Symposia in Biology* 23: 366–70.
- O’Leary, Valerie Bríd, Saak Victor Ovsepiyan, Laura Garcia Carrascosa, Fabian Andreas Buske, Vanja Radulovic, Maximilian Niyazi, Simone Moertl, Matt Trau, Michael John Atkinson, and Nataša Anastasov. 2015. “PARTICLE, a Triplex-Forming Long ncRNA, Regulates Locus-Specific Methylation in Response to Low-Dose Irradiation.” *Cell Reports* 11 (3): 474–85.
- Onyango, David O., Gabriella Lee, and Jeremy M. Stark. 2017. “PRPF8 Is Important for BRCA1-Mediated Homologous Recombination.” *Oncotarget* 8 (55): 93319–37.
- Ota, Toshio, Yutaka Suzuki, Tetsuo Nishikawa, Tetsuji Otsuki, Tomoyasu Sugiyama, Ryotaro Irie, Ai Wakamatsu, et al. 2004. “Complete Sequencing and Characterization of 21,243 Full-Length Human cDNAs.” *Nature Genetics* 36 (1): 40–45.

- Oughtred, Rose, Chris Stark, Bobby-Joe Breitkreutz, Jennifer Rust, Lorrie Boucher, Christie Chang, Nadine Kolas, et al. 2019. "The BioGRID Interaction Database: 2019 Update." *Nucleic Acids Research* 47 (D1): D529–41.
- Pandey, Poomy, Satyanarayana Rachagani, Srustidhar Das, Parthasarathy Seshacharyulu, Yuri Sheinin, Naava Naslavsky, Zenggang Pan, et al. 2015. "Amyloid Precursor-like Protein 2 (APLP2) Affects the Actin Cytoskeleton and Increases Pancreatic Cancer Growth and Metastasis." *Oncotarget* 6 (4): 2064–75.
- Pardi, Norbert, Michael J. Hogan, Frederick W. Porter, and Drew Weissman. 2018. "mRNA Vaccines - a New Era in Vaccinology." *Nature Reviews. Drug Discovery* 17 (4): 261–79.
- Park, Eddie, Zhicheng Pan, Zijun Zhang, Lan Lin, and Yi Xing. 2018. "The Expanding Landscape of Alternative Splicing Variation in Human Populations." *American Journal of Human Genetics* 102 (1): 11–26.
- Parker, Joel S., Michael Mullins, Maggie C. U. Cheang, Samuel Leung, David Voduc, Tammi Vickery, Sherri Davies, et al. 2009. "Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes." *Journal of Clinical Oncology*. <https://doi.org/10.1200/jco.2008.18.1370>.
- Park, Sunghye, Mattia Brugiolo, Martin Akerman, Shipra Das, Laura Urbanski, Adam Geier, Anil K. Kesarwani, et al. 2019. "Differential Functions of Splicing Factors in Mammary Transformation and Breast Cancer Metastasis." *Cell Reports* 29 (9): 2672–88.e7.
- Parl, F. F., B. P. Schmidt, W. D. Dupont, and R. K. Wagner. 1984. "Prognostic Significance of Estrogen Receptor Status in Breast Cancer in Relation to Tumor Stage, Axillary Node Metastasis, and Histopathologic Grading." *Cancer* 54 (10): 2237–42.
- Patro, Rob, Geet Duggal, Michael I. Love, Rafael A. Irizarry, and Carl Kingsford. 2017. "Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression." *Nature Methods* 14 (4): 417–19.
- Patro, Rob, Stephen M. Mount, and Carl Kingsford. 2014. "Sailfish Enables Alignment-Free Isoform Quantification from RNA-Seq Reads Using Lightweight Algorithms." *Nature Biotechnology* 32 (5): 462–64.
- Pavet, V., M. M. Portal, J. C. Moulin, R. Herbrecht, and H. Gronemeyer. 2011. "Towards Novel Paradigms for Cancer Therapy." *Oncogene* 30 (1): 1–20.
- Peart, Natoya J., Jae Yeon Hwang, Mathieu Quesnel-Vallières, Matthew J. Sears, Yuequin Yang, Peter Stoilov, Yoseph Barash, Juwon Park, and Russ P. Carstens. n.d. "The Global Protein-RNA Interaction Map of Epithelial Splicing Regulatory Protein 1 Defines a Post-Transcriptional Program That Is Essential for Epithelial Cell Function." <https://doi.org/10.1101/2021.05.18.444719>.
- Pedroza, Diego A., Ramadevi Subramani, and Rajkumar Lakshmanaswamy. 2020. "Classical and Non-Classical Progesterone Signaling in Breast Cancers." *Cancers* 12 (9). <https://doi.org/10.3390/cancers12092440>.
- Pennisi, Elizabeth. 2012. "Genomics. ENCODE Project Writes Eulogy for Junk DNA." *Science* 337 (6099): 1159, 1161.
- Penot, Graziella, Christine Le Péron, Yohann Mérot, Eva Grimaud-Fanouillère, François Ferrière, Nouredine Boujrad, Olivier Kah, et al. 2005. "The Human Estrogen Receptor-Alpha Isoform hERalpha46 Antagonizes the Proliferative Influence of hERalpha66 in MCF7 Breast Cancer Cells." *Endocrinology* 146 (12): 5474–84.

- Perou, Charles M., Therese Sørlie, Michael B. Eisen, Matt van de Rijn, Stefanie S. Jeffrey, Christian A. Rees, Jonathan R. Pollack, et al. 2000. "Molecular Portraits of Human Breast Tumours." *Nature*. <https://doi.org/10.1038/35021093>.
- Peters, Haley L., Ying Yan, and Joyce C. Solheim. 2013. "APLP2 Regulates the Expression of MHC Class I Molecules on Irradiated Ewing's Sarcoma Cells." *Oncology Immunology*. <https://doi.org/10.4161/onci.26293>.
- Phillips, John W., Yang Pan, Brandon L. Tsai, Zhijie Xie, Levon Demirdjian, Wen Xiao, Harry T. Yang, et al. 2020. "Pathway-Guided Analysis Identifies Myc-Dependent Alternative Pre-mRNA Splicing in Aggressive Prostate Cancers." *Proceedings of the National Academy of Sciences of the United States of America* 117 (10): 5269–79.
- Piunti, Andrea, and Ali Shilatifard. 2016. "Epigenetic Balance of Gene Expression by Polycomb and COMPASS Families." *Science* 352 (6290): aad9780.
- Podszycalow-Bartnicka, Paulina, Michalina Kosiorek, Katarzyna Piwocka, Ewa Sikora, Krzysztof Zablocki, and Slawomir Pikula. 2010. "Role of Annexin A6 Isoforms in Catecholamine Secretion by PC12 Cells: Distinct Influence on Calcium Response." *Journal of Cellular Biochemistry*. <https://doi.org/10.1002/jcb.22685>.
- Power, R., S. Mani, J. Codina, O. Conneely, and B. O'Malley. 1991. "Dopaminergic and Ligand-Independent Activation of Steroid Hormone Receptors." *Science*. <https://doi.org/10.1126/science.1749936>.
- Prokunina-Olsson, Ludmila, Cullan Welch, Ola Hansson, Neeta Adhikari, Laura J. Scott, Nicolle Usher, Maurine Tong, et al. 2009. "Tissue-Specific Alternative Splicing of TCF7L2." *Human Molecular Genetics* 18 (20): 3795–3804.
- Proudfoot, Nick J., Andre Furger, and Michael J. Dye. 2002. "Integrating mRNA Processing with Transcription." *Cell*. [https://doi.org/10.1016/s0092-8674\(02\)00617-7](https://doi.org/10.1016/s0092-8674(02)00617-7).
- Qi, Houbao, Shuqing Liu, Chunmei Guo, Jiasheng Wang, Frederick T. Greenaway, and Ming-Zhong Sun. 2015. "Role of Annexin A6 in Cancer." *Oncology Letters* 10 (4): 1947–52.
- Qin, Hai, Haiwei Ni, Yichen Liu, Yaqin Yuan, Tao Xi, Xiaoman Li, and Lufeng Zheng. 2020. "RNA-Binding Proteins in Tumor Progression." *Journal of Hematology & Oncology* 13 (1): 90.
- Quattrone, Alessandro, and Erik Dassi. 2019. "The Architecture of the Human RNA-Binding Protein Regulatory Network." *iScience* 21 (November): 706–19.
- Quinn, Jeffrey J., Qiangfeng C. Zhang, Plamen Georgiev, Ibrahim A. Ilik, Asifa Akhtar, and Howard Y. Chang. 2016. "Rapid Evolutionary Turnover Underlies Conserved lncRNA–genome Interactions." *Genes & Development*. <https://doi.org/10.1101/gad.272187.115>.
- Raffetseder, Ute, Björn Frye, Thomas Rauen, Karsten Jürchott, Hans-Dieter Royer, Petra Lynen Jansen, and Peter R. Mertens. 2003. "Splicing Factor SRp30c Interaction with Y-Box Protein-1 Confers Nuclear YB-1 Shuttling and Alternative Splice Site Selection." *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.m212518200>.
- Raghavachari, Balaji, Asba Tasneem, Teresa M. Przytycka, and Raja Jothi. 2008. "DOMINE: A Database of Protein Domain Interactions." *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkm761>.

- Rahman, Mohammad Alinoor, Adrian R. Krainer, and Omar Abdel-Wahab. 2020. "SnapShot: Splicing Alterations in Cancer." *Cell* 180 (1): 208–208.e1.
- Ransohoff, Julia D., Yuning Wei, and Paul A. Khavari. 2018. "The Functions and Unique Features of Long Intergenic Non-Coding RNA." *Nature Reviews. Molecular Cell Biology* 19 (3): 143–57.
- Ray, Debashish, Kevin C. H. Ha, Kate Nie, Hong Zheng, Timothy R. Hughes, and Quaid D. Morris. 2017. "RNAcompete Methodology and Application to Determine Sequence Preferences of Unconventional RNA-Binding Proteins." *Methods* 118-119 (April): 3–15.
- Ray, Debashish, Hilal Kazan, Esther T. Chan, Lourdes Peña Castillo, Sidharth Chaudhry, Shaheynoor Talukder, Benjamin J. Blencowe, Quaid Morris, and Timothy R. Hughes. 2009. "Rapid and Systematic Analysis of the RNA Recognition Specificities of RNA-Binding Proteins." *Nature Biotechnology* 27 (7): 667–70.
- Ray, Debashish, Hilal Kazan, Kate B. Cook, Matthew T. Weirauch, Hamed S. Najafabadi, Xiao Li, Serge Gueroussov, et al. 2013. "A Compendium of RNA-Binding Motifs for Decoding Gene Regulation." *Nature* 499 (7457): 172–77.
- Reinke, Lauren M., Yilin Xu, and Chonghui Cheng. 2012. "Snail Represses the Splicing Regulator Epithelial Splicing Regulatory Protein 1 to Promote Epithelial-Mesenchymal Transition." *The Journal of Biological Chemistry* 287 (43): 36435–42.
- Reis-Filho, Jorge S., and Lajos Pusztai. 2011. "Gene Expression Profiling in Breast Cancer: Classification, Prognostication, and Prediction." *The Lancet* 378 (9805): 1812–23.
- Ritchie, Marylyn D., Emily R. Holzinger, Ruowang Li, Sarah A. Pendergrass, and Dokyoon Kim. 2015. "Methods of Integrating Data to Uncover Genotype-Phenotype Interactions." *Nature Reviews. Genetics* 16 (2): 85–97.
- Rivera, Chloe M., and Bing Ren. 2013. "Mapping Human Epigenomes." *Cell*. <https://doi.org/10.1016/j.cell.2013.09.011>.
- Rodriguez, Jose Manuel, Juan Rodriguez-Rivas, Tomás Di Domenico, Jesús Vázquez, Alfonso Valencia, and Michael L. Tress. 2018. "APPRIS 2017: Principal Isoforms for Multiple Gene Sets." *Nucleic Acids Research* 46 (D1): D213–17.
- Romero-Barrios, Natali, Maria Florencia Legascue, Moussa Benhamed, Federico Ariel, and Martin Crespi. 2018. "Splicing Regulation by Long Noncoding RNAs." *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gky095>.
- Romfo, C. M., C. J. Alvarez, W. J. van Heeckeren, C. J. Webb, and J. A. Wise. 2000. "Evidence for Splice Site Pairing via Intron Definition in *Schizosaccharomyces Pombe*." *Molecular and Cellular Biology* 20 (21): 7955–70.
- Rose, Alan B. 2018. "Introns as Gene Regulators: A Brick on the Accelerator." *Frontiers in Genetics* 9: 672.
- Rosenberg, Alexander B., Rupali P. Patwardhan, Jay Shendure, and Georg Seelig. 2015. "Learning the Sequence Determinants of Alternative Splicing from Millions of Random Sequences." *Cell* 163 (3): 698–711.

- Rosen, J. M. 2012. "On Hormone Action in the Mammary Gland." *Cold Spring Harbor Perspectives in Biology*. <https://doi.org/10.1101/cshperspect.a013086>.
- Rossbach, Oliver, Lee-Hsueh Hung, Ekaterina Khrameeva, Silke Schreiner, Julian König, Tomaž Curk, Blaž Zupan, Jernej Ule, Mikhail S. Gelfand, and Albrecht Bindereif. 2014. "Crosslinking-Immunoprecipitation (iCLIP) Analysis Reveals Global Regulatory Roles of hnRNP L." *RNA Biology*. <https://doi.org/10.4161/ma.27991>.
- Ross-Innes, Caryn S., Rory Stark, Andrew E. Teschendorff, Kelly A. Holmes, H. Raza Ali, Mark J. Dunning, Gordon D. Brown, et al. 2012. "Differential Oestrogen Receptor Binding Is Associated with Clinical Outcome in Breast Cancer." *Nature* 481 (7381): 389–93.
- Rouzier, Roman, Charles M. Perou, W. Fraser Symmans, Nuhad Ibrahim, Massimo Cristofanilli, Keith Anderson, Kenneth R. Hess, et al. 2005. "Breast Cancer Molecular Subtypes Respond Differently to Preoperative Chemotherapy." *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 11 (16): 5678–85.
- Ryan, Michael C., James Cleland, Ryanguk Kim, Wing Chung Wong, and John N. Weinstein. 2012. "SpliceSeq: A Resource for Analysis and Visualization of RNA-Seq Data on Alternative Splicing and Its Functional Impacts." *Bioinformatics* 28 (18): 2385–87.
- Ryan, Michael, Wing Chung Wong, Robert Brown, Rehan Akbani, Xiaoping Su, Bradley Broom, James Melott, and John Weinstein. 2016. "TCGASpliceSeq a Compendium of Alternative mRNA Splicing in Cancer." *Nucleic Acids Research* 44 (D1): D1018–22.
- Saha, Ashis, Yungil Kim, Ariel D. H. Gewirtz, Brian Jo, Chuan Gao, Ian C. McDowell, GTEx Consortium, Barbara E. Engelhardt, and Alexis Battle. 2017. "Co-Expression Networks Reveal the Tissue-Specific Regulation of Transcription and Splicing." *Genome Research* 27 (11): 1843–58.
- Salama, Salama A., Mahmoud A. Mohammad, Concepcion R. Diaz-Arrastia, Marwa W. Kamel, Gokhan S. Kilic, Bih T. Ndofor, Mohamed S. Abdel-Baki, and Shaleen K. Theiler. 2014. "Estradiol-17 $\beta$  Upregulates Pyruvate Kinase M2 Expression to Coactivate Estrogen Receptor- $\alpha$  and to Integrate Metabolic Reprogramming with the Mitogenic Response in Endometrial Cells." *The Journal of Clinical Endocrinology and Metabolism* 99 (10): 3790–99.
- Salmena, Leonardo, Laura Poliseno, Yvonne Tay, Lev Kats, and Pier Paolo Pandolfi. 2011. "A ceRNA Hypothesis: The Rosetta Stone of a Hidden RNA Language?" *Cell* 146 (3): 353–58.
- Sanawar, Rahul, Vipin Mohan Dan, Thankayyan R. Santhoshkumar, Rakesh Kumar, and M. Radhakrishna Pillai. 2019. "Estrogen Receptor- $\alpha$  Regulation of microRNA-590 Targets FAM171A1-a Modifier of Breast Cancer Invasiveness." *Oncogenesis* 8 (1): 5.
- Sandbrink, R., C. L. Masters, and K. Beyreuther. 1994. "Similar Alternative Splicing of a Non-Homologous Domain in Beta A4-Amyloid Protein Precursor-like Proteins." *Journal of Biological Chemistry*. [https://doi.org/10.1016/s0021-9258\(17\)36778-9](https://doi.org/10.1016/s0021-9258(17)36778-9).
- Sanford, J. R., J. Ellis, and J. F. Cáceres. 2005. "Multiple Roles of Arginine/serine-Rich Splicing Factors in RNA Processing." *Biochemical Society Transactions* 33 (Pt 3): 443–46.

- Sarropoulos, Ioannis, Ray Marin, Margarida Cardoso-Moreira, and Henrik Kaessmann. 2019. “Developmental Dynamics of lncRNAs across Mammalian Organs and Species.” *Nature* 571 (7766): 510–14.
- Schafer, Sebastian, Kui Miao, Craig C. Benson, Matthias Heinig, Stuart A. Cook, and Norbert Hubner. 2015. “Alternative Splicing Signatures in RNA-Seq Data: Percent Spliced in (PSI).” *Current Protocols in Human Genetics / Editorial Board, Jonathan L. Haines ... [et Al.]* 87 (October): 11.16.1–11.16.14.
- Scheidereit, C., P. Krauter, D. Von Der Ahe, S. Janich, O. Rabenau, A. C. B. Cato, G. Suske, H. M. Westphal, and M. Beato. 1986. “Mechanism of Gene Regulation by Steroid Hormones.” *Journal of Steroid Biochemistry*. [https://doi.org/10.1016/0022-4731\(86\)90026-9](https://doi.org/10.1016/0022-4731(86)90026-9).
- Schlackow, Margarita, Takayuki Nojima, Tomas Gomes, Ashish Dhir, Maria Carmo-Fonseca, and Nick J. Proudfoot. 2017. “Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs.” *Molecular Cell* 65 (1): 25–38.
- Schmucker, Dietmar, James C. Clemens, Huidy Shu, Carolyn A. Worby, Jian Xiao, Marco Muda, Jack E. Dixon, and S. Lawrence Zipursky. 2000. “Drosophila Dscam Is an Axon Guidance Receptor Exhibiting Extraordinary Molecular Diversity.” *Cell*. [https://doi.org/10.1016/s0092-8674\(00\)80878-8](https://doi.org/10.1016/s0092-8674(00)80878-8).
- Schnelzer, A., D. Prechtel, U. Knaus, K. Dehne, M. Gerhard, H. Graeff, N. Harbeck, M. Schmitt, and E. Lengyel. 2000. “Rac1 in Human Breast Cancer: Overexpression, Mutation Analysis, and Characterization of a New Isoform, Rac1b.” *Oncogene* 19 (26): 3013–20.
- Schwabe, J. W. R., and S. A. Teichmann. 2004. “Nuclear Receptors: The Evolution of Diversity.” *Science Signaling*. <https://doi.org/10.1126/stke.2172004pe4>.
- Schwanhäusser, Björn, Dorothea Busse, Na Li, Gunnar Dittmar, Johannes Schuchhardt, Jana Wolf, Wei Chen, and Matthias Selbach. 2011. “Global Quantification of Mammalian Gene Expression Control.” *Nature*. <https://doi.org/10.1038/nature10098>.
- Sebestyén, Endre, Babita Singh, Belén Miñana, Amadís Pagès, Francesca Mateo, Miguel Angel Pujana, Juan Valcárcel, and Eduardo Eyras. 2016. “Large-Scale Analysis of Genome and Transcriptome Alterations in Multiple Tumors Unveils Novel Cancer-Relevant Splicing Networks.” *Genome Research* 26 (6): 732–44.
- Sebestyén, Endre, Michał Zawisza, and Eduardo Eyras. 2015. “Detection of Recurrent Alternative Splicing Switches in Tumor Samples Reveals Novel Signatures of Cancer.” *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gku1392>.
- Seo, Ji-Young, Do-Yeon Kim, Seong-Hoon Kim, Hyo-Jin Kim, Hye Guk Ryu, Juhyun Lee, Kyung-Ha Lee, and Kyong-Tai Kim. 2017. “Heterogeneous Nuclear Ribonucleoprotein (hnRNP) L Promotes DNA Damage-Induced Cell Apoptosis by Enhancing the Translation of p53.” *Oncotarget* 8 (31): 51108–22.
- Shapiro, Irina M., Albert W. Cheng, Nicholas C. Flytzanis, Michele Balsamo, John S. Condeelis, Maja H. Oktay, Christopher B. Burge, and Frank B. Gertler. 2011. “An EMT-Driven Alternative Splicing Program Occurs in Human Breast Cancer and Modulates Cellular Phenotype.” *PLoS Genetics*. <https://doi.org/10.1371/journal.pgen.1002218>.
- Sheeler, Cameron Q., David W. Singleton, and Sobhaib A. Khan. 2003. “Mutation of Serines 104, 106, and 118 Inhibits Dimerization of the Human Estrogen Receptor in Yeast.” *Endocrine Research* 29 (2): 237–55.

- Shen, Shihao, Juw Won Park, Jian Huang, Kimberly A. Dittmar, Zhi-Xiang Lu, Qing Zhou, Russ P. Carstens, and Yi Xing. 2012. "MATS: A Bayesian Framework for Flexible Detection of Differential Alternative Splicing from RNA-Seq Data." *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkr1291>.
- Shen, Shihao, Juw Won Park, Zhi-Xiang Lu, Lan Lin, Michael D. Henry, Ying Nian Wu, Qing Zhou, and Yi Xing. 2014. "rMATS: Robust and Flexible Detection of Differential Alternative Splicing from Replicate RNA-Seq Data." *Proceedings of the National Academy of Sciences of the United States of America* 111 (51): E5593–5601.
- Shukla, Chinmay J., Alexandra L. McCorkindale, Chiara Gerhardinger, Keegan D. Korthauer, Moran N. Cabili, David M. Shechner, Rafael A. Irizarry, Philipp G. Maass, and John L. Rinn. 2018. "High-Throughput Identification of RNA Nuclear Enrichment Sequences." *The EMBO Journal* 37 (6). <https://doi.org/10.15252/emj.201798452>.
- Shults, Cody L., Caitlin B. Dingwall, Chun K. Kim, Elena Pinceti, Yathindar S. Rao, and Toni R. Pak. 2018. "17 $\beta$ -Estradiol Regulates the RNA-Binding Protein Nova1, Which Then Regulates the Alternative Splicing of Estrogen Receptor  $\beta$  in the Aging Female Rat Brain." *Neurobiology of Aging* 61 (January): 13–22.
- Shults, Cody L., Elena Pinceti, Yathindar S. Rao, and Toni R. Pak. 2015. "Aging and Loss of Circulating 17 $\beta$ -Estradiol Alters the Alternative Splicing of ER $\beta$  in the Female Rat Brain." *Endocrinology*. <https://doi.org/10.1210/en.2015-1514>.
- Sinha, Abhilasha, Sumit Agarwal, Deepak Parashar, Archana Verma, Shikha Saini, Nirmala Jagadish, Abdul S. Ansari, Nirmal K. Lohiya, and Anil Suri. 2013. "Down Regulation of SPAG9 Reduces Growth and Invasive Potential of Triple-Negative Breast Cancer Cells: Possible Implications in Targeted Therapy." *Journal of Experimental & Clinical Cancer Research: CR* 32 (September): 69.
- Slepicka, Priscila Ferreira, Amritha Varshini Hanasoge Somasundara, and Camila O. Dos Santos. 2020. "The Molecular Basis of Mammary Gland Development and Epithelial Differentiation." *Seminars in Cell & Developmental Biology*, October. <https://doi.org/10.1016/j.semcdb.2020.09.014>.
- Society for Experimental Biology (Great Britain). 1958. *The Biological Replication of Macromolecules*.
- Soneson, Charlotte, Michael I. Love, and Mark D. Robinson. 2015. "Differential Analyses for RNA-Seq: Transcript-Level Estimates Improve Gene-Level Inferences." *F1000Research* 4 (December): 1521.
- Sørli, Therese, Robert Tibshirani, Joel Parker, Trevor Hastie, J. S. Marron, Andrew Nobel, Shihong Deng, et al. 2003. "Repeated Observation of Breast Tumor Subtypes in Independent Gene Expression Data Sets." *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.0932692100>.
- Sperling, Joseph, Maia Azubel, and Ruth Sperling. 2008. "Structure and Function of the Pre-mRNA Splicing Machine." *Structure* 16 (11): 1605–15.
- Statello, Luisa, Chun-Jie Guo, Ling-Ling Chen, and Maite Huarte. 2021. "Gene Regulation by Long Non-Coding RNAs and Its Biological Functions." *Nature Reviews. Molecular Cell Biology* 22 (2): 96–118.
- Statello, Luisa, Marco Maugeri, Elena Garre, Muhammad Nawaz, Jessica Wahlgren, Alexandros Papadimitriou, Christina Lundqvist, et al. 2018. "Identification of RNA-Binding Proteins in Exosomes



- Capable of Interacting with Different Types of RNA: RBP-Facilitated Transport of RNAs into Exosomes.” *PloS One* 13 (4): e0195969.
- Stender, Joshua D., Jonna Frasar, Barry Komm, Ken C. N. Chang, W. Lee Kraus, and Benita S. Katzenellenbogen. 2007. “Estrogen-Regulated Gene Networks in Human Breast Cancer Cells: Involvement of E2F1 in the Regulation of Cell Proliferation.” *Molecular Endocrinology* 21 (9): 2112–23.
- Sterne-Weiler, Timothy, Robert J. Weatheritt, Andrew J. Best, Kevin C. H. Ha, and Benjamin J. Blencowe. 2018. “Efficient and Accurate Quantitative Profiling of Alternative Splicing Patterns of Any Complexity on a Laptop.” *Molecular Cell* 72 (1): 187–200.e6.
- Stricker, Thomas P., Christopher D. Brown, Chaitanya Bandlamudi, Megan Mc Nerney, Ralf Kittler, Vanessa Montoya, April Peterson, Robert Grossman, and Kevin P. White. 2017. “Robust Stratification of Breast Cancer Subtypes Using Differential Patterns of Transcript Isoform Expression.” *PLoS Genetics* 13 (3): e1006589.
- Stumpf, Craig R., Melissa V. Moreno, Adam B. Olshen, Barry S. Taylor, and Davide Ruggero. 2013. “The Translational Landscape of the Mammalian Cell Cycle.” *Molecular Cell*. <https://doi.org/10.1016/j.molcel.2013.09.018>.
- Sugnet, C. W., W. J. Kent, M. Ares Jr, and D. Haussler. 2004. “Transcriptome and Genome Conservation of Alternative Splicing Events in Humans and Mice.” *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 66–77.
- Sun, Guoliang, Hui Zhou, Ke Chen, Jin Zeng, Yangjun Zhang, Libin Yan, Weimin Yao, et al. 2020. “HnRNP A1 - Mediated Alternative Splicing of CCDC50 Contributes to Cancer Progression of Clear Cell Renal Cell Carcinoma via ZNF395.” *Journal of Experimental & Clinical Cancer Research: CR* 39 (1): 116.
- Suzuki, Toshiharu, Kanae Ando, Toshio Isohara, Masaki Oishi, Gloria S. Lim, Yasushi Satoh, Wilma Wasco, et al. 1997. “Phosphorylation of Alzheimer  $\beta$ -Amyloid Precursor-like Proteins†.” *Biochemistry*. <https://doi.org/10.1021/bi962618k>.
- Szklarczyk, Damian, Annika L. Gable, David Lyon, Alexander Junge, Stefan Wyder, Jaime Huerta-Cepas, Milan Simonovic, et al. 2019. “STRING v11: Protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets.” *Nucleic Acids Research* 47 (D1): D607–13.
- Tabaglio, Tommaso, Diana Hp Low, Winnie Koon Lay Teo, Pierre Alexis Goy, Piotr Cywoniuk, Heike Wollmann, Jessica Ho, et al. 2018. “MBNL1 Alternative Splicing Isoforms Play Opposing Roles in Cancer.” *Life Science Alliance* 1 (5): e201800157.
- Tang, Liang, Peng Zhao, and Dalu Kong. 2019. “Muscleblind-like 1 Destabilizes Snail mRNA and Suppresses the Metastasis of Colorectal Cancer Cells via the Snail/E-cadherin Axis.” *International Journal of Oncology*. <https://doi.org/10.3892/ijo.2019.4691>.
- Tan, Si Kee, Zhen Hua Lin, Cheng Wei Chang, Vipin Varang, Kern Rei Chng, You Fu Pan, Eu Leong Yong, Wing Kin Sung, and Edwin Cheung. 2011. “AP-2 $\gamma$  Regulates Oestrogen Receptor-Mediated Long-Range Chromatin Interaction and Gene Transcription.” *The EMBO Journal* 30 (13): 2569–81.

- Tarallo, Roberta, Angela Bamundo, Giovanni Nassa, Ernesto Nola, Ornella Paris, Concetta Ambrosino, Angelo Facchiano, Marc Baumann, Tuula A. Nyman, and Alessandro Weisz. 2011. "Identification of Proteins Associated with Ligand-Activated Estrogen Receptor  $\alpha$  in Human Breast Cancer Cell Nuclei by Tandem Affinity Purification and Nano LC-MS/MS." *Proteomics* 11 (1): 172–79.
- Tarrero, Lucia Coscujuela, Giulio Ferrero, Valentina Miano, Carlo De Intinis, Laura Ricci, Maddalena Arigoni, Federica Riccardo, et al. 2018. "Luminal Breast Cancer-Specific Circular RNAs Uncovered by a Novel Tool for Data Analysis." *Oncotarget* 9 (18): 14580.
- Taylor, Siân E., Pierre L. Martin-Hirsch, and Francis L. Martin. 2010. "Oestrogen Receptor Splice Variants in the Pathogenesis of Disease." *Cancer Letters* 288 (2): 133–48.
- Terenzi, Fulvia, and Andrea N. Ladd. 2010. "Conserved Developmental Alternative Splicing of Muscleblind-like (MBNL) Transcripts Regulates MBNL Localization and Activity." *RNA Biology* 7 (1): 43–55.
- Theodorou, Vasiliki, Rory Stark, Suraj Menon, and Jason S. Carroll. 2013. "GATA3 Acts Upstream of FOXA1 in Mediating ESR1 Binding by Shaping Enhancer Accessibility." *Genome Research* 23 (1): 12–22.
- Therneau, Terry M., and Patricia M. Grambsch. 2000. "Modeling Survival Data: Extending the Cox Model." *Statistics for Biology and Health*. <https://doi.org/10.1007/978-1-4757-3294-8>.
- Thorsen, Kasper, Karina D. Sørensen, Anne Sofie Brems-Eskildsen, Charlotte Modin, Mette Gaustadnes, Anne-Mette K. Hein, Mogens Kruhøffer, et al. 2008. "Alternative Splicing in Colon, Bladder, and Prostate Cancer Identified by Exon Array Analysis." *Molecular & Cellular Proteomics*. <https://doi.org/10.1074/mcp.m700590-mcp200>.
- Tian, Bin, and James L. Manley. 2017. "Alternative Polyadenylation of mRNA Precursors." *Nature Reviews Molecular Cell Biology* 18 (1): 18–30.
- Tranchevent, Léon-Charles, Fabien Aubé, Louis Dulaurier, Clara Benoit-Pilven, Amandine Rey, Arnaud Poret, Emilie Chautard, et al. 2017. "Identification of Protein Features Encoded by Alternative Exons Using Exon Ontology." *Genome Research* 27 (6): 1087–97.
- Trapnell, Cole, Adam Roberts, Loyal Goff, Geo Pertea, Daehwan Kim, David R. Kelley, Harold Pimentel, Steven L. Salzberg, John L. Rinn, and Lior Pachter. 2012. "Differential Gene and Transcript Expression Analysis of RNA-Seq Experiments with TopHat and Cufflinks." *Nature Protocols*. <https://doi.org/10.1038/nprot.2012.016>.
- Traunmüller, Lisa, Andrea M. Gomez, Thi-Minh Nguyen, and Peter Scheiffele. 2016. "Control of Neuronal Synapse Specification by a Highly Dedicated Alternative Splicing Program." *Science* 352 (6288): 982–86.
- Trincado, Juan L., Juan C. Entizne, Gerald Hysenaj, Babita Singh, Miha Skalic, David J. Elliott, and Eduardo Eyras. 2018. "SUPPA2: Fast, Accurate, and Uncertainty-Aware Differential Splicing Analysis across Multiple Conditions." *Genome Biology* 19 (1): 40.
- Tripathi, Vidisha, Jonathan D. Ellis, Zhen Shen, David Y. Song, Qun Pan, Andrew T. Watt, Susan M. Freier, et al. 2010. "The Nuclear-Retained Noncoding RNA MALAT1 Regulates Alternative Splicing by Modulating SR Splicing Factor Phosphorylation." *Molecular Cell*. <https://doi.org/10.1016/j.molcel.2010.08.011>.

- Tuli, Amit, Mahak Sharma, Xiaojian Wang, Laura C. Simone, Haley L. Capek, Steven Cate, William H. Hildebrand, Naava Naslavsky, Steve Caplan, and Joyce C. Solheim. 2009. "Amyloid Precursor-like Protein 2 Association with HLA Class I Molecules." *Cancer Immunology, Immunotherapy*. <https://doi.org/10.1007/s00262-009-0657-z>.
- Turpin, J., C. Ling, E. J. Crosby, Z. C. Hartman, A. M. Simond, L. A. Chodosh, J. P. Rennhack, et al. 2016. "The ErbB2 $\Delta$ Ex16 Splice Variant Is a Major Oncogenic Driver in Breast Cancer That Promotes a pro-Metastatic Tumor Microenvironment." *Oncogene*. <https://doi.org/10.1038/onc.2016.129>.
- Ule, Jernej, and Robert B. Darnell. 2006. "RNA Binding Proteins and the Regulation of Neuronal Synaptic Plasticity." *Current Opinion in Neurobiology*. <https://doi.org/10.1016/j.conb.2006.01.003>.
- Ungefroren, Hendrik, Ulrich F. Wellner, Tobias Keck, Hendrik Lehnert, and Jens-Uwe Marquardt. 2020. "The Small GTPase RAC1B: A Potent Negative Regulator of- and Useful Tool to Study-TGF $\beta$  Signaling." *Cancers* 12 (11). <https://doi.org/10.3390/cancers12113475>.
- Uszyczynska-Ratajczak, Barbara, Julien Lagarde, Adam Frankish, Roderic Guigó, and Rory Johnson. 2018. "Towards a Complete Map of the Human Long Non-Coding RNA Transcriptome." *Nature Reviews. Genetics* 19 (9): 535–48.
- Valdivia, H. H. 2007. "One Gene, Many Proteins: Alternative Splicing of the Ryanodine Receptor Gene Adds Novel Functions to an Already Complex Channel Protein." *Circulation Research* 100 (6). <https://doi.org/10.1161/01.RES.0000263400.64391.37>.
- Van Nostrand, Eric L., Peter Freese, Gabriel A. Pratt, Xiaofeng Wang, Xintao Wei, Rui Xiao, Steven M. Blue, et al. 2020. "A Large-Scale Binding and Functional Map of Human RNA-Binding Proteins." *Nature* 583 (7818): 711–19.
- Vaquero-Garcia, Jorge, Alejandro Barrera, Matthew R. Gazzara, Juan González-Vallinas, Nicholas F. Lahens, John B. Hogenesch, Kristen W. Lynch, and Yoseph Barash. 2016. "A New View of Transcriptome Complexity and Regulation through the Lens of Local Splicing Variations." *eLife* 5 (February): e11752.
- Vasaikar, Suhas V., Abhijeet P. Deshmukh, Petra den Hollander, Sridevi Addanki, Nick Allen Kuburich, Sriya Kudaravalli, Robiya Joseph, Jeffrey T. Chang, Rama Soundararajan, and Sendurai A. Mani. 2021. "EMTome: A Resource for Pan-Cancer Analysis of Epithelial-Mesenchymal Transition Genes and Signatures." *British Journal of Cancer* 124 (1): 259–69.
- Vilá de Muga, S., P. Timpson, L. Cubells, R. Evans, T. E. Hayes, C. Rentero, A. Hegemann, et al. 2009. "Annexin A6 Inhibits Ras Signalling in Breast Cancer Cells." *Oncogene* 28 (3): 363–77.
- Villemin, Jean-Philippe, Claudio Lorenzi, Marie-Sarah Cabrillac, Andrew Oldfield, William Ritchie, and Reini F. Luco. 2021. "A Cell-to-Patient Machine Learning Transfer Approach Uncovers Novel Basal-like Breast Cancer Prognostic Markers amongst Alternative Splice Variants." *BMC Biology* 19 (1): 70.
- Vitting-Seerup, Kristoffer, and Albin Sandelin. 2017a. "The Landscape of Isoform Switches in Human Cancers." *Molecular Cancer Research*. <https://doi.org/10.1158/1541-7786.mcr-16-0459>.
- . 2017b. "The Landscape of Isoform Switches in Human Cancers." *Molecular Cancer Research: MCR* 15 (9): 1206–20.

- . 2019. “IsoformSwitchAnalyzeR: Analysis of Changes in Genome-Wide Patterns of Alternative Splicing and Its Functional Consequences.” *Bioinformatics* 35 (21): 4469–71.
- Voena, Claudia, Lydia M. Varesio, Liye Zhang, Matteo Menotti, Teresa Poggio, Elena Panizza, Qi Wang, et al. 2016. “Oncogenic ALK Regulates EMT in Non-Small Cell Lung Carcinoma through Repression of the Epithelial Splicing Regulatory Protein 1.” *Oncotarget* 7 (22): 33316–30.
- Vos, Seychelle M., Lucas Farnung, Marc Boehning, Christoph Wigge, Andreas Linden, Henning Urlaub, and Patrick Cramer. 2018. “Structure of Activated Transcription Complex Pol II–DSIF–PAF–SPT6.” *Nature*. <https://doi.org/10.1038/s41586-018-0440-4>.
- Vrtačnik, Peter, Barbara Ostanek, Simona Mencej-Bedrač, and Janja Marc. 2014. “The Many Faces of Estrogen Signaling.” *Biochemia Medica: Casopis Hrvatskoga Društva Medicinskih Biokemicara / HDMB* 24 (3): 329–42.
- Vuong, Celine K., Douglas L. Black, and Sika Zheng. 2016. “The Neurogenetics of Alternative Splicing.” *Nature Reviews Neuroscience*. <https://doi.org/10.1038/nrn.2016.27>.
- Wang, Eric T., Rickard Sandberg, Shujun Luo, Irina Khrebtkova, Lu Zhang, Christine Mayr, Stephen F. Kingsmore, Gary P. Schroth, and Christopher B. Burge. 2008. “Alternative Isoform Regulation in Human Tissue Transcriptomes.” *Nature* 456 (7221): 470–76.
- Wang, Hong, Chris Zhiyi Zhang, Shi-Xun Lu, Mei-Fang Zhang, Li-Li Liu, Rong-Zhen Luo, Xia Yang, et al. 2019. “A Coiled-Coil Domain Containing 50 Splice Variant Is Modulated by Serine/Arginine-Rich Splicing Factor 3 and Promotes Hepatocellular Carcinoma in Mice by the Ras Signaling Pathway.” *Hepatology* 69 (1): 179–95.
- Wang, Jingjing, Xianwen Meng, Hongjun Chen, Chunhui Yuan, Xue Li, Yincong Zhou, and Ming Chen. 2016. “Exploring the Mechanisms of Genome-Wide Long-Range Interactions: Interpreting Chromosome Organization.” *Briefings in Functional Genomics* 15 (5): 385–95.
- Wang, Yiwei, Zipeng Liu, and Jian Shen. 2019. “MicroRNA-421-Targeted PDCD4 Regulates Breast Cancer Cell Proliferation.” *International Journal of Molecular Medicine* 43 (1): 267–75.
- Wan, Yilv, Feng Liang, Minjun Wei, and Ying Liu. 2020. “Long Non-Coding RNA LINC00525 Regulates the Proliferation and Epithelial to Mesenchymal Transition of Human Glioma Cells by Sponging miR-338-3p.” *AMB Express*. <https://doi.org/10.1186/s13568-020-01094-4>.
- Wärnmark, Anette, Eckardt Treuter, Anthony P. H. Wright, and Jan-Ake Gustafsson. 2003. “Activation Functions 1 and 2 of Nuclear Receptors: Molecular Strategies for Transcriptional Activation.” *Molecular Endocrinology* 17 (10): 1901–9.
- Warren, Chloe F. A., Michelle W. Wong-Brown, and Nikola A. Bowden. 2019. “BCL-2 Family Isoforms in Apoptosis and Cancer.” *Cell Death & Disease* 10 (3): 177.
- Warzecha, Claude C., Peng Jiang, Karine Amirikian, Kimberly A. Dittmar, Hezhe Lu, Shihao Shen, Wei Guo, Yi Xing, and Russ P. Carstens. 2010. “An ESRP-Regulated Splicing Programme Is Abrogated during the Epithelial-Mesenchymal Transition.” *The EMBO Journal* 29 (19): 3286–3300.

- Warzecha, Claude C., Shihao Shen, Yi Xing, and Russ P. Carstens. 2009. "The Epithelial Splicing Factors ESRP1 and ESRP2 Positively and Negatively Regulate Diverse Types of Alternative Splicing Events." *RNA Biology* 6 (5): 546–62.
- Watermann, Dirk O., Yesheng Tang, Axel zur Hausen, Markus Jäger, Stefan Stamm, and Elmar Stickeler. 2006. "Splicing Factor Tra2- $\beta$ 1 Is Specifically Induced in Breast Cancer and Regulates Alternative Splicing of the CD44 Gene." *Cancer Research*. <https://doi.org/10.1158/0008-5472.can-04-3294>.
- Weigelt, Britta, and Jorge S. Reis-Filho. 2013. "Activating Mutations in HER2: Neu Opportunities and Neu Challenges." *Cancer Discovery*. <https://doi.org/10.1158/2159-8290.cd-12-0585>.
- Wen, Ji, Kevin H. Toomer, Zhibin Chen, and Xiaodong Cai. 2015. "Genome-Wide Analysis of Alternative Transcripts in Human Breast Cancer." *Breast Cancer Research and Treatment* 151 (2): 295–307.
- Whalen, Diva S., Sarrah E. Widatalla, Olga Y. Korolkova, Gladys S. Nangami, Heather K. Beasley, Stephen D. Williams, Carlos Virgous, Brian D. Lehmann, Josiah Ochieng, and Amos M. Sakwe. 2019. "Implication of Calcium Activated RasGRF2 in Annexin A6-Mediated Breast Tumor Cell Growth and Motility." *Oncotarget* 10 (2): 133–51.
- Widatalla, Sarrah E., Olga Y. Korolkova, Diva S. Whalen, J. Shawn Goodwin, Kevin P. Williams, Josiah Ochieng, and Amos M. Sakwe. 2019. "Lapatinib-Induced Annexin A6 Upregulation as an Adaptive Response of Triple-Negative Breast Cancer Cells to EGFR Tyrosine Kinase Inhibitors." *Carcinogenesis* 40 (8): 998–1009.
- Wilkinson, Leland. 2011. "ggplot2: Elegant Graphics for Data Analysis by WICKHAM, H." *Biometrics*. <https://doi.org/10.1111/j.1541-0420.2011.01616.x>.
- Wilusz, C. J., M. Wormington, and S. W. Peltz. 2001. "The Cap-to-Tail Guide to mRNA Turnover." *Nature Reviews. Molecular Cell Biology* 2 (4): 237–46.
- Woodley, Louise, and Juan Valcárcel. 2002. "Regulation of Alternative Pre-mRNA Splicing." *Briefings in Functional Genomics & Proteomics* 1 (3): 266–77.
- Wu, Douglas C., Jun Yao, Kevin S. Ho, Alan M. Lambowitz, and Claus O. Wilke. 2018. "Limitations of Alignment-Free Tools in Total RNA-Seq Quantification." *BMC Genomics* 19 (1): 510.
- Wutz, Anton. 2011. "Gene Silencing in X-Chromosome Inactivation: Advances in Understanding Facultative Heterochromatin Formation." *Nature Reviews. Genetics* 12 (8): 542–53.
- Wu, Xiang-Song, Fang Wang, Huai-Feng Li, Yun-Ping Hu, Lin Jiang, Fei Zhang, Mao-Lan Li, et al. 2017. "LncRNA-PAGBC Acts as a microRNA Sponge and Promotes Gallbladder Tumorigenesis." *EMBO Reports* 18 (10): 1837–53.
- Xia, Heng, Dong Chen, Qijia Wu, Gang Wu, Yanhong Zhou, Yi Zhang, and Libin Zhang. 2017. "CELF1 Preferentially Binds to Exon-Intron Boundary and Regulates Alternative Splicing in HeLa Cells." *Biochimica et Biophysica Acta, Gene Regulatory Mechanisms* 1860 (9): 911–21.
- Xiang, Jian-Feng, Qing-Fei Yin, Tian Chen, Yang Zhang, Xiao-Ou Zhang, Zheng Wu, Shaofeng Zhang, et al. 2014. "Human Colorectal Cancer-Specific CCAT1-L lncRNA Regulates Long-Range Chromatin Interactions at the MYC Locus." *Cell Research* 24 (5): 513–31.

- Xiao, Jian, Linling Lin, Dakui Luo, Liang Shi, Wangwang Chen, Hao Fan, Zengliang Li, et al. 2020. “Long Noncoding RNA TRPM2-AS Acts as a microRNA Sponge of miR-612 to Promote Gastric Cancer Progression and Radioresistance.” *Oncogenesis* 9 (3): 29.
- Xue, Zhuyi, René L. Warren, Ewan A. Gibb, Daniel MacMillan, Johnathan Wong, Readman Chiu, S. Austin Hammond, et al. 2018. “Recurrent Tumor-Specific Regulation of Alternative Polyadenylation of Cancer-Related Genes.” *BMC Genomics* 19 (1): 536.
- Xu, Hang, Shijie Zhang, Xianfu Yi, Dariusz Plewczynski, and Mulin Jun Li. 2020. “Exploring 3D Chromatin Contacts in Gene Regulation: The Evolution of Approaches for the Identification of Functional Enhancer-Promoter Interaction.” *Computational and Structural Biotechnology Journal* 18 (February): 558–70.
- Xu, Qiang, Barmak Modrek, and Christopher Lee. 2002. “Genome-Wide Detection of Tissue-Specific Alternative Splicing in the Human Transcriptome.” *Nucleic Acids Research* 30 (17): 3754–66.
- Yae, Toshifumi, Kenji Tsuchihashi, Takatsugu Ishimoto, Takeshi Motohara, Momoko Yoshikawa, Go J. Yoshida, Takeyuki Wada, et al. 2012. “Alternative Splicing of CD44 mRNA by ESRP1 Enhances Lung Colonization of Metastatic Cancer Cell.” *Nature Communications*. <https://doi.org/10.1038/ncomms1892>.
- Yamnik, Rachel L., Alla Digilova, Daphne C. Davis, Z. Nilly Brodt, Christopher J. Murphy, and Marina K. Holz. 2009. “S6 Kinase 1 Regulates Estrogen Receptor  $\alpha$  in Control of Breast Cancer Cell Proliferation.” *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.m807532200>.
- Yang, Xiping, Jasmin Coulombe-Huntington, Shuli Kang, Gloria M. Sheynkman, Tong Hao, Aaron Richardson, Song Sun, et al. 2016. “Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing.” *Cell* 164 (4): 805–17.
- Yang, Yan, Jian Zhu, Tiantian Zhang, Jing Liu, Yumei Li, Yue Zhu, Lingjie Xu, et al. 2018. “PHD-Finger Domain Protein 5A Functions as a Novel Oncoprotein in Lung Adenocarcinoma.” *Journal of Experimental & Clinical Cancer Research*. <https://doi.org/10.1186/s13046-018-0736-0>.
- Yang, Yueqin, Juwon Park, Thomas W. Bebee, Claude C. Warzecha, Yang Guo, Xuequn Shang, Yi Xing, and Russ P. Carstens. 2016. “Determination of a Comprehensive Alternative Splicing Regulatory Network and Combinatorial Regulation by Key Factors during the Epithelial-to-Mesenchymal Transition.” *Molecular and Cellular Biology*. <https://doi.org/10.1128/mcb.00019-16>.
- Yao, Run-Wen, Yang Wang, and Ling-Ling Chen. 2019. “Cellular Functions of Long Noncoding RNAs.” *Nature Cell Biology*. <https://doi.org/10.1038/s41556-019-0311-8>.
- Yap, Kyoko L., Side Li, Ana M. Muñoz-Cabello, Selina Raguz, Lei Zeng, Shiraz Mujtaba, Jesús Gil, Martin J. Walsh, and Ming-Ming Zhou. 2010. “Molecular Interplay of the Noncoding RNA ANRIL and Methylated Histone H3 Lysine 27 by Polycomb CBX7 in Transcriptional Silencing of INK4a.” *Molecular Cell* 38 (5): 662–74.
- Yen, Meng-Chi, Yung-Chi Huang, Jung-Yu Kan, Po-Lin Kuo, Ming-Feng Hou, and Ya-Ling Hsu. 2018. “S100B Expression in Breast Cancer as a Predictive Marker for Cancer Metastasis.” *International Journal of Oncology* 52 (2): 433–40.

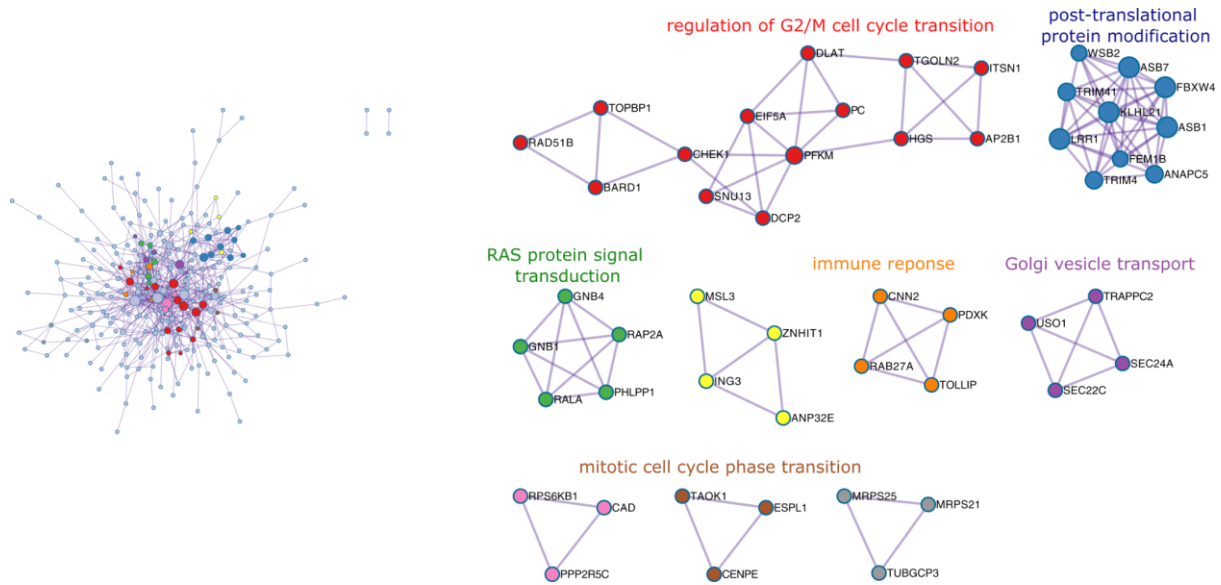
- Yi, Ping, Zhao Wang, and Bert W. O'Malley. 2019. "Structural Studies with Coactivators for the Estrogen Receptor." *Estrogen Receptor and Breast Cancer*. [https://doi.org/10.1007/978-3-319-99350-8\\_4](https://doi.org/10.1007/978-3-319-99350-8_4).
- Zeng, Chao, and Michiaki Hamada. 2018. "Identifying Sequence Features That Drive Ribosomal Association for lncRNA." *BMC Genomics* 19 (Suppl 10): 906.
- Zhang, Bin, Kamesh R. Babu, Chun You Lim, Zhi Hao Kwok, Jia Li, Siqin Zhou, Henry Yang, and Yvonne Tay. 2020. "A Comprehensive Expression Landscape of RNA-Binding Proteins (RBPs) across 16 Human Cancer Types." *RNA Biology* 17 (2): 211–26.
- Zhang, Chi, Baohong Zhang, Michael S. Vincent, and Shanrong Zhao. 2016. "Bioinformatics Tools for RNA-Seq Gene and Isoform Quantification." *Journal of Next Generation Sequencing & Applications*. <https://doi.org/10.4172/2469-9853.1000140>.
- Zhang, Honghong, Rhonda L. Brown, Yong Wei, Pu Zhao, Sali Liu, Xuan Liu, Yu Deng, et al. 2019. "CD44 Splice Isoform Switching Determines Breast Cancer Stem Cell State." *Genes & Development*. <https://doi.org/10.1101/gad.319889.118>.
- Zhang, Minzhe, Tao Wang, Guanghua Xiao, and Yang Xie. 2020. "Large-Scale Profiling of RBP-circRNA Interactions from Public CLIP-Seq Datasets." *Genes* 11 (1). <https://doi.org/10.3390/genes11010054>.
- Zhang, Peng, Xiao-Ou Zhang, Tingting Jiang, Lingling Cai, Xiao Huang, Qi Liu, Dan Li, et al. 2020. "Comprehensive Identification of Alternative Back-Splicing in Human Tissue Transcriptomes." *Nucleic Acids Research* 48 (4): 1779–89.
- Zhang, Shanyong, Lei Ding, Feng Gao, and Hongwu Fan. 2020. "Long Non-Coding RNA DSCAM-AS1 Upregulates Expression through Sponging miR-101-3p to Accelerate Osteosarcoma Progression." *Biochemistry and Cell Biology = Biochimie et Biologie Cellulaire* 98 (5): 600–611.
- Zhang, Siyuan, Wen-Chien Huang, Ping Li, Hua Guo, Say-Bee Poh, Samuel W. Brady, Yan Xiong, et al. 2011. "Combating Trastuzumab Resistance by Targeting SRC, a Common Node Downstream of Multiple Resistance Pathways." *Nature Medicine*. <https://doi.org/10.1038/nm.2309>.
- Zhang, Xiaona, Yanchun Zhou, Shaoying Chen, Wei Li, Weibing Chen, and Wei Gu. 2019. "LncRNA MACC1-AS1 Sponges Multiple miRNAs and RNA-Binding Protein PTBP1." *Oncogenesis* 8 (12): 73.
- Zhang, Yin, Yong-Xin Huang, Dan-Lan Wang, Bing Yang, Hai-Yan Yan, Le-Hang Lin, Yun Li, et al. 2020. "LncRNA DSCAM-AS1 Interacts with YBX1 to Promote Cancer Progression by Forming a Positive Feedback Loop That Activates FOXA1 Transcription Network." *Theranostics* 10 (23): 10823–37.
- Zhao, Shanrong, Ying Zhang, William Gordon, Jie Quan, Hualin Xi, Sarah Du, David von Schack, and Baohong Zhang. 2015. "Comparison of Stranded and Non-Stranded RNA-Seq Transcriptome Profiling and Investigation of Gene Overlap." *BMC Genomics* 16 (September): 675.
- Zhao, Zhenzhou, Xuejie Li, Chuanyu Gao, Dongdong Jian, Peiyuan Hao, Lixin Rao, and Muwei Li. 2017. "Peripheral Blood Circular RNA hsa\_circ\_0124644 Can Be Used as a Diagnostic Biomarker of Coronary Artery Disease." *Scientific Reports*. <https://doi.org/10.1038/srep39918>.

- Zheng, Rongbin, Changxin Wan, Shenglin Mei, Qian Qin, Qiu Wu, Hanfei Sun, Chen-Hao Chen, et al. 2019. "Cistrome Data Browser: Expanded Datasets and New Tools for Gene Regulatory Analysis." *Nucleic Acids Research* 47 (D1): D729–35.
- Zheng, Yi-Zi, Meng-Zhu Xue, Hong-Jie Shen, Xiao-Guang Li, Ding Ma, Yue Gong, Yi-Rong Liu, et al. 2018. "PHF5A Epigenetically Inhibits Apoptosis to Promote Breast Cancer Progression." *Cancer Research*. <https://doi.org/10.1158/0008-5472.can-17-3514>.
- Zhou, Baoguo, Yuli Wang, Jinpeng Jiang, Hongpeng Jiang, Jianwei Song, Taotao Han, Juan Shi, and Haiquan Qiao. 2016. "The Long Noncoding RNA Colon Cancer-Associated Transcript-1/miR-490 Axis Regulates Gastric Cancer Cell Migration by Targeting hnRNPA1." *IUBMB Life*. <https://doi.org/10.1002/iub.1474>.
- Zhou, C., S. Licciulli, J. L. Avila, M. Cho, S. Troutman, P. Jiang, A. V. Kossenkov, et al. 2013. "The Rac 1 Splice Form Rac 1b Promotes K-Ras-Induced Lung Tumorigenesis." *Oncogene* 32 (7): 903–9.
- Zhou, Yingyao, Bin Zhou, Lars Pache, Max Chang, Alireza Hadj Khodabakhshi, Olga Tanaseichuk, Christopher Benner, and Sumit K. Chanda. 2019. "Metascape Provides a Biologist-Oriented Resource for the Analysis of Systems-Level Datasets." *Nature Communications* 10 (1): 1523.
- Zhu, Yumin, Gang Xu, Yucheng T. Yang, Zhiyu Xu, Xinduo Chen, Binbin Shi, Daoxin Xie, Zhi John Lu, and Pengyuan Wang. 2019. "POSTAR2: Deciphering the Post-Transcriptional Regulatory Logics." *Nucleic Acids Research* 47 (D1): D203–11.
- Zikherman, Julie, and Arthur Weiss. 2008. "Alternative Splicing of CD45: The Tip of the Iceberg." *Immunity*.
- Zinn, Rabea, Hannah Otterbein, Hendrik Lehnert, and Hendrik Ungefroren. 2019. "RAC1B: A Guardian of the Epithelial Phenotype and Protector Against Epithelial-Mesenchymal Transition." *Cells* 8 (12). <https://doi.org/10.3390/cells8121569>.
- Zong, Feng-Yang, Xing Fu, Wen-Juan Wei, Ya-Ge Luo, Monika Heiner, Li-Juan Cao, Zhaoyuan Fang, et al. 2014. "The RNA-Binding Protein QKI Suppresses Cancer-Associated Aberrant Splicing." *PLoS Genetics* 10 (4): e1004289.
- Zuckerman, Binyamin, Maya Ron, Martin Mikl, Eran Segal, and Igor Ulitsky. 2020. "Gene Architecture and Sequence Composition Underpin Selective Dependency of Nuclear Export of Long RNAs on NXF1 and the TREX Complex." *Molecular Cell* 79 (2): 251–67.e6.
- Zwart, Wilbert, Alexander Griekspoor, Valeria Berno, Kim Lakeman, Kees Jalink, Michael Mancini, Jacques Neefjes, and Rob Michalides. 2007. "PKA-Induced Resistance to Tamoxifen Is Associated with an Altered Orientation of ERalpha towards Co-Activator SRC-1." *The EMBO Journal* 26 (15): 3534–44.

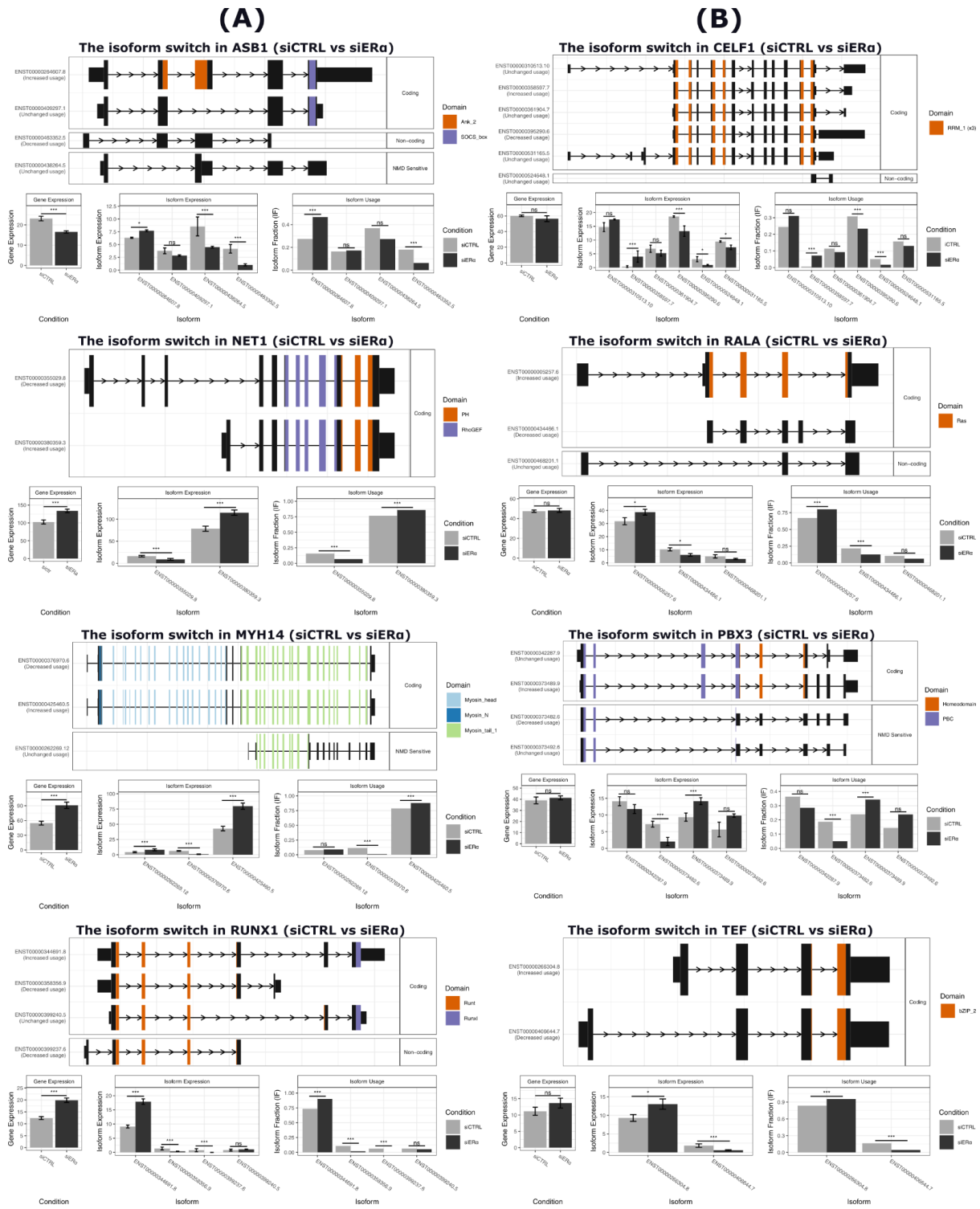


## Appendices

### Appendix A: Supplementary Figures :



**Supplementary Figure 1:** Protein-protein interaction Network of genes harbouring isoform switching events with downstream consequences upon ERa silencing. The small subnetworks are the most significant protein complexes identified by MCODE algorithm from Metascape (Zhou et al. 2019).

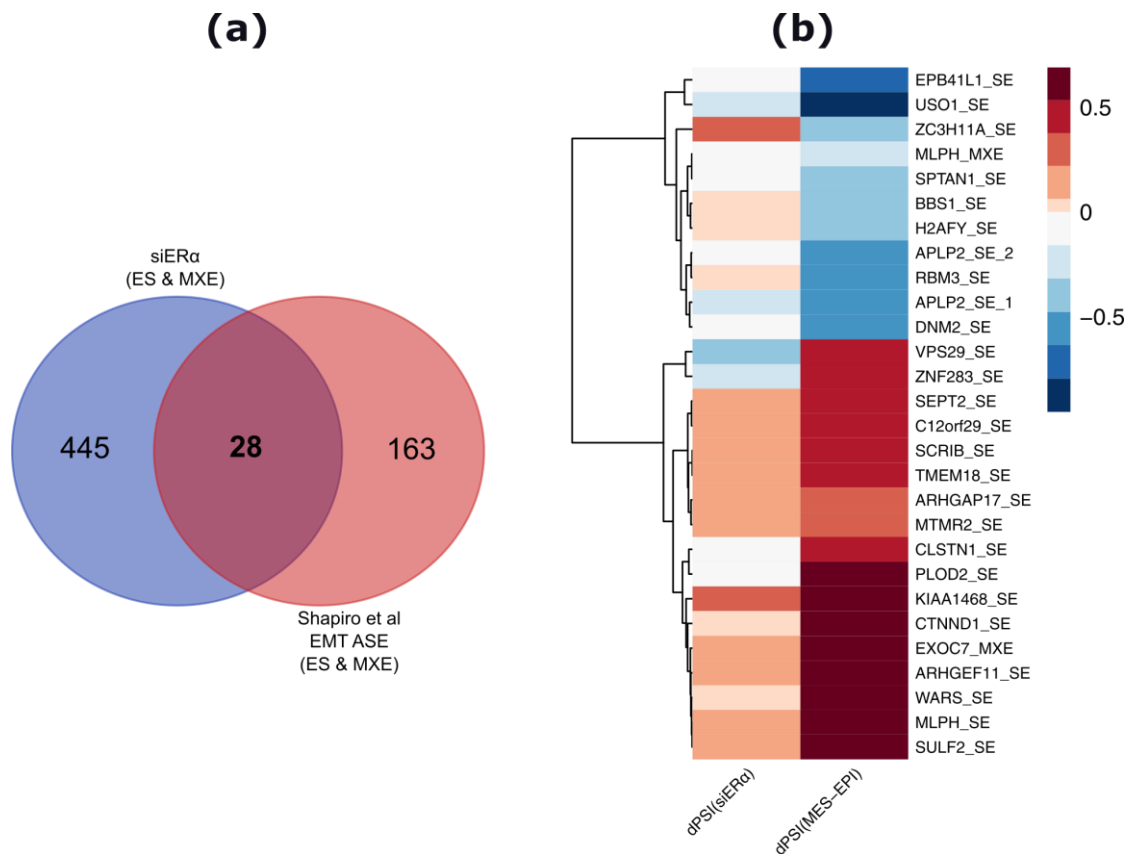


**Supplementary Figure 2:** 8 selected example genes with isoform expression switching behaviour induced by ERA silencing. (A) examples of regulated genes (up or down) with isoforms responding in opposite direction to ERA silencing. (B) examples of genes not regulated at gene level, but regulated at isoform level.

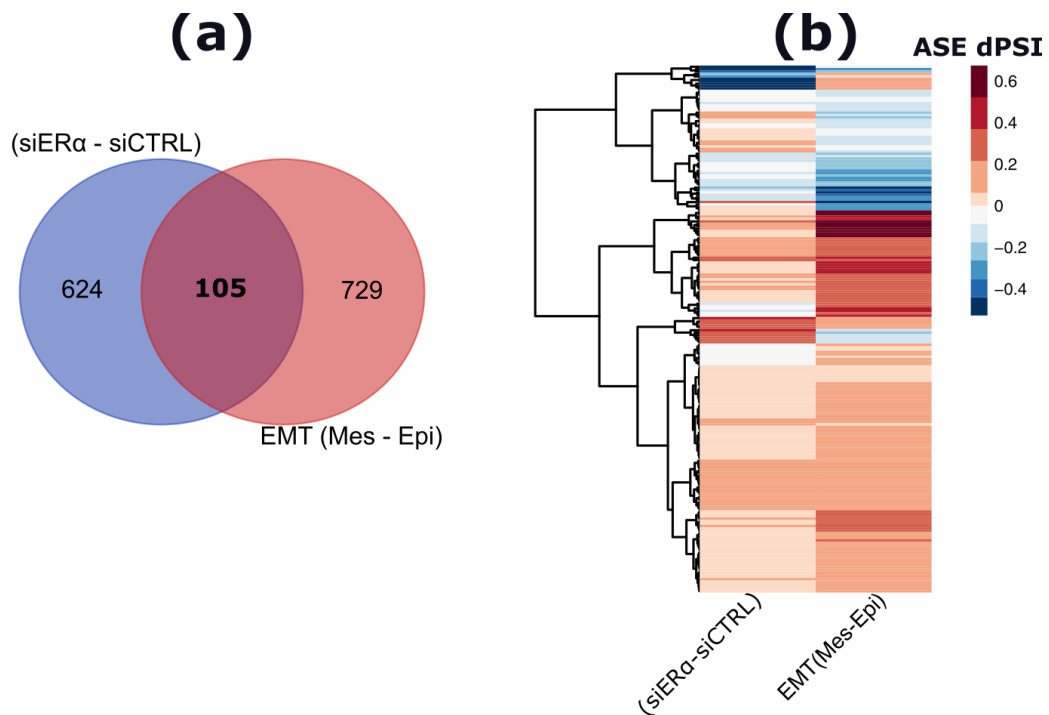
In these supplementary plots, selected examples of genes with an isoform differential usage behaviour are reported. For instance, on panel A, a significant switch in the ankyrin repeat and SOCS box containing 1 (*ASB1*) gene isoforms resulted in the upregulation of the longer, canonical isoform

(ENST00000264607.8) and downregulation of the non-coding (ENST00000463352.5) and NMD-sensitive (ENST00000438264.5) isoforms. Consequently, this switch enriched the expression of isoforms with coding potential, having two protein domain encoding exons, insensitive to NMD, and having longer 5'UTR and 3'UTR regions. Another isoform switching event representing a different type of isoform expression regulation by alternative promoter usage is the neuroepithelial cell transforming 1 (*NET1*) gene where the longest isoform (ENST00000355029.8) is repressed while the shortest isoform (ENST00000380359.3) is being induced. Another switching event representing the skipping of a single exon encoding for a functional protein domain (Myosin\_head) is in the Myosin heavy chain 14 (*MYH14*) isoforms, where the inclusion form of the exon (ENST00000376970.6) is repressed while the exclusion form (ENST00000425460.5) is induced. Interestingly, the annotation of the skipped exon in this switching event using DIGGER (Louadi et al. 2021) reveals that the protein domain encoded by this exon is implicated in 12 PPIs, 9 of which are partially affected (30% to 67% lost domain-domain interaction) and 3 PPIs are completely (100% lost domain-domain interactions) suppressed by the skipping event. Another switching event in the runt related transcription factor 1 (*RUNX1*) illustrates the differential usage of the last exon (ENST00000344691.8) is induced and (ENST00000358356.9) is repressed, which consequently results in the expression of a novel protein domain (Runx inhibition domain, Runxl). On the other hand, as shown in panel **B**, cases also included switching genes that show no regulation at gene level, but are exclusively regulated at isoform level only. For instance, the splicing factor encoding gene *CELF1* which is not regulated at gene level has two isoforms responding in opposite directions to ER $\alpha$  silencing, resulting in the enrichment of isoforms with shorter 3'UTR and IR loss events. Another example is the RAS like proto-oncogene A (*RALA*) gene which while it was not regulated at gene level, it had a significant switching event at the level of isoforms, promoting the expression of the principal, NMD-insensitive isoform (ENST00000005257.6) and repressing the NMD-sensitive isoform (ENST00000434466.1). This switch in the *RALA* gene results in the enrichment of the protein domain Ras (PF00071). The annotation of the Ras domain revealed that it is involved in 24 different PPIs. A similar isoform switching event promoting the expression of the NMD-insensitive isoform (ENST00000373489.9) and repressing the NMD-sensitive one (ENST00000373482.6) occurred in the PBX homeobox 3 (*PBX3*) gene. In particular, the fifth exon (ENSE00003699340) of the upregulated, NMD-insensitive isoform (ENST00000373482.6) encodes for two important protein domains (Homeodomain and Homeobox\_KN) involved in 6 and 4 PPIs, respectively, including the interaction with the splicing factor hnRNPL. Importantly, the exon (ENSE00003699340) encodes for

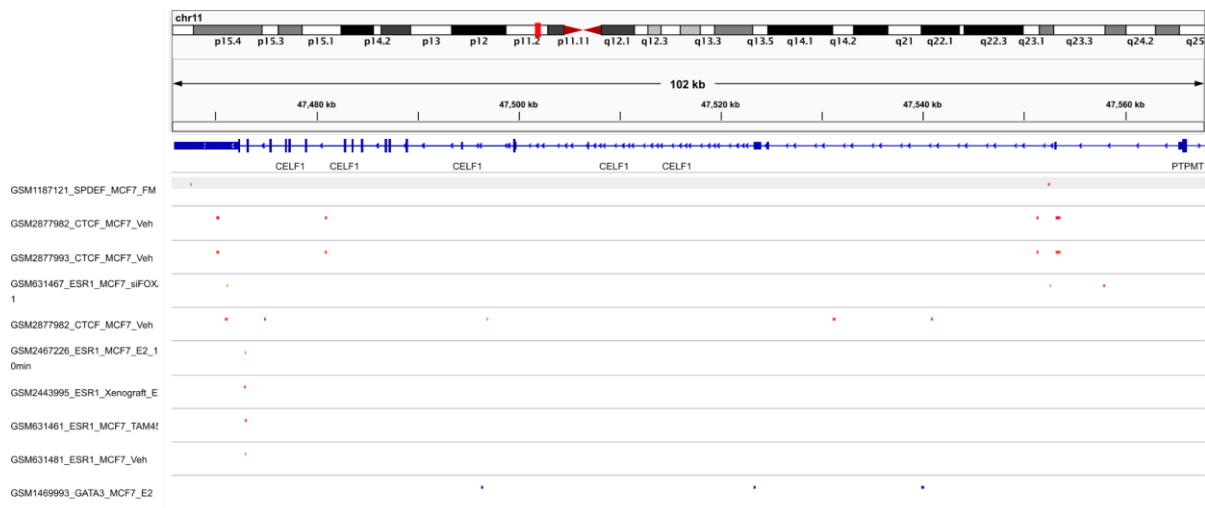
the residues that lie on the active sites involved in three PPIs including the interaction with HOXB1, HOXA9, and the HOXA10-HOXA9 readthrough proteins. Together, these examples illustrate the importance of considering isoform regulation rather than focussing the attention on gene expression changes alone.



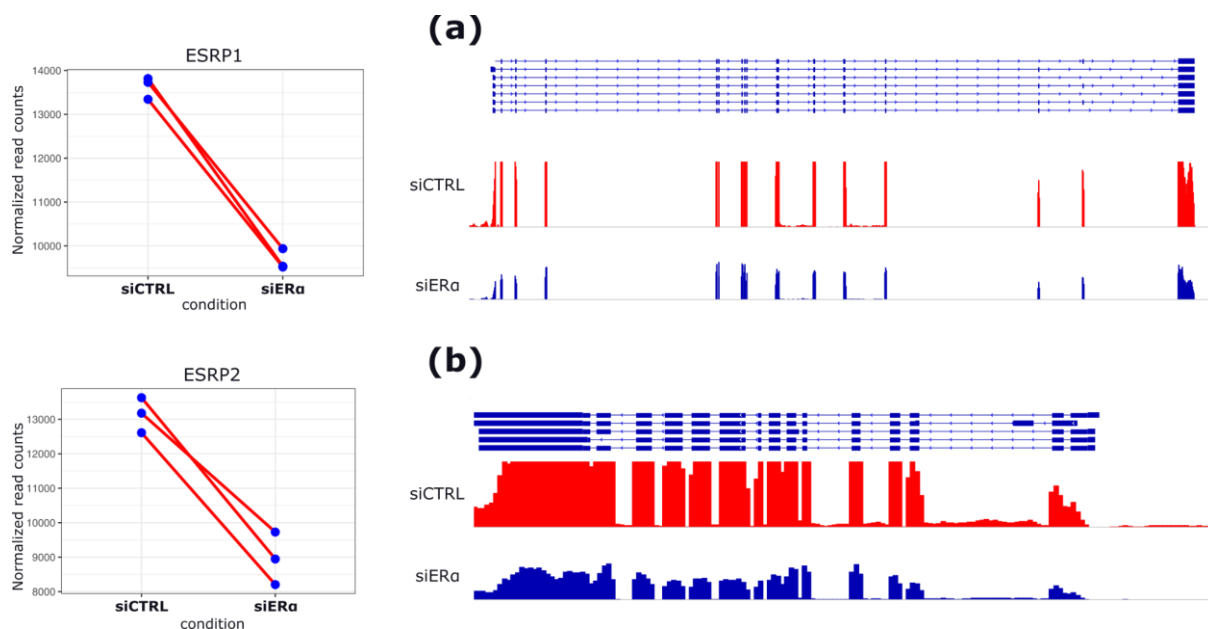
**Supplementary Figure 3:** (a) Number of ERα-ASEs (ES and MXE) identified in this study overlapping the ASEs reported by (Shapiro et al. 2011) to be regulated during EMT process comparing. (b) Heat map reporting the coherence of the dPSI of overlapping events. In (Shapiro et al. 2011), the inclusion level of ASEs was calculated comparing mesenchymal to epithelial BC cells dPSI(Mes - Epi). The comparison is limited to ES and MXE events as they were the only events provided by the authors. Full list of events identified in (Shapiro et al. 2011) is provided in **Supplementary Table 7**.



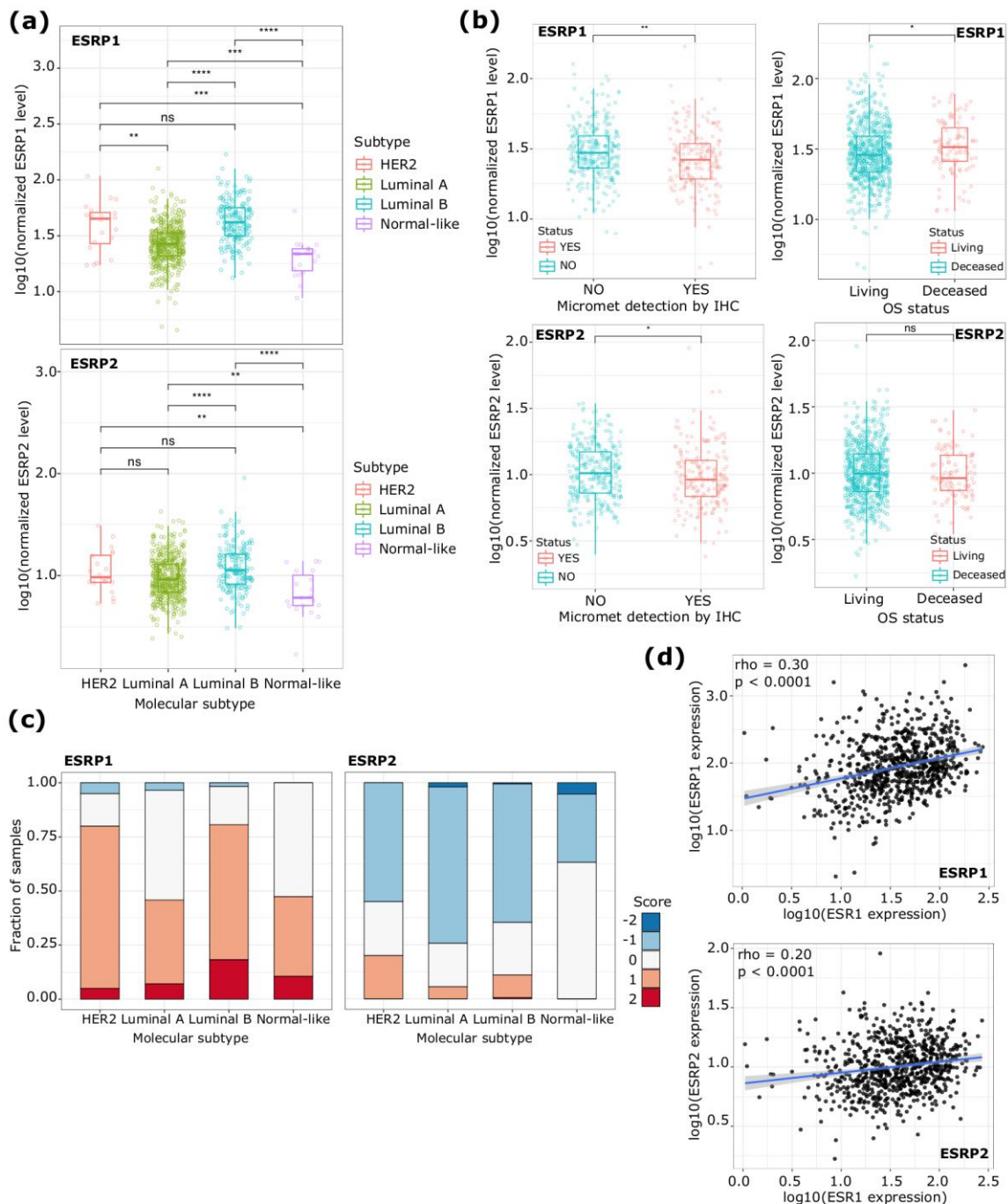
**Supplementary Figure 4:** Overlapping AEs between siRNA-mediated silencing of ER $\alpha$  in hormone deprived MCF-7 cells and EMT time course RNA-seq from (Yueqin Yang et al. 2016) study. EMT was induced by the stable overexpression of ZEB1 in epithelial cells. Total RNA was extracted and changes of AS patterns were measured over a 7-days time course. **(a)** Venn diagram reporting the number of overlapping AEs. **(b)** Heat map plot showing the dPSI (siER $\alpha$  - siCTRL) and (Mes - Epi) of overlapping AEs shown in **(a)**.



**Supplementary Figure 5:** Screenshot of the integrative genome viewer (IGV) reporting ChIP-Seq binding peaks of selected TFs including  $ER\alpha$ , SPDEF, CTCF, and GATA3 showing a binding peak at the body and 3'UTR region of CELF1 in MCF-7 BC cells under Full Medium (FM), Vehicle (Veh), or under estrogen (E2) treatment. The 3'UTR region of CELF1 is differentially spliced under  $ER\alpha$  silencing.

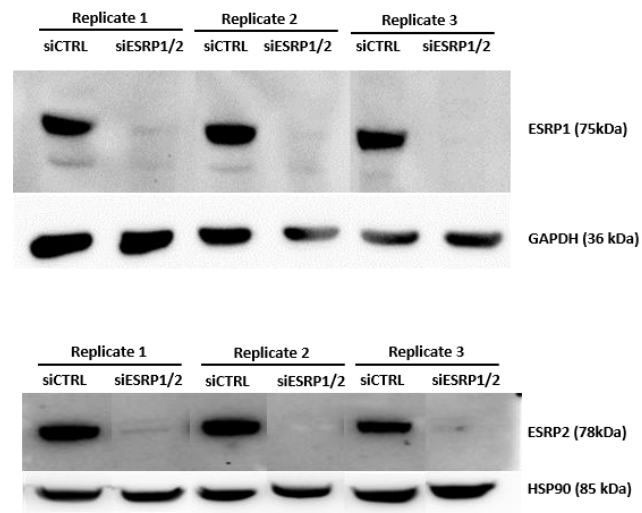


**Supplementary figure 6:** Overview of gene expression changes of ESRP1 and ESRP2 genes upon the silencing of ERA in MCF-7 cells cultured under hormone deprivation conditions. **(a-b)** left panels: line plot reporting the expression levels in normalized read counts of ESRP1 and ESRP2 genes in the triplicates of siCTRL and siERA conditions, respectively. Right panels: genome representation taken from the interactive genome browser (IGV) showing the read coverage signal over the gene body of ESRP1 and ESRP2 in both conditions, respectively.

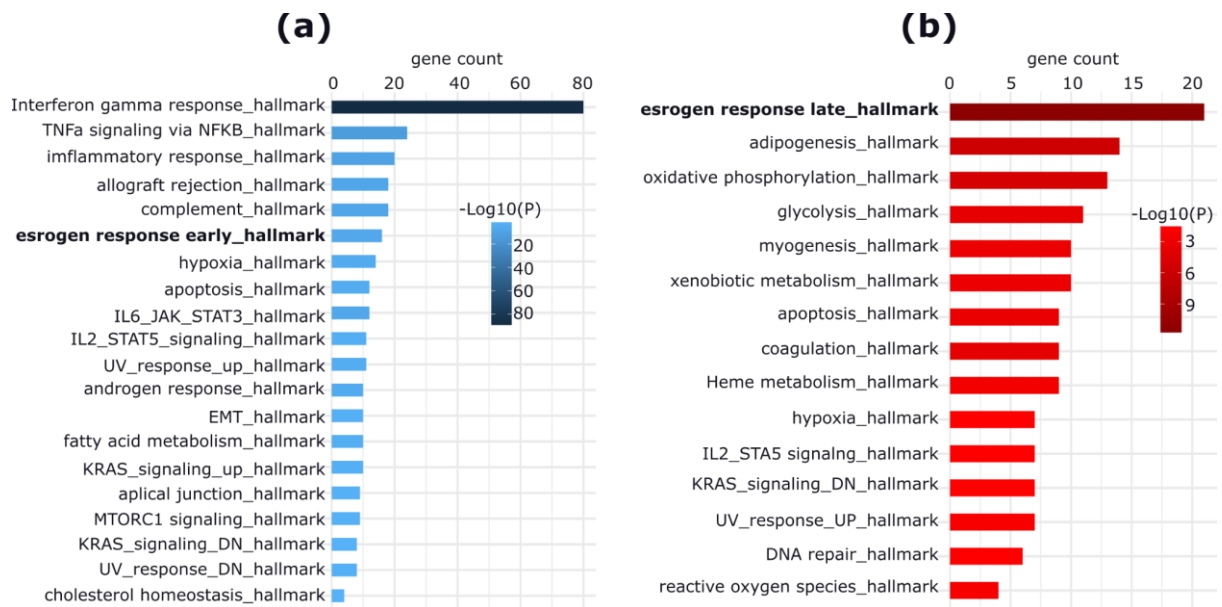


**Supplementary figure 7:** Gene expression and genome wide alteration analysis of both ESRP1 and ESRP2 genes in BCs classified based on their molecular profiles. **(a)** Box plots reporting the expression level of ESRP1 (top) and ESRP2 (bottom) genes, respectively (y-axis), in different BC subtypes including HER2-positive, luminal A and B, and normal-like subtypes (x-axis). (Wilcoxon p-value, ns, not significant; \*\*, p < 0.01; \*\*\*, p < 0.0001; \*\*\*\*, p < 0.00001). **(b)** Boxplots representing the expression levels of ESRP1 (top) and ESRP2 (bottom) genes as a function of micrometastasis detection and overall survival time of ERa+ patients. **(c)** copy number variation profiles of ESRP1 (left) and ESRP2 (right) genes in ER+ samples classified based on their molecular profiles into HER2-positive, luminal A and B, and normal-like subtypes, respectively. **(d)** Scatter plots reporting the gene expression correlation analysis performed for ESRP1 (top), and ESRP2 (bottom) and ESR1 gene in ER+ BCs.

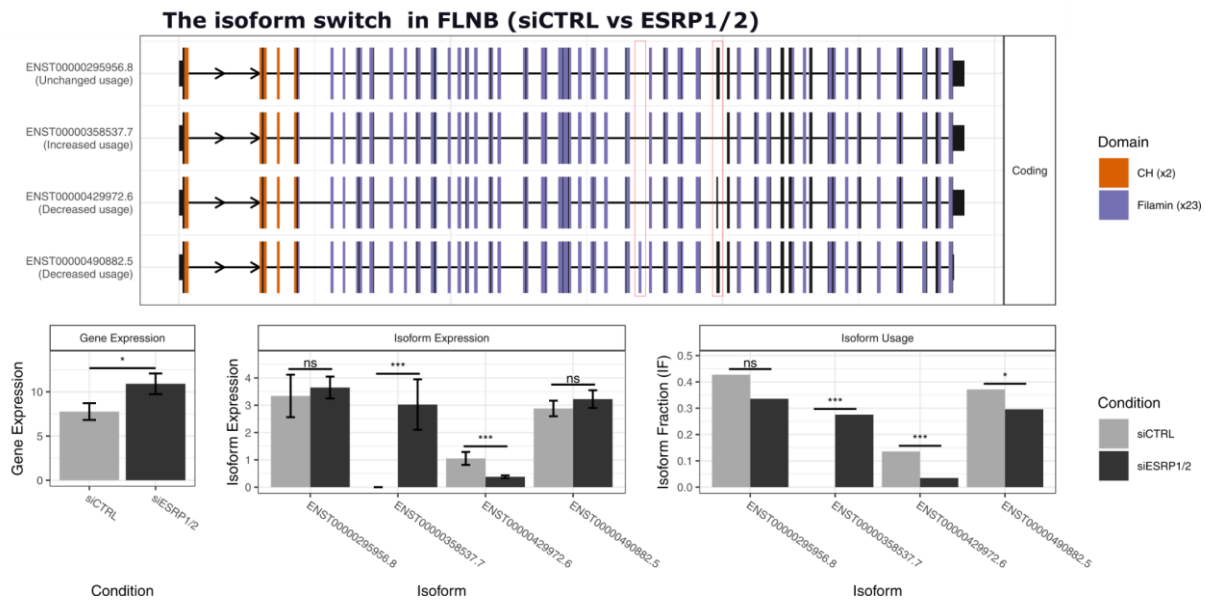




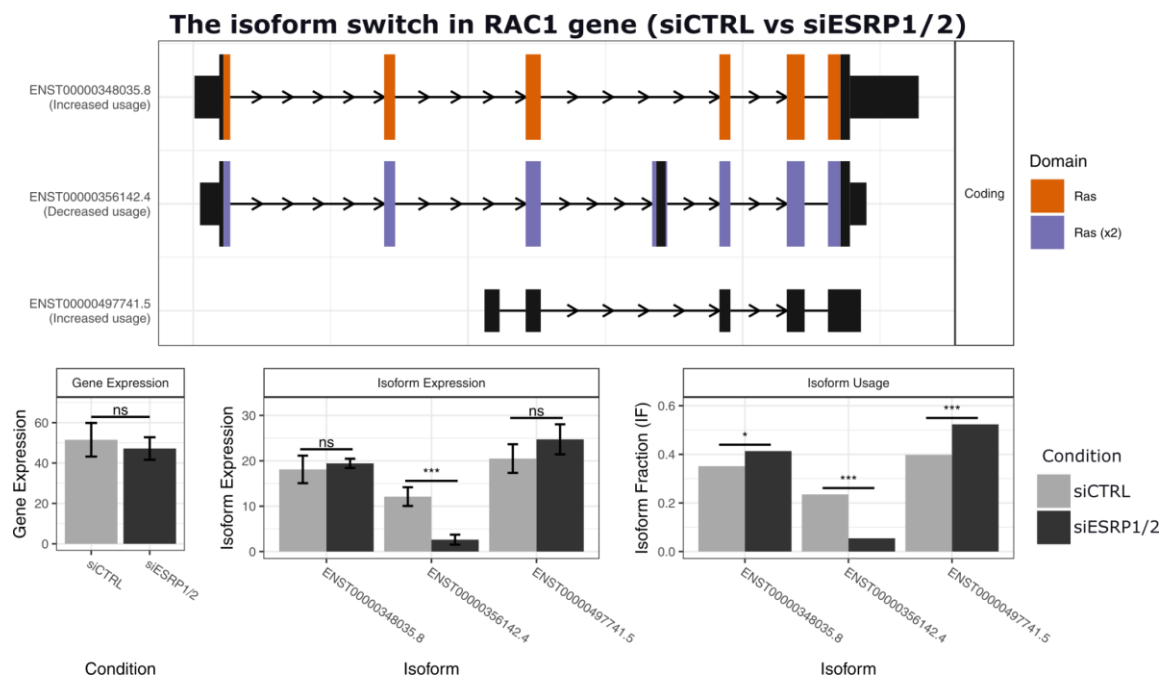
*Supplementary figure 8: Western Blot experiments showing the efficacy of the siRNAs-mediated silencing of ESRP1 and ESRP2 genes, respectively, in the RNA-seq replicates. Each experiment is performed in triplicates. Total RNA was extracted from these samples and subjected to sequencing.*



**Supplementary figure 9:** Bar plots reporting gene sets hallmarks enriched for downregulated (a) and upregulated genes (b) upon the combined silencing of ESRP1 and ESRP2 genes. Estrogen related hallmarks are highlighted in bold. Bar length is proportional to the number of genes overlapping each term, and the color intensities are representing the significance ( $p$ -value) of the enrichment.

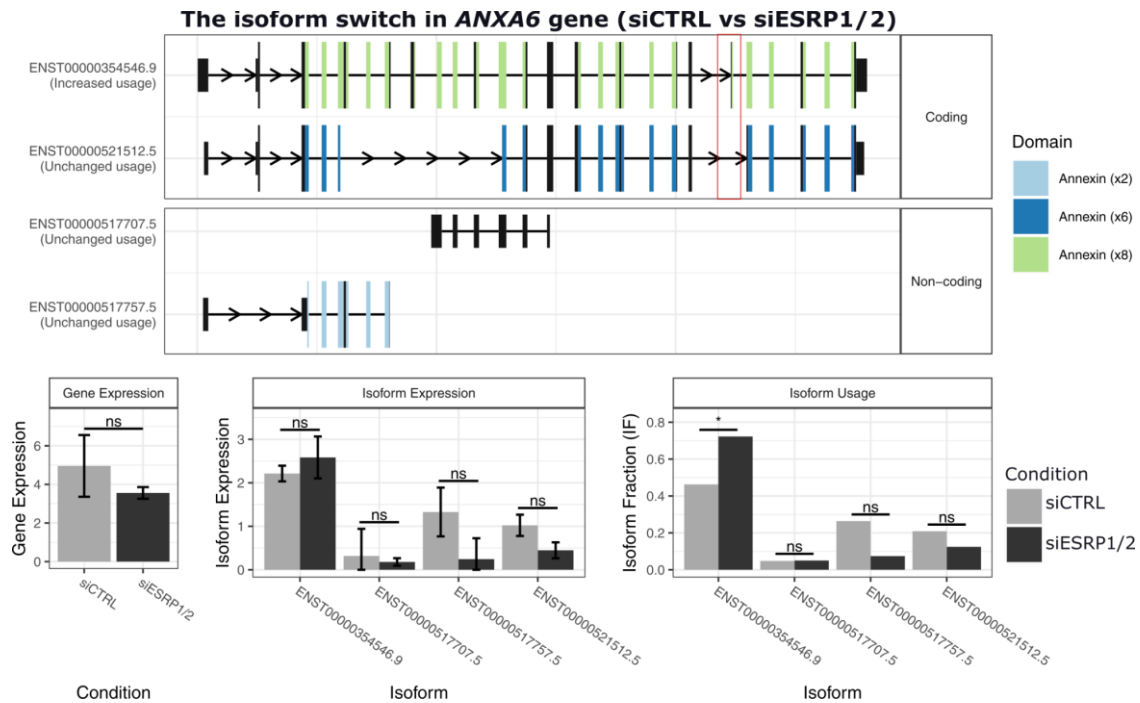


**Supplementary Figure 10:** Isoform switch plot showing isoform switching event in the relative abundance of FLNB gene isoform pairs and their downstream consequences induced by the combined silencing of ESRP1 and ESRP2 genes in MCF-7 BC cells. Switching isoform pairs (induced versus repressed isoforms) differ by the inclusion/exclusion of exons 26 and 30 (and an A5' splice site at exon 30) highlighted in the plot with red rectangles. Protein domains (CH, Filamin) encoded by each exonic region of the gene are reported. Histograms represent the expression (in normalized TPM units) of the gene and its isoform pairs, and their DE status (gene expression, isoform expression) and differential usage (isoform usage) status are indicated. ns, not significant; \*\*, adj-p < 0.001; \*\*\*, adj-p < 0.00001.

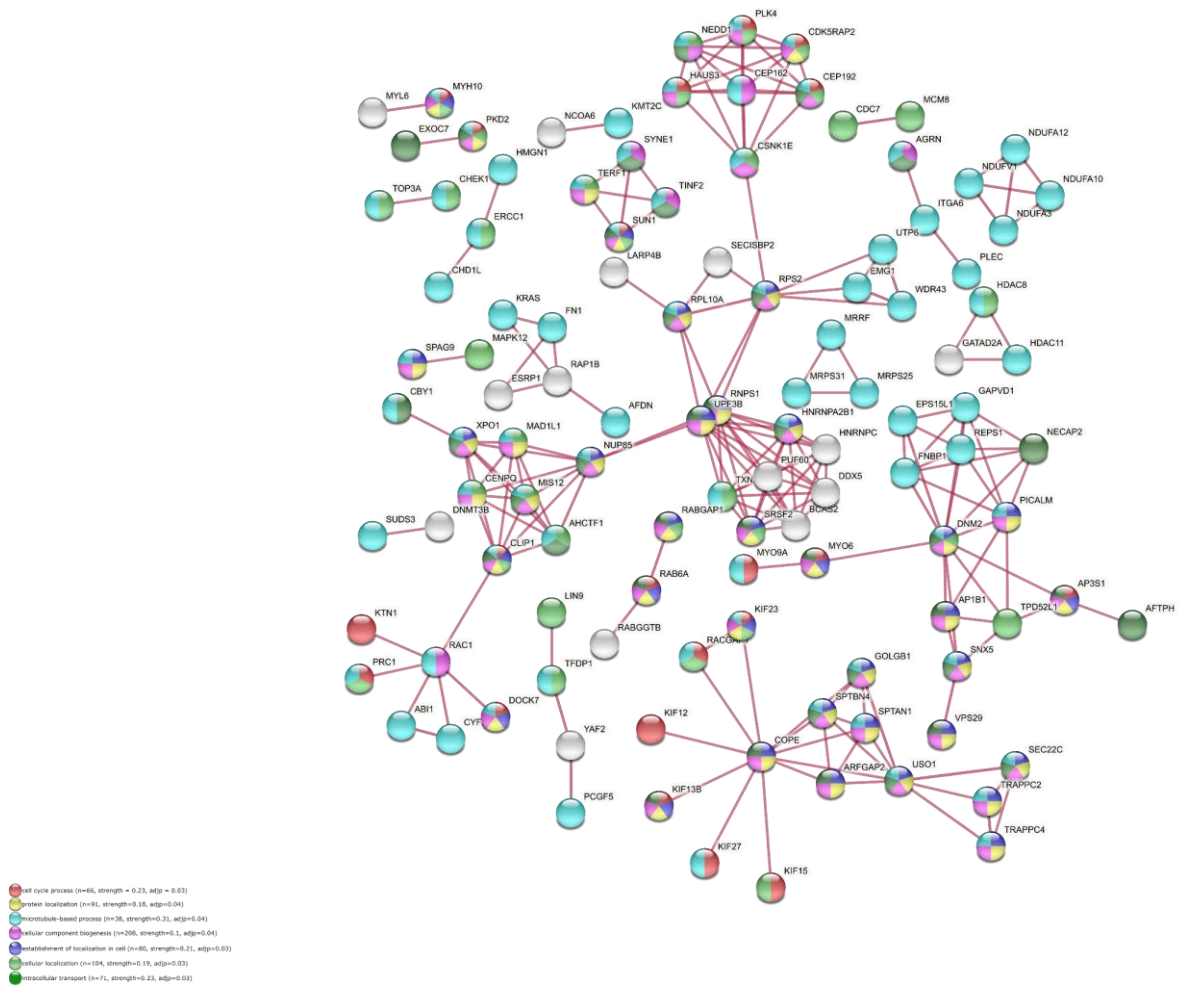


**Supplementary Figure 11:** Isoform switch plot showing isoform switching event in the relative abundance of RAC1 gene isoform pairs and their downstream consequences induced by the combined silencing of ESRP1 and ESRP2 genes

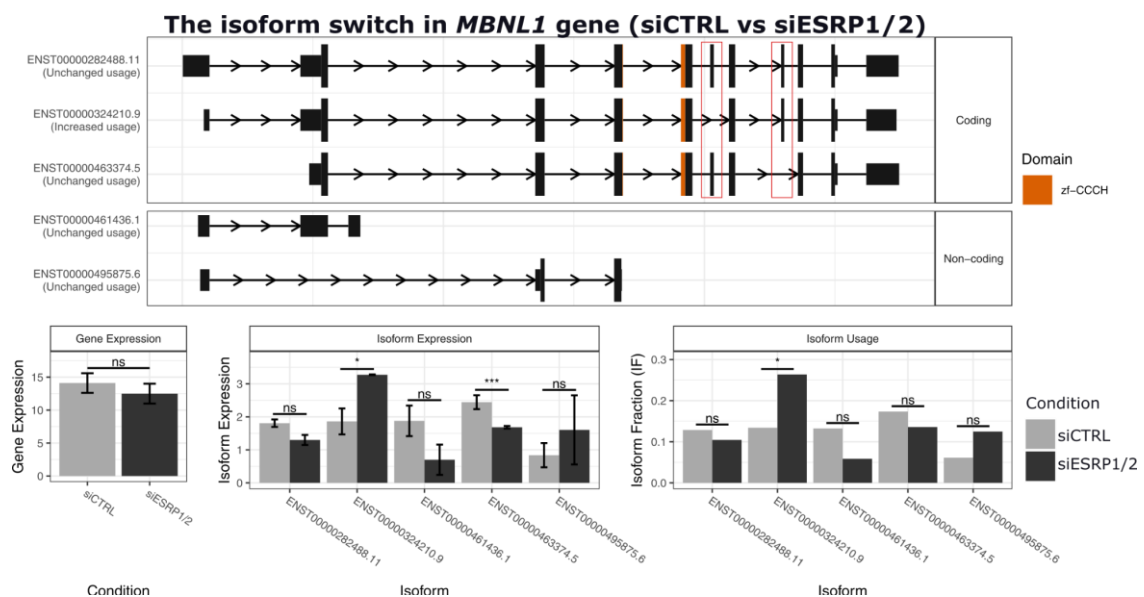
in MCF-7 BC cells. Switching isoform pairs (induced versus repressed isoforms) differ by the inclusion/exclusion of exon 4 (also known as 3b). Protein domains (Ras, Ras (x2)) encoded by each exonic region of the gene are reported. Histograms represent the expression (in normalized TPM units) of the gene and its isoform pairs, and their DE status (gene expression, isoform expression) and isoform differential usage (isoform usage) are indicated. ns, not significant; \*\*, adj-p < 0.001; \*\*\*, adj-p < 0.00001.



**Supplementary Figure 12:** Isoform switch plot showing ANXA6 gene switching isoforms pairs and their downstream consequences induced by the combined silencing of ESRP1 and ESRP2 genes in MCF-7 BC cells. Switching isoform pairs (induced versus repressed isoforms) differ by the inclusion/exclusion of exon 21 (hg38) highlighted with a red rectangle. Protein domains (Annexin) encoded by each exonic region of the gene are reported. Histograms represent the expression (in normalized TPM units) of the gene and its isoform pairs, and their DE status (gene expression, isoform expression) and isoform differential usage (isoform usage) are indicated. ns, not significant; \*\*, adj-p < 0.001; \*\*\*, adj-p < 0.00001.

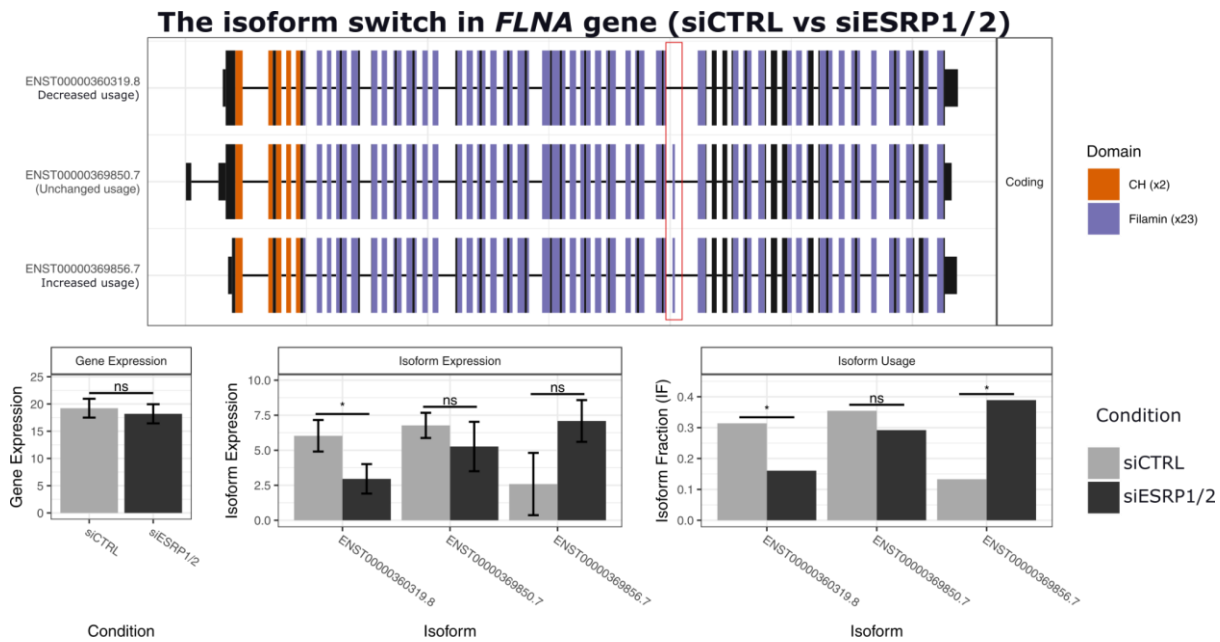


**Supplementary Figure 13:** Physical protein-protein interactions network enriched for genes harbouring an ES event upon the combined silencing of *ESRP1* and *ESRP2* genes. Nodes represent enriched proteins, where each color represents a specific enriched biological process. Connections/edges between the nodes are based on published experimental evidence.

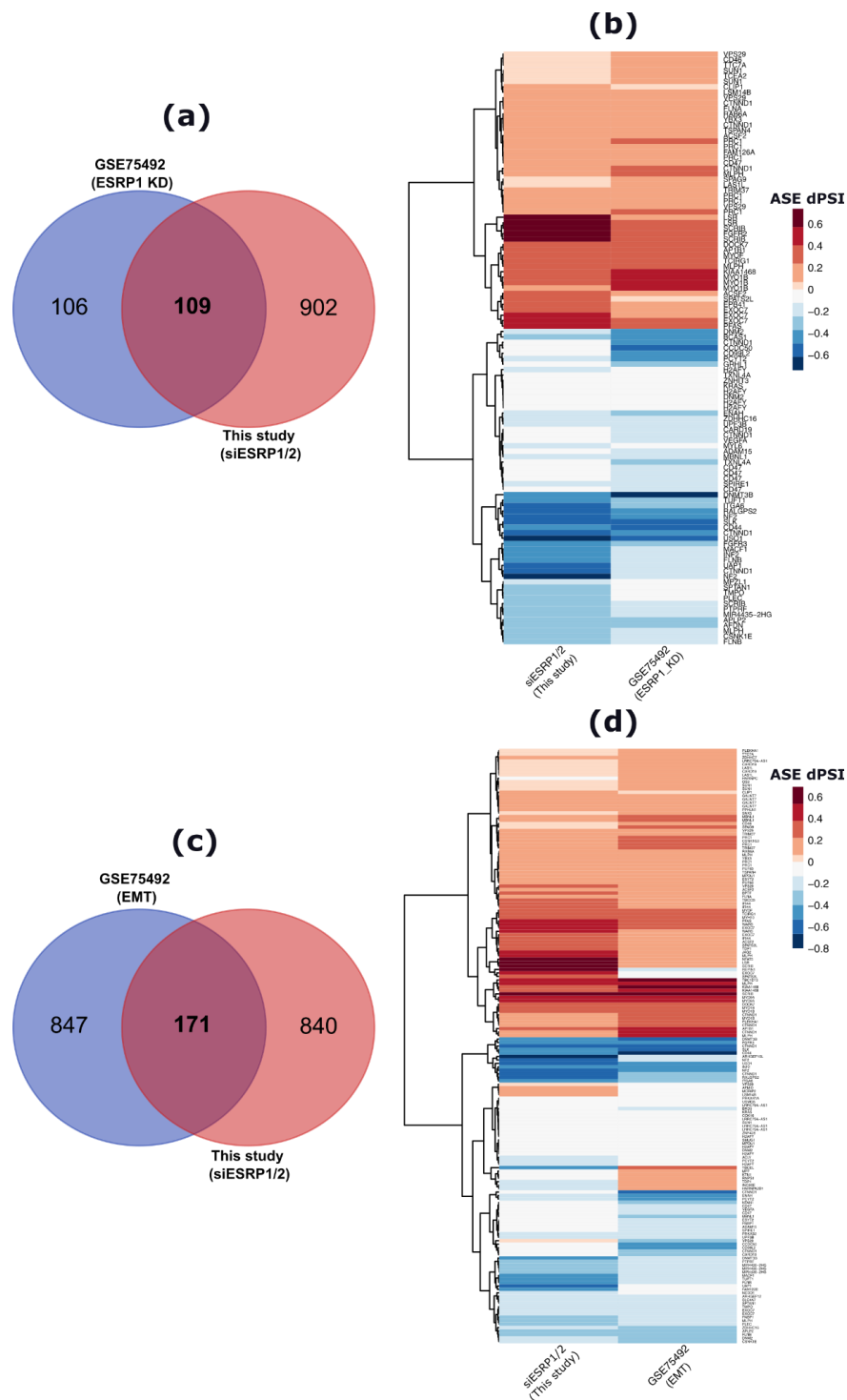


**Supplementary Figure 14:** Isoform switch plot showing *MBNL1* gene switching isoforms pairs and their downstream consequences induced by the combined silencing of *ESRP1* and *ESRP2* genes in MCF-7 BC cells. Switching isoform pairs (induced versus repressed isoforms) differ by the inclusion/exclusion of exons 6 and 8 (hg38) highlighted with red

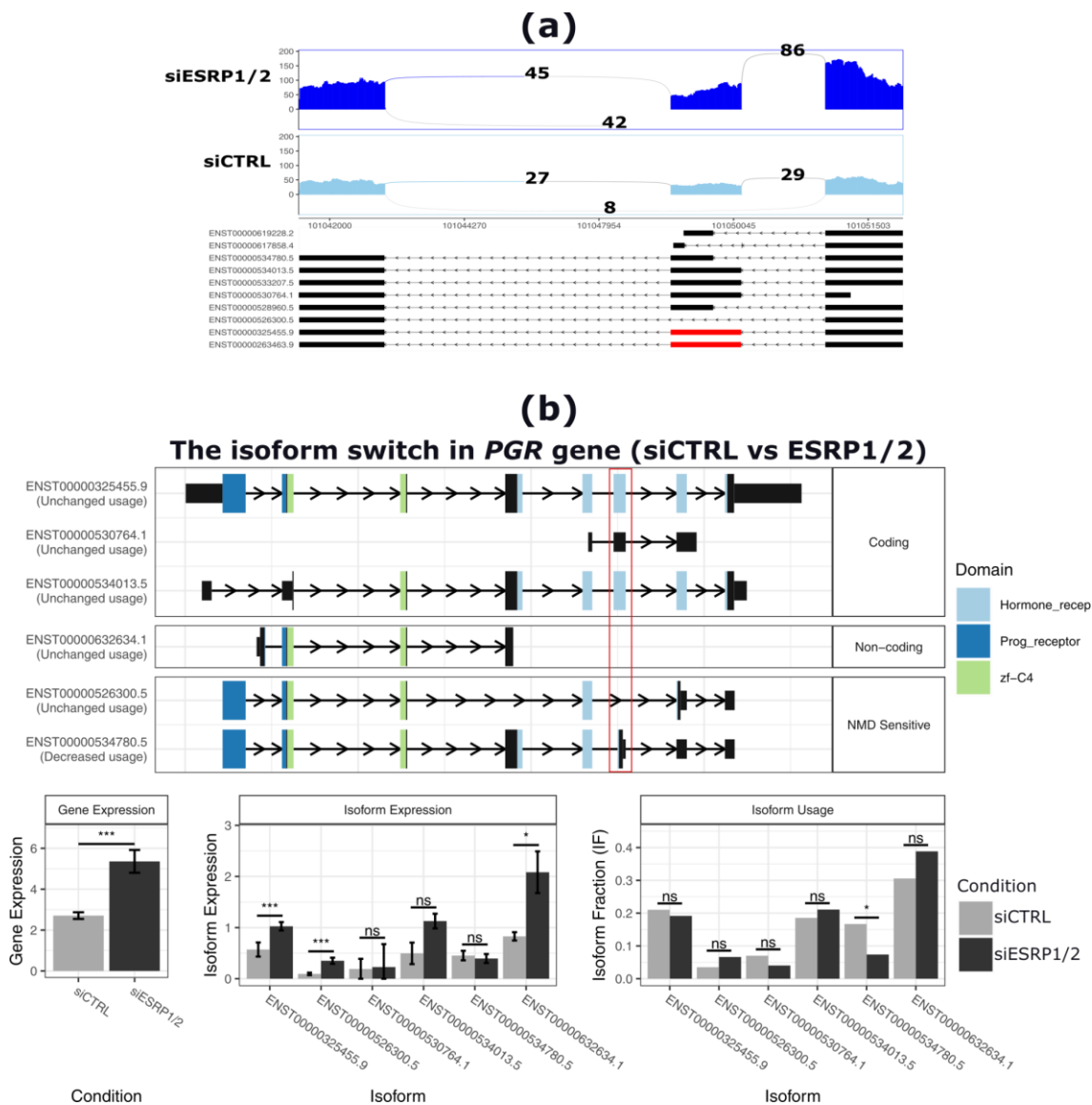
rectangles. Protein domains (zf-CCCH) encoded by each exonic region of the gene are reported. Histograms represent the expression (in normalized TPM units) of the gene and its isoform pairs, and their DE status (gene expression, isoform expression) and isoform differential usage (isoform usage) are indicated. ns, not significant; \*\*,  $adj-p < 0.001$ ; \*\*\*,  $adj-p < 0.00001$ .



**Supplementary Figure 15:** Isoform switch plot showing *FLNA* gene switching isoforms pairs induced by the combined silencing of *ESRP1* and *ESRP2* genes in MCF-7 BC cells. Switching isoform pairs (induced versus repressed isoforms) differ by the inclusion/exclusion of exon 32 (hg38) highlighted with red rectangles. Protein domains (CH, Filamin) encoded by each exonic region of the gene are reported. Histograms represent the expression (in normalized TPM units) of the gene and its isoform pairs, and their DE status (gene expression, isoform expression) and isoform differential usage (isoform usage) are indicated. ns, not significant; \*,  $adj-p < 0.05$ ; \*\*,  $adj-p < 0.001$ ; \*\*\*,  $adj-p < 0.00001$ .

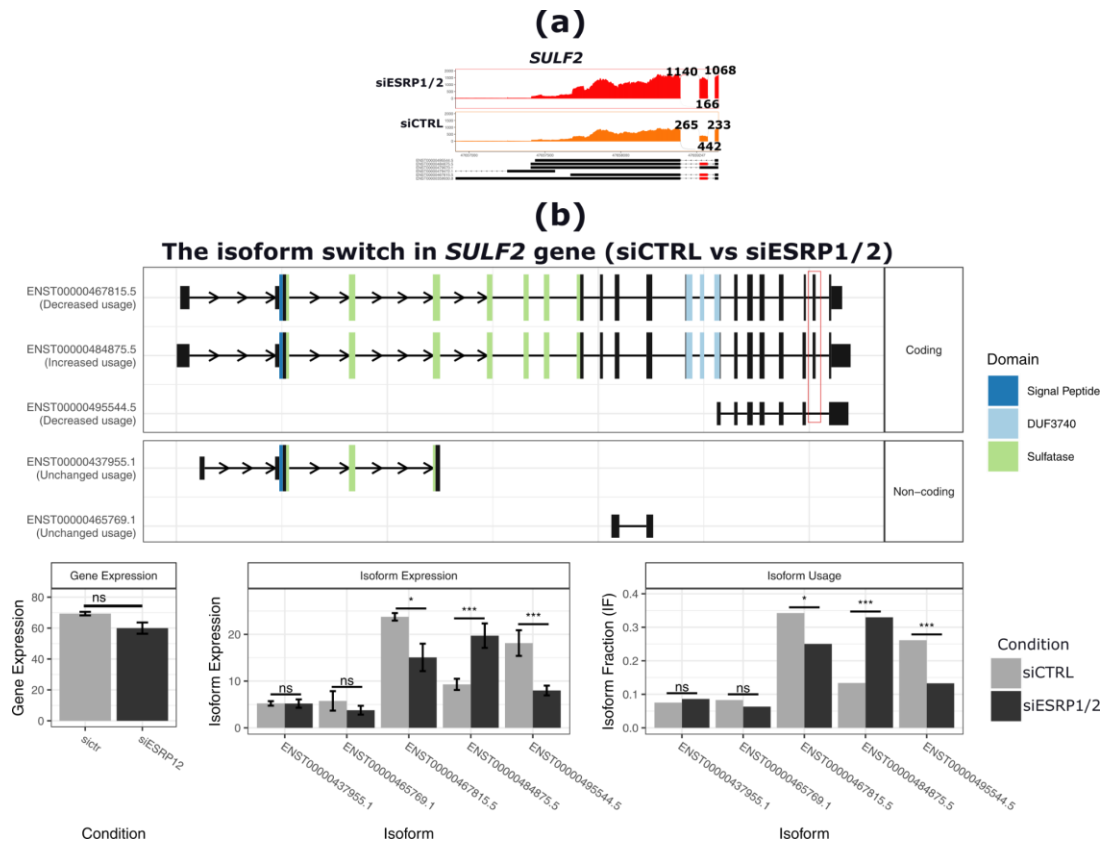


**Supplementary Figure 16:** Overview of the overlap in AS changes between this study and the (GSE75492) study. (a) Venn diagram showing the number of overlapping ASEs between this study (siESRP1/2) and the ESRP1 knock down (ESRP1 KD) experiment by (Yueqin Yang et al. 2016). (b) Heat map plot reporting the dPSI of the 109 overlapping ASEs shown in (a). (c) Venn diagram representing the number of overlapping ASEs between this study and the EMT induction experiment by (Yueqin Yang et al. 2016). (d) Heat map plot representing the dPSI of the 171 overlapping ASEs shown in (c).

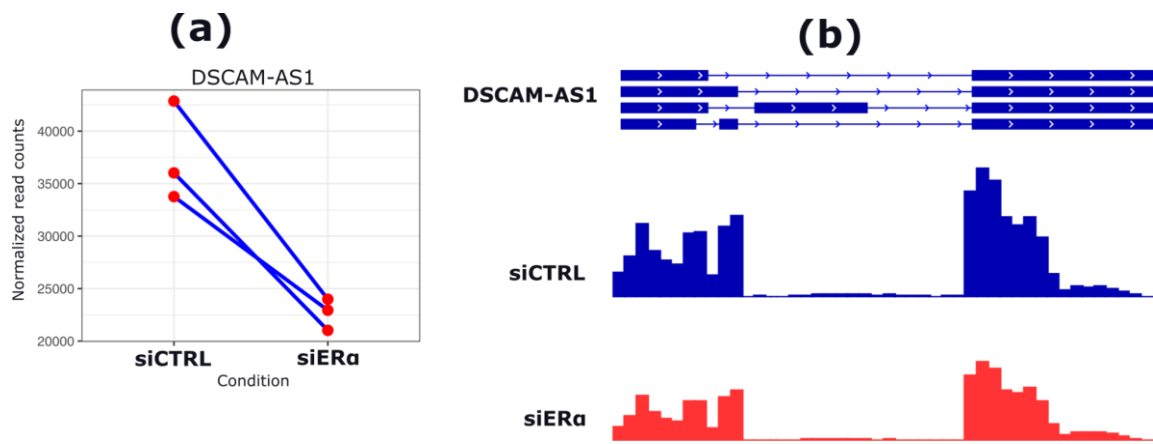


**Supplementary Figure 17:** Transcriptional and posttranscriptional regulation of *PGR* gene induced by *ESRP1/2* silencing in MCF-7 BC cells. **(a)** Sashimi plot reporting an exon inclusion event induced by *ESRP1/2* silencing. Numbers above junctions indicate the normalized read counts supporting either inclusion or exclusion of the event in siCTRL and siESRP1/2 conditions. **(b)** Isoform switching plot showing the differential regulation *PGR* isoforms by *ESRP1/2* silencing. The differentially spliced exon reported in (a) is highlighted with a red rectangle. ns, not significant; \*,  $adj-p < 0.05$ ; \*\*,  $adj-p < 0.001$ ; \*\*\*,  $adj-p < 0.00001$ .

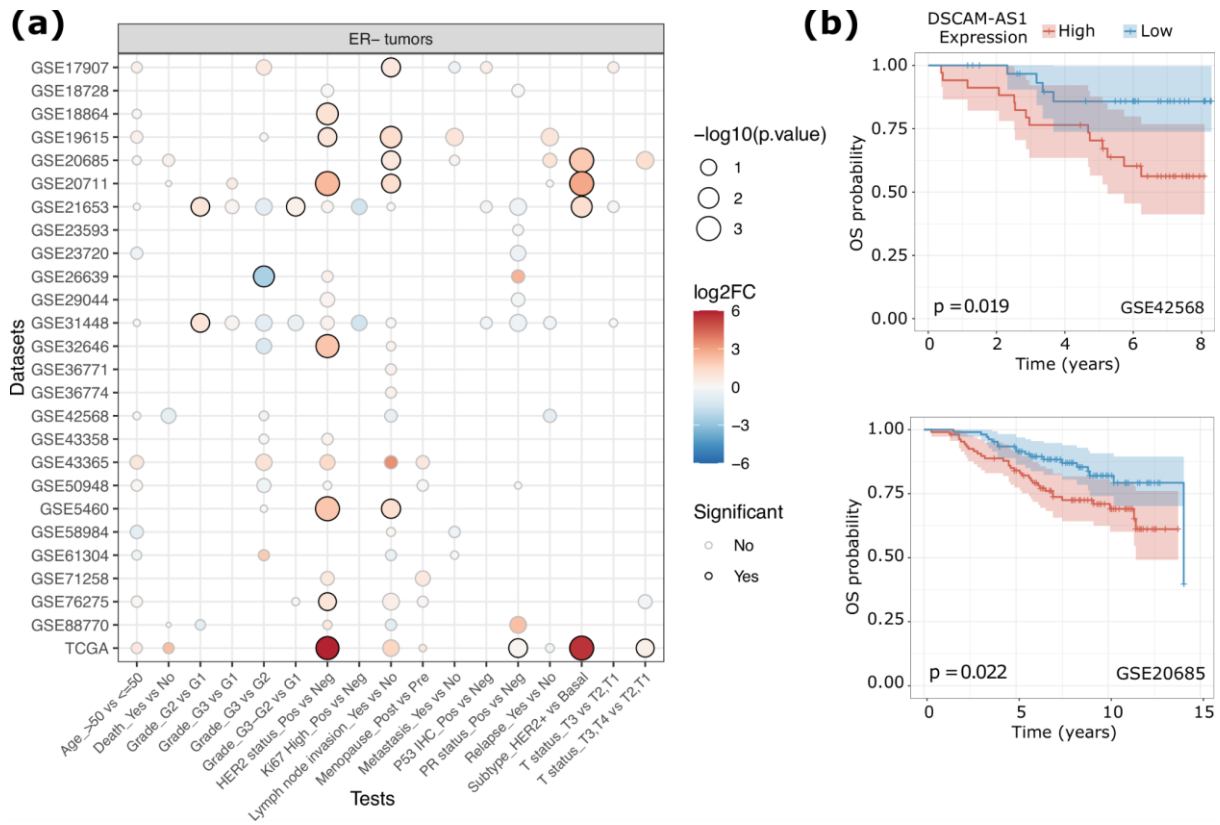




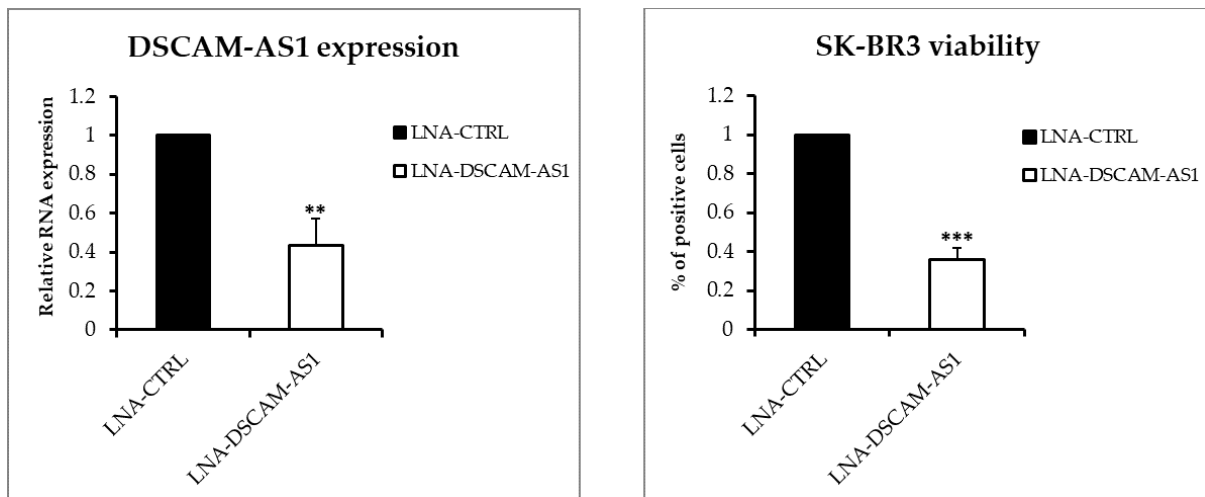
**(a)** Sashimi plot reporting an exon inclusion event induced by *ESRP1/2* silencing. Numbers above junctions indicate the normalized read counts supporting either inclusion or exclusion of the event in siCTRL and siESRP1/2 conditions. **(b)** Isoform switching plot showing the differential regulation *SULF2* isoforms by *ESRP1/2* silencing. The differentially spliced exon reported in (a) is highlighted with a red rectangle. ns, not significant; \*,  $adj-p < 0.05$ ; \*\*,  $adj-p < 0.001$ ; \*\*\*,  $adj-p < 0.00001$ .



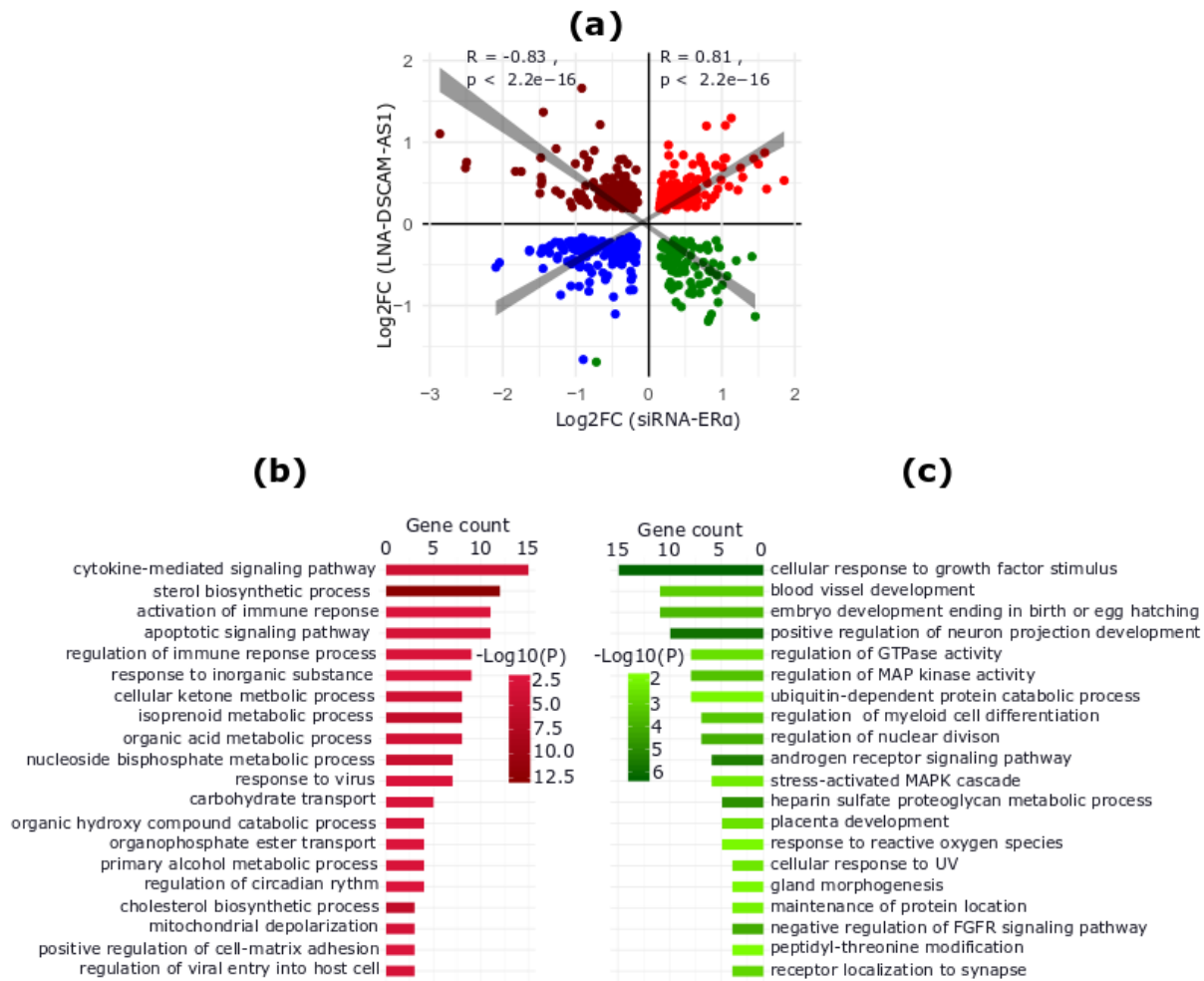
**Supplementary Figure 19** Overview of gene expression changes of DSCAM-AS1 gene upon ER $\alpha$  silencing in MCF-7 cells cultured under hormone-deprived conditions. **(a)** Line plot reporting the expression levels in normalized read counts of DSCAM-AS1 genes in the triplicates of siCTRL and siER $\alpha$  conditions, respectively. **(b)** Genome representation of read coverage over DSCAM-AS1 gene body from the interactive genome browser (IGV).



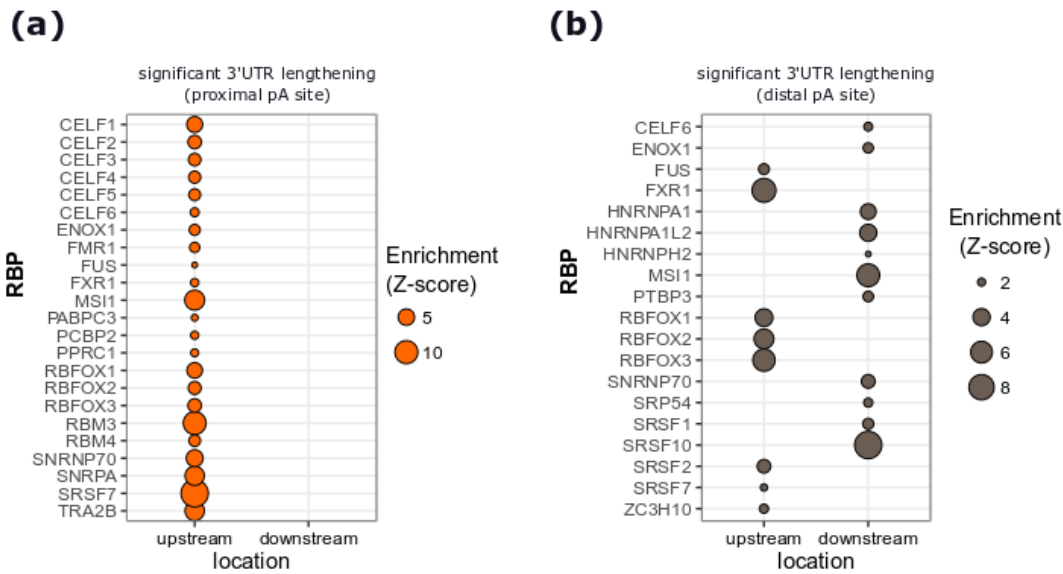
**Supplementary Figure 20:** (a) Dot plot reporting the level of statistical significance of the differential DSCAM-AS1 expression analyses between groups of ER-negative BC patients separated with respect to specific clinical data. The size of the dot is proportional to the significance of the results while the color code represents the log<sub>2</sub>FC of expression. ER, Estrogen Receptor; Pos, positive; Neg, negative; PR, Progesterone Receptor. (b) Kaplan-Meier curves representing the Overall Survival (OS) of BC patients based on the median level of DSCAM-AS1 expression. P-value by log-rank test.



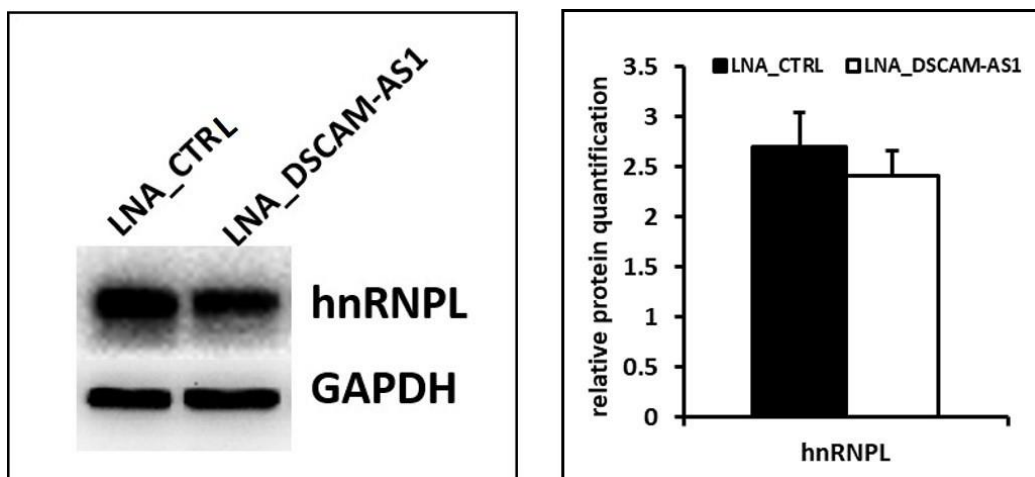
**Supplementary Figure 21:** (a) Expression levels of DSCAM-AS1 and (b) viability measured by Crystal Violet Assay in SK-BR-3 cells upon transection of DSCAM-AS1-targeting or control LNA. Error bars represent the standard deviation of three biological replicates. Significance from T-test: \*\*,  $p$ -value < 0.01; \*\*\*,  $p$ -value < 0.001.



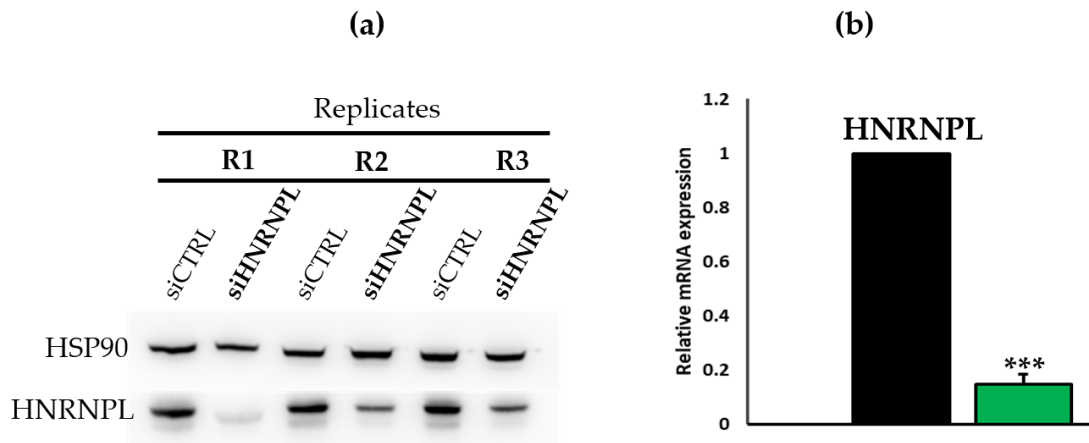
**Supplementary Figure 22:** (a) The scatter plot shows the  $\text{log}_2$  fold change of DE genes between this study and those obtained upon ERA silencing (siERA) (Miano et al. 2018b). Dark red and green dots represent genes upregulated in this study while downregulated in the siER experiment, and those downregulated in this study and upregulated in the siERA experiment, respectively. Blue and red dots represent those genes downregulated or upregulated in both studies, respectively. (b) GO terms enriched for genes upregulated in this study and downregulated in the siERA experiment. (c) GO terms enriched for genes downregulated in this study while upregulated in the siERA experiment.



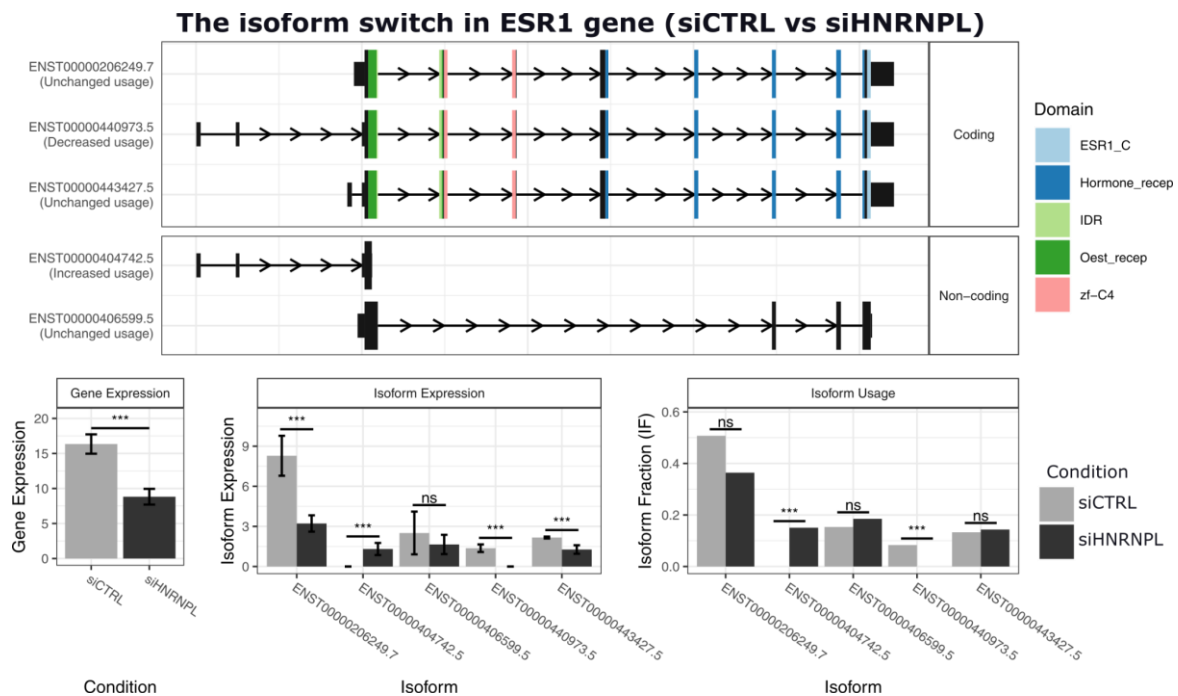
**Supplementary Figure 23:** The list of RBPs predicted to have an enrichment of their binding motifs in the 3'UTR lengthening events upon DSCAM-AS1 silencing. The enrichment is shown for a selected region upstream and downstream of proximal (a) and distal (b) APA sites, respectively. Significance: z-score > 1.96



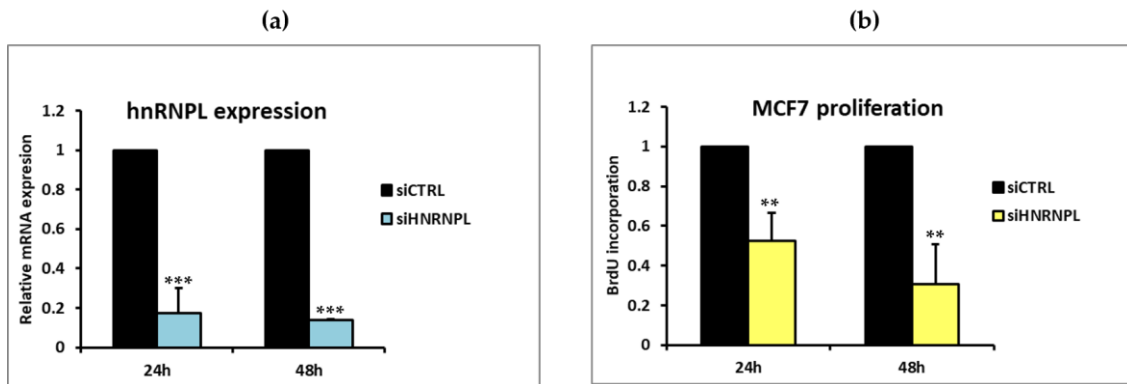
**Supplementary Figure 24. Left.** Representative image of Western blot analysis of hnRNPL protein, GAPDH: loading control. **Right.** The histogram shows the protein level of hnRNPL in MCF-7 cells transfected with DSCAM-AS1 or control LNA. hnRNPL values are relative to GAPDH. Error bars represent the standard error of three biological replicates.



**Supplementary Figure 25.** Left, Representative image of WB analysis of *hnRNPL* protein expression on MCF-7 transfected with specific siRNA, HSP90: loading control. Right, The histogram shows the mRNA expression level of HNRNPL in MCF-7 cells transfected with HNRNPL or control siRNA. *hnRNPL* values are relative to HSP90. Error bars represent the standard error of the three biological replicates.



**Supplementary Figure 26:** Isoform switching plot showing the transcriptional and posttranscriptional regulation of *ESR1* gene induced by HNRNPL silencing in MCF-7 BC cells. *ns*, not significant; \*, *adj-p* < 0.05; \*\*, *adj-p* < 0.001; \*\*\*, *adj-p* < 0.00001.










**Supplementary Figure 27:** silencing HNRNPL reduces proliferative potential of MCF-7 cells. (a) Expression levels of HNRNPL 24h and 48h upon transfection of MCF-7 cells with control or HNRNPL-targeting siRNAs, respectively. (b) MCF-7 proliferation changes measured by BrdU incorporation assay measured by cells 24h and 48h upon transfection of cells with HNRNPL-targeting or control siRNAs, respectively. Error bars represent the standard deviation of three biological replicates. Significance from T-test: \*\*,  $p$ -value < 0.01; \*\*\*,  $p$ -value < 0.001.



Article

# DSCAM-AS1-Driven Proliferation of Breast Cancer Cells Involves Regulation of Alternative Exon Splicing and 3'-End Usage

Jamal Elhasnaoui <sup>1,2,†</sup> , Valentina Miano <sup>1,2,3,†</sup> , Giulio Ferrero <sup>1,2,4</sup> , Elena Doria <sup>2</sup>, Antonette E. Leon <sup>5</sup>, Aline S. C. Fabricio <sup>5</sup> , Laura Annaratone <sup>6</sup> , Isabella Castellano <sup>6</sup>, Anna Sapino <sup>6,7</sup>  and Michele De Bortoli <sup>1,2,\*</sup> 

- <sup>1</sup> Center for Molecular Systems Biology, University of Turin, Orbassano, 10043 Turin, Italy; jamal.elhasnaoui@unito.it (J.E.); valentina.miano@unito.it (V.M.); giulio.ferrero@unito.it (G.F.)
- <sup>2</sup> Department of Clinical and Biological Sciences, University of Turin, Orbassano, 10043 Turin, Italy; elena.doria@edu.unito.it
- <sup>3</sup> Division of Cellular and Molecular Pathology, Department of Pathology, University of Cambridge, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK
- <sup>4</sup> Department of Computer Science, University of Turin, 10149 Turin, Italy
- <sup>5</sup> Regional Center for Biomarkers, Department of Clinical Pathology, Azienda ULSS 3 Serenissima, Campo SS Giovanni e Paolo 6777, 30122 Venice, Italy; antonette.leon@aulss3.veneto.it (A.E.L.); aline.fabricio@aulss3.veneto.it (A.S.C.F.)
- <sup>6</sup> Department of Medical Sciences, University of Turin, 10126 Turin, Italy; laura.annaratone@unito.it (L.A.); isabella.castellano@unito.it (I.C.); anna.sapino@unito.it (A.S.)
- <sup>7</sup> Candiolo Cancer Institute, FPO-IRCCS, 10060 Candiolo, Turin, Italy
- \* Correspondence: michele.debortoli@unito.it; Tel.: +39-0116-7050-58
- † These authors contributed equally to this work.

Received: 7 April 2020; Accepted: 31 May 2020; Published: 3 June 2020



**Abstract:** *DSCAM-AS1* is a cancer-related long noncoding RNA with higher expression levels in Luminal A, B, and HER2-positive Breast Carcinoma (BC), where its expression is strongly dependent on Estrogen Receptor Alpha (ER $\alpha$ ). *DSCAM-AS1* expression is analyzed in 30 public datasets and, additionally, by qRT-PCR in tumors from 93 BC patients, to uncover correlations with clinical data. Moreover, the effect of *DSCAM-AS1* knockdown on gene expression and alternative splicing is studied by RNA-Seq in MCF-7 cells. We confirm *DSCAM-AS1* overexpression in high grade Luminal A, B, and HER2+ BCs and find a significant correlation with disease relapse. In total, 908 genes are regulated by *DSCAM-AS1*-silencing, primarily involved in the cell cycle and inflammatory response. Noteworthy, the analysis of alternative splicing and isoform regulation reveals 2085 splicing events regulated by *DSCAM-AS1*, enriched in alternative polyadenylation sites, 3'UTR (untranslated region) shortening and exon skipping events. Finally, the *DSCAM-AS1*-interacting splicing factor heterogeneous nuclear ribonucleoprotein L (hnRNPL) is predicted as the most enriched RBP for exon skipping and 3'UTR events. The relevance of *DSCAM-AS1* overexpression in BC is confirmed by clinical data and further enhanced by its possible involvement in the regulation of RNA processing, which is emerging as one of the most important dysfunctions in cancer.

**Keywords:** lncRNA; breast cancer; alternative splicing; estrogen receptor; RNA-Seq

## 1. Introduction

Non-coding RNAs are an established layer of regulation in the molecular pathophysiology of complex diseases, including cancer [1]. A large amount of evidence on the oncogenic or

oncosuppressor activities of non-coding RNA transcripts longer than 200 bp, defined as long noncoding RNAs (lncRNAs), has been accumulated in recent years thanks to the diffusion of deep-sequencing technologies [2,3]. In the context of Breast Carcinoma (BC), lncRNA expression and molecular activity were related to different stages of the disease as well as to the different BC subtypes [4]. Among these subtypes, the Estrogen Receptor alpha (ER $\alpha$ )-positive BC (Luminal A and B subtypes) represents the most frequent breast neoplasm with over 270,000 estimated new cases in the US population for 2020 [5]. Despite these tumors being characterized by a less aggressive phenotype and better patient outcome, a growing number of cases showed drug resistance and disease relapse [6]. Among the lncRNA genes related to the onset and progression of luminal BC, *HOTAIR*, *MIAT*, and *DSCAM-AS1* were described in multiple studies [7].

*DSCAM-AS1* was originally described as a lncRNA overexpressed in invasive BCs compared to normal adjacent tissue [8]. A subsequent analysis of ER $\alpha$  regulated lncRNAs in the luminal BC model MCF-7 performed by our group evidenced *DSCAM-AS1* as the most significantly ER $\alpha$ -regulated lncRNA in these cells [9]. In the same study, the *DSCAM-AS1* gene was evidenced as overexpressed in ER $\alpha$ -positive tumors, particularly of the luminal B subtype. Furthermore, we showed that siRNA-mediated silencing of *DSCAM-AS1* induced a reduction in MCF-7 proliferation with an increase in cell death. These results were further confirmed by different groups [10–12]. Niknafs and colleagues reported the overexpression of the lncRNA in tamoxifen-resistant cells whose proliferation decreases upon *DSCAM-AS1* silencing [10]. Alongside these results, our group showed that, in BC cells grown in hormone-deprived medium, ER $\alpha$  binding occurs in a super-enhancer region upstream of the *DSCAM-AS1* locus, which promotes the lncRNA overexpression in these cells [13]. These results, in addition to the evidence of *DSCAM-AS1* overexpression in BC patients with poor outcomes [11], make the understanding of the functional role of this lncRNA relevant.

The first evidence of interaction between *DSCAM-AS1* and RNA-binding proteins (RBPs) was reported by Niknafs and colleagues, demonstrating that *DSCAM-AS1* physically interacts with heterogeneous nuclear ribonucleoprotein L (hnRNPL) [10]. However, the functional role of this interaction was not elucidated. hnRNPL is a well-known splicing factor belonging to the heterogeneous nuclear ribonucleoprotein protein family and it is involved in the regulation of alternative splicing by binding to C/A-rich elements particularly at gene intron and 3'UTR [14,15]. Furthermore, hnRNPL activity was related to the maintenance of mRNA stability by the regulation of the nonsense-mediated mRNA decay (NMD) pathway through the binding at gene 3'UTR [16]. This activity was particularly relevant in cancer biology since the mRNA stability of well-known oncogenes like B-cell 2 (Bcl2) and Serine/arginine-rich splicing factor 3 (SRSF3), as well as the tumor suppressor protein 53 (p53), were demonstrated to be regulated by hnRNPL [16–18].

From a computational point of view, the recent advances in bioinformatics tools now allow the accurate characterization and quantification of gene isoforms from RNA sequencing data. These mRNA molecules, originating from the same locus, have different exon compositions and lengths and may encode for different corresponding proteins [19]. These isoforms may result from the differential usage of alternative transcription start sites (aTSSs), termination sites (aTTSs), or Alternative PolyA (APA) sites, or may be the consequence of the alternative splicing (AS) of internal exons [20]. AS is a regulatory mechanism which allows the fine-tuning of gene isoform expression in different cell types and tissues or under specific conditions such as cancer [21]. AS may affect mRNA localization, stability, and may change the open reading frame, resulting in protein isoforms with diverse functions or localization [22]. Accumulating evidence has indeed shown that different isoforms can be differentially used under distinct conditions and that this may have a substantial biological impact due to differences in their functional potentials [23,24]. These mechanisms of differential usage of isoforms, usually referred to as isoform switching, has been shown to be implicated in many diseases and is especially prominent in cancer, where it affects the expression of isoforms of genes involved in almost all cancer hallmarks [25,26].

In this study, we performed an exploratory analysis of *DSCAM-AS1* gene expression in multiple BC tumors, confirming the relation between the lncRNA expression and the poor survival rate of ER $\alpha$ -positive BC patients. Then, to functionally investigate the activity of this lncRNA, we analyzed the effects of *DSCAM-AS1* downregulation by RNA-seq followed by gene- and isoform-level analyses. The knockdown of *DSCAM-AS1* strongly hampers cell growth and proliferation pathways in MCF-7 cells. Interestingly, *DSCAM-AS1* knockdown not only affects gene expression but also drives remarkable changes in isoform expression and alternative splicing through its physical interaction with the splicing factor hnRNPL. Altogether, these data highlight the functional role of *DSCAM-AS1* in MCF-7 cells and shed light on the importance and diverse roles of this lncRNA in regulating isoform expression.

## 2. Results

### 2.1. *DSCAM-AS1* is Overexpressed in More Aggressive ER $\alpha$ -Positive BCs

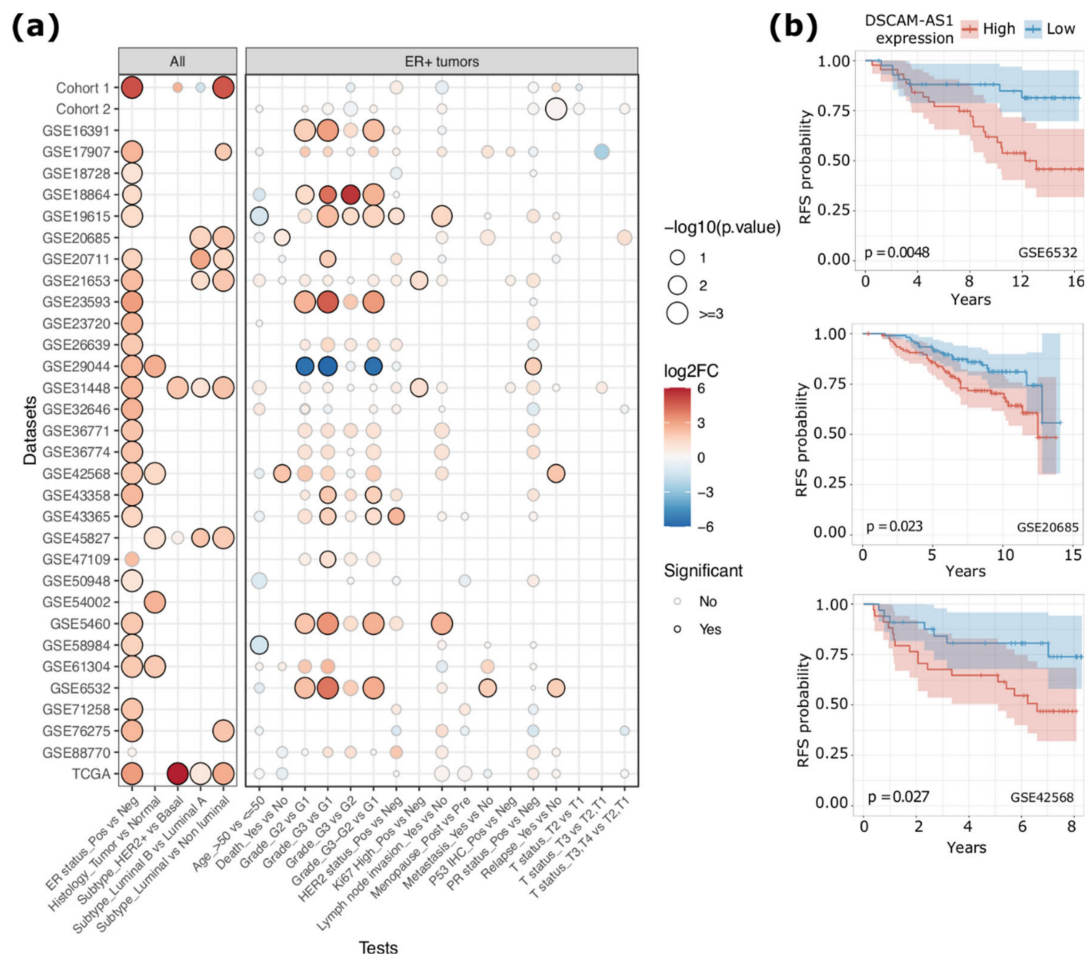
To investigate the expression of *DSCAM-AS1* in tumor tissues, we analyzed gene expression data from 30 public microarray datasets (Table S1a) and RNA-Seq data from The Cancer Genome Atlas (TCGA). In addition to these microarray and RNA-seq public data, we quantified, by qRT-PCR, the *DSCAM-AS1* expression in RNA samples derived from primary cancer tissues from two BC cohorts composed, respectively, of 42 (Cohort\_1) and 51 (Cohort\_2) subjects (details are reported in Materials and Methods and Table S1b,c).

Initially, we evaluated the differential *DSCAM-AS1* expression level between ER+ and ER- tumors observing, as expected, a consistent overexpression of this lncRNA in ER+ tumors (average log<sub>2</sub>FC ER+ vs. ER- = 2.07) (Figure 1a and Table S1d). This difference was statistically significant in 23 out of the 30 microarray datasets analyzed ( $p$ -value < 0.05), in the TCGA RNA-Seq dataset ( $p$ -value <  $2.2 \times 10^{-16}$ ) and in Cohort\_1 ( $p$ -value =  $8.01 \times 10^{-10}$ ). In the TCGA dataset, the gene was detectable (fragment per kilobase per million mapped reads (FPKM) > 1) in 381 out of 803 ER+ tumors while it was detectable in 29 out of 237 ER- tumors. Noteworthy, 93% of *DSCAM-AS1* positive BCs are ER+. In coherence with this result, *DSCAM-AS1* was observed as recurrently upregulated in luminal BCs compared to non-luminal BCs (average log<sub>2</sub>FC = 2.37), with higher expression in luminal B tumors compared to luminal A BCs (average log<sub>2</sub>FC = 1.20) as previously reported [9,12]. Furthermore, in non-luminal BCs, *DSCAM-AS1* was overexpressed in HER2+ BCs (Figure 1a and Figure S1a).

The association between *DSCAM-AS1* expression levels and different clinical data was instead evaluated separately for ER+ and ER- tumor groups. As reported in Figure 1a, in ER+ BCs, *DSCAM-AS1* emerged as recurrently overexpressed in high-grade BCs (G2 or G3, 10 significant datasets). In a subset of datasets, *DSCAM-AS1* was significantly overexpressed in BC with lymph node invasion (two datasets), high Ki67 levels (two datasets), HER2 overexpression (two datasets) and when diagnosed in young patients (two datasets). Finally, in four datasets, *DSCAM-AS1* was overexpressed in BCs of patients associated with a higher death rate, while in three datasets, *DSCAM-AS1* was overexpressed in BCs characterized by a higher relapse rate.

To further explore the relation between *DSCAM-AS1* expression and the patient survival rate, we performed a survival analysis using Overall Survival (OS) or Recurrence-Free Survival (RFS) data provided in nine microarray datasets and in the BRCA TCGA cohort (Table S1e). As reported in Figure 1b, the *DSCAM-AS1* overexpression was significantly associated with a higher relapse rate in GSE6532 (HR = 3.21,  $p$ -value = 0.0048), GSE20685 (HR = 1.9,  $p$ -value = 0.023) and GSE42568 (HR = 2.6,  $p$ -value = 0.027) datasets. In the remaining microarray datasets, patients with higher *DSCAM-AS1* expression showed a non-significant lower relapse free survival rate (Log-rank-test  $p$ -value range from 0.34 to 0.93), even though, in three of them (GSE88770, GSE58984, GSE19615), the separation of RFS curves was concordant (Table S1e and Figure S1b). A nonsignificant difference ( $p$ -value = 0.99) was also observed in the analysis of TCGA data, as previously reported [10]. A significant difference in relapse rate was also observed among subjects from Cohort\_2 ( $p$ -value = 0.031). With respect to the overall survival rate, *DSCAM-AS1* overexpression was significantly associated with a lower survival

rate in two datasets, GSE42568 (HR = 3.47,  $p$ -value = 0.019) and GSE20685 (HR = 1.96,  $p$ -value = 0.022) (Figure S1c).



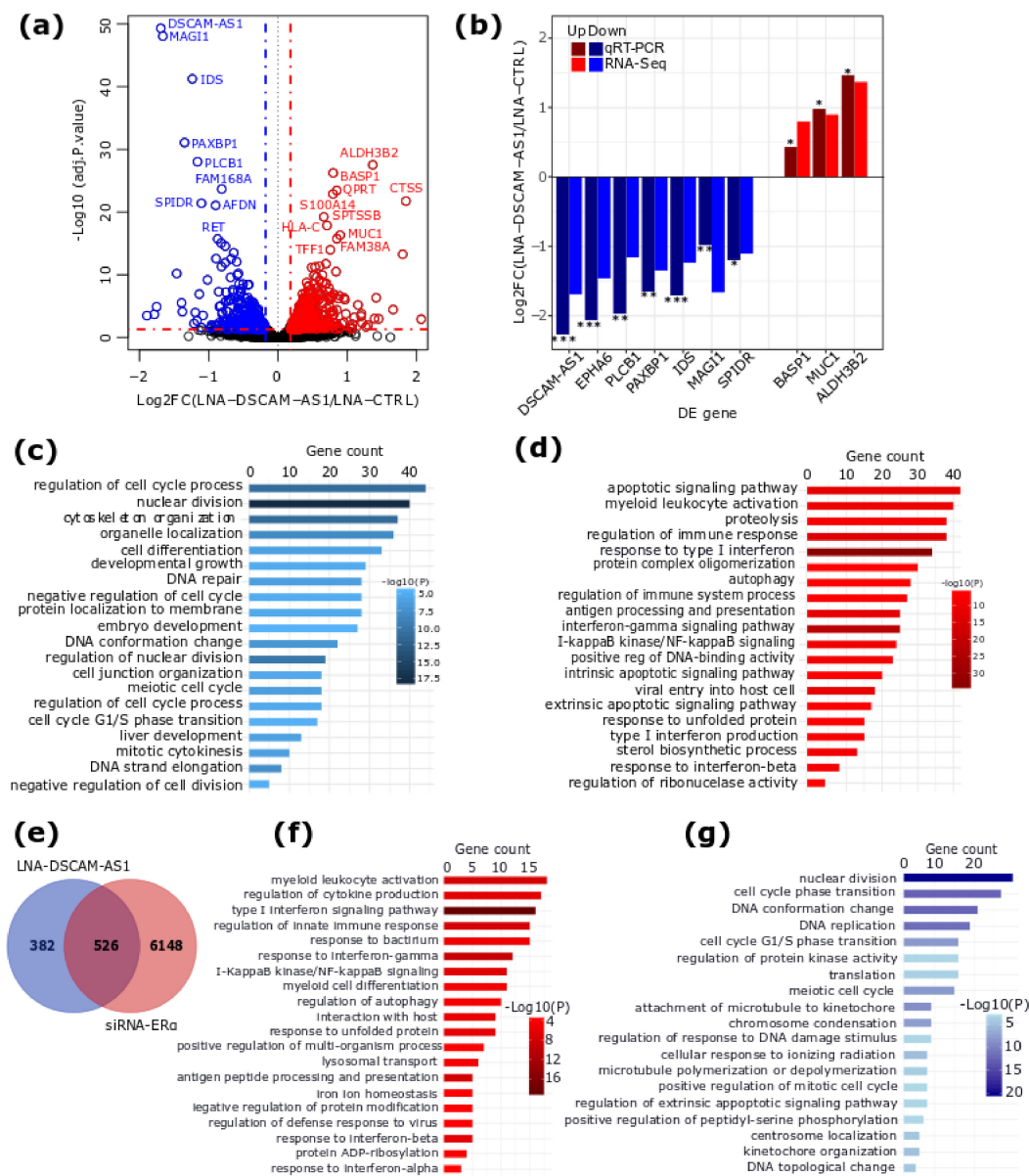
**Figure 1.** (a) Dot plot reporting the level of statistical significance of the differential *DSCAM-AS1* expression analyses between groups of Breast Carcinoma (BC) patients classified based on specific clinical data. The size of the dot is proportional to the significance of the results while the color code represents the  $\log_2FC$  of *DSCAM-AS1* expression. The left panel reports the results obtained considering all the samples, while the right panel reports the results of tests performed considering only the ER+ tumors. Estrogen Receptor (ER); positive (Pos); negative (Neg); Progesterone Receptor (PR). (b) Kaplan-Meier curves representing the Relapse Free Survival (RFS) of BC patients grouped by the median level of *DSCAM-AS1* expression.  $p$ -value by log-rank test.

## 2.2. *DSCAM-AS1* Knockdown Induces a Downregulation of Cell Cycle-Related Genes in BC Cells

To investigate the functional role of *DSCAM-AS1* in ER+ BC, we performed an RNA-Seq of MCF-7 cells transfected with control or *DSCAM-AS1*-targeting LNA (locked nucleic acids) GapmeRs. The analysis of the RNA-Seq data evidenced 908 genes differentially expressed (DE,  $|\log_2FC| > 0.20$ , adjusted  $p$ -value  $< 0.05$ ) upon *DSCAM-AS1* silencing (Figure 2a and Table S2). Specifically, 420 genes showed a significant decrease in expression, while 488 genes were upregulated. Notably, the differential expression of highly significant DE genes from the RNA-Seq analysis was confirmed by qRT-PCR analysis (Figure 2b). Furthermore, we observed, in our RNA-seq dataset, the same regulation trend for four out of six genes (*BCL2*, *ESR1*, *CDC6*, *E2F7*, *FEN1*, *TOP2A*) that were previously reported to be downregulated in MCF-7 by *DSCAM-AS1* silencing using siRNA [11]. Noteworthy, *DSCAM-AS1*-silenced cells showed a lower proliferation rate compared to the control (data not shown), even though *ESR1* expression was not perturbed (Table S2), confirming previous results [9,10]. Since the

overexpression of *DSCAM-AS1* was observed not only in ER $\alpha$ + but also in HER2+ breast tumors, we investigated the effects of *DSCAM-AS1* knockdown in the SK-BR-3 cell line, a commonly used HER2+ BC cell model. We observed that *DSCAM-AS1*-silencing significantly reduces the proliferation rate of SK-BR-3 cells (Figure S2), confirming that *DSCAM-AS1* is crucial for the survival of different BC cells.

A functional enrichment analysis of the DE genes showed distinct biological processes related to downregulated and upregulated genes. Specifically, downregulated genes were mainly involved in cell cycle progression, cell growth and proliferation (Figure 2c), while upregulated genes were mainly enriched in terms related to the inflammatory response, evoking type I and type II interferon signaling pathways, autophagy and the ER-phagosome pathway, apoptosis, and the regulation of cholesterol biosynthesis (Figure 2d and Table S3).



**Figure 2.** (a) Volcano plot showing the log<sub>2</sub>FC of gene expression and the statistical significance of the differential expression (DE) analysis performed between MCF-7 cells transfected with control or *DSCAM-AS1*-targeting LNA GapmeRs. In red are the upregulated genes, while in blue are the downregulated ones. The top first 20 significant DE genes are labeled. (b) Bar plot showing the expression log<sub>2</sub>FC of 10 DE genes whose expression was measured by qRT-PCR. Data from three biological replicates and *p*-value by T-test test: \*\*\*, *p*-value < 0.001; \*\*, *p*-value < 0.01; \*, *p*-value < 0.05.

(c,d) Bar plot reporting the top 20 significantly enriched biological processes related to downregulated and upregulated genes, respectively. The number of genes per GO (gene ontology) term is shown and the color code is proportional to significance. (e) Venn Diagram showing the overlap between genes DE in this study and those DE upon ER $\alpha$  silencing [13]. (f,g) Top 20 enriched GO terms of overlapping DE genes showing concordant regulation are shown for genes upregulated (f) and downregulated (g) in both datasets.

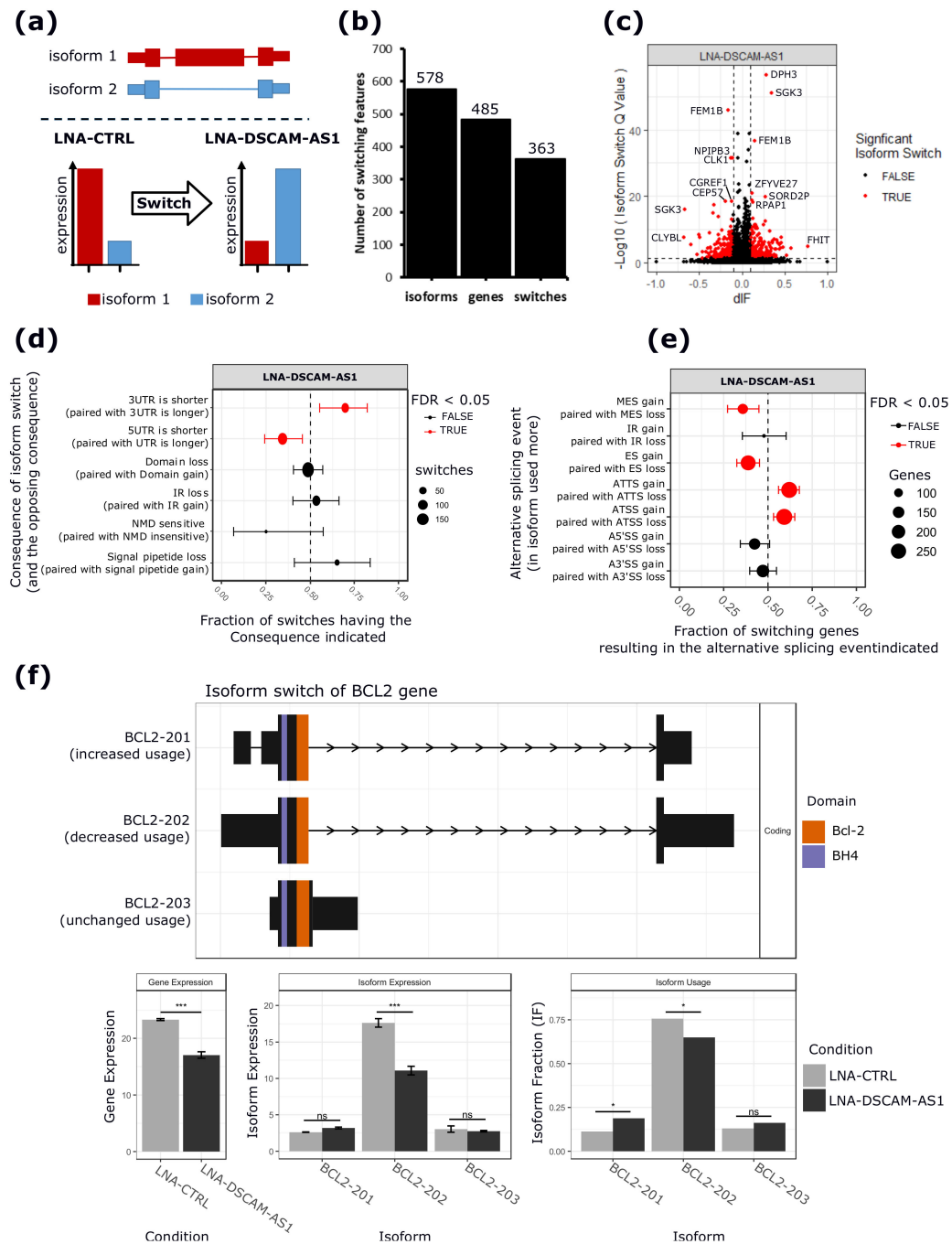
We previously demonstrated that *DSCAM-AS1* is an ER $\alpha$ -regulated lncRNA [9,13]. We thus compared the list of DE genes obtained in this study with those DE upon ER $\alpha$  silencing [13]. Notably, 526 genes were detected as DE in both datasets (hypergeometric test  $p < 4.36 \times 10^{-48}$ , representation factor = 1.7) (Figure 2e and Table S4). Of this, 291 genes (156 downregulated and 135 upregulated) showed coherent expression changes in the two datasets, while 235 genes showed an opposite regulation direction (Figure S3a). The GO enrichment analysis of the four categories of overlapping genes showed that they are related to distinct biological processes (Table S5). Notably, genes that showed an upregulated expression in both datasets are involved in interferon signaling pathways and the activation of the immune response, as well as autophagy regulation (Figure 2f), while genes that were downregulated in both datasets are involved in the regulation of the cell cycle process (Figure 2g). Genes upregulated only in this study but downregulated in the siRNA-ER $\alpha$  experiment are related to metabolic processes such as sterol and cholesterol biosynthesis (Figure S3b). Finally, genes downregulated in this study but upregulated in the siRNA-ER $\alpha$  experiment are related to developmental processes such as gland morphogenesis and neuron projection development pathway (Figure S3c).

When considering differential expression at the isoform level, the silencing of *DSCAM-AS1* perturbed the expression of 1035 isoforms, of which 439 were downregulated and 596 were upregulated (Figure S4a and Table S6). The DE isoforms were transcribed from 898 genes of which 381 (42%) genes were detected as DE in our gene-level analysis (Figure S4b). The GO enrichment analysis of the parent genes of these DE isoforms showed enrichment in terms similar to those obtained when considering the overall gene expression changes only (Figure S4c,d and Table S7).

### 2.3. *DSCAM-AS1* Physically Interacts with hnRNPL to Regulate AS in MCF-7 Cells

*DSCAM-AS1* was reported to interact with the RNA-Binding Protein (RBP) hnRNPL [10], which regulates AS and mRNA stability [15,16]. We also confirmed this interaction in our MCF-7 cellular model with an hnRNPL Cross-Linking Immunoprecipitation experiment (CLIP) (Figure S5). Therefore, in order to test whether *DSCAM-AS1* silencing displayed any effect on the AS events regulated by hnRNPL, we explored the possible regulation at the RNA isoform expression level in our dataset, as a result of either alternative exon splicing, or alternative transcription start site, or alternative poly(A) site. This was evaluated using two approaches: (i) an analysis of the isoform switching events, also called differential isoform usage, and (ii) an analysis of local AS changes occurring upon *DSCAM-AS1* silencing.

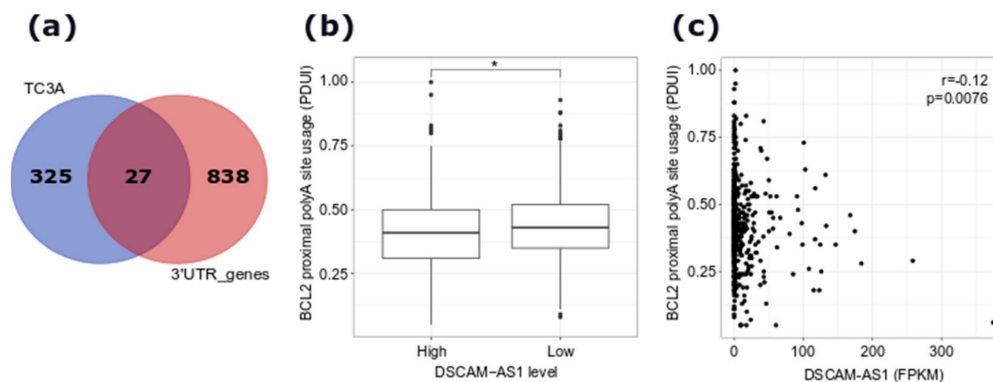
As shown in Figure 3a, using the IsoformSwitchAnalyseR tool [27], the relative contribution of the individual isoforms to the expression of the gene was evaluated by calculating an Isoform Fraction (IF) value, dividing the expression of each individual isoform by the expression level of the gene, where the latter is the sum of the expression of all of its isoforms. Next, changes in isoform fraction (dIF) upon *DSCAM-AS1* silencing were measured as  $dIF = IF_{\text{silencing}} - IF_{\text{control}}$  and significant changes were defined as isoform switching events (details are given in the Materials and Methods, Section 4.6).



**Figure 3.** Isoform switching analysis. (a) Schematic overview of isoform switching concept used by the IsoformSwitchAnalyzeR tool [27]. An isoform switching event is defined as a case where the relative contribution of the isoforms to the parent gene expression changes significantly between conditions. (b) Bar plot reporting the number of isoforms and genes showing significant switching events upon *DSCAM-AS1* silencing. (c) Volcano plot showing the differential isoform fraction (dIF) and the significance of the switching isoforms. In red are reported those isoforms with a  $|dIF| > 10\%$  and an  $FDR < 0.05$ . (d) Summary plot showing the enrichment of specific isoform features (consequence) resulting from the observed isoform switching events. From left to right, the x-axis of the plot shows the fraction of switches having the indicated consequence, where  $<0.5$  means depleted while  $>0.5$  means enriched upon *DSCAM-AS1* silencing. (e) AS event enrichment involved in isoform switches upon *DSCAM-AS1* silencing. The x-axis shows the fraction of genes showing enrichment of a specific AS event upon *DSCAM-AS1* silencing (from left to right). (f) Isoform switching event of the *BCL2* gene

upon *DSCAM-AS1* silencing. The short 3'UTR isoform *BCL2-201* is more commonly used, while the longest 3'UTR isoform *BCL2-202* shows a decreased level upon *DSCAM-AS1* silencing. The different protein domains, *Bcl-2* and *BH-4*, of the isoforms are represented by different colors. The bar plots below show changes in the expression of the *BCL2* gene, *BCL2* isoforms, and their usage upon *DSCAM-AS1* silencing. Multiple Exon Skipping (MES); Intron Retention (IR); Exon Skipping (ES); Alternative Last (AL); Alternative First (AF); Alternative 5' Splicing Site (A5'SS); Alternative 3' Splicing Site (A3'SS); \*\*\*, adjusted-*p*-value < 0.001; \*\*, adjusted-*p*-value < 0.01; \*, adjusted-*p*-value < 0.05; ns, non-significant.

As shown in Figure 3b,c, the isoform switching analysis revealed 485 genes, showing 363 significant switching events involving 578 isoforms (Table S8a). Interestingly, 254 genes (60%) have an isoform switching event with downstream consequences affecting the functional properties of 320 isoforms (Table S8a). Notably, the annotation of the isoforms upregulated upon *DSCAM-AS1* silencing revealed the enrichment of 5'UTR lengthening and 3'UTR-shortening events (Figure 3d). Equally, the analysis revealed that the switching events involved AF and AL prevalently compared to multiple or single exon skipping events (Figure 3e). The switching genes were involved in distinct pathways, including the gland development pathway (*PTPN3*, *RTN4*), cell cycle progression pathway (*POLA1*, *MYCBP*), positive regulation of cell growth (*ZFYVE27*, *TGFBR1*, *TNFRSF12A*, *RPTOR*), apoptosis pathway (*FEM1B*, *BCL2*), and pre-mRNA splicing pathway (*SRSF5*). Interestingly, the *BCL2* gene, whose RNA stability was previously demonstrated to be regulated by hnRNPL [16], was characterized by 3'UTR shortening upon *DSCAM-AS1* downregulation (Figure 3f). The list of switching events affecting the 3'UTR length is reported in Table S8b. An analysis of alternative 3'UTR usage from TCGA BC-data revealed a significant relationship between *DSCAM-AS1* expression and the alternative use of the 3'UTR of 360 genes (Table S9). Among them, 27 were also detected in our analysis in MCF-7 cells (Figure 4a), including the shorter *BCL2* 3'UTR, whose usage was negatively correlated with the *DSCAM-AS1* expression (Figure 4b,c), in coherence with our isoform-switching analysis.

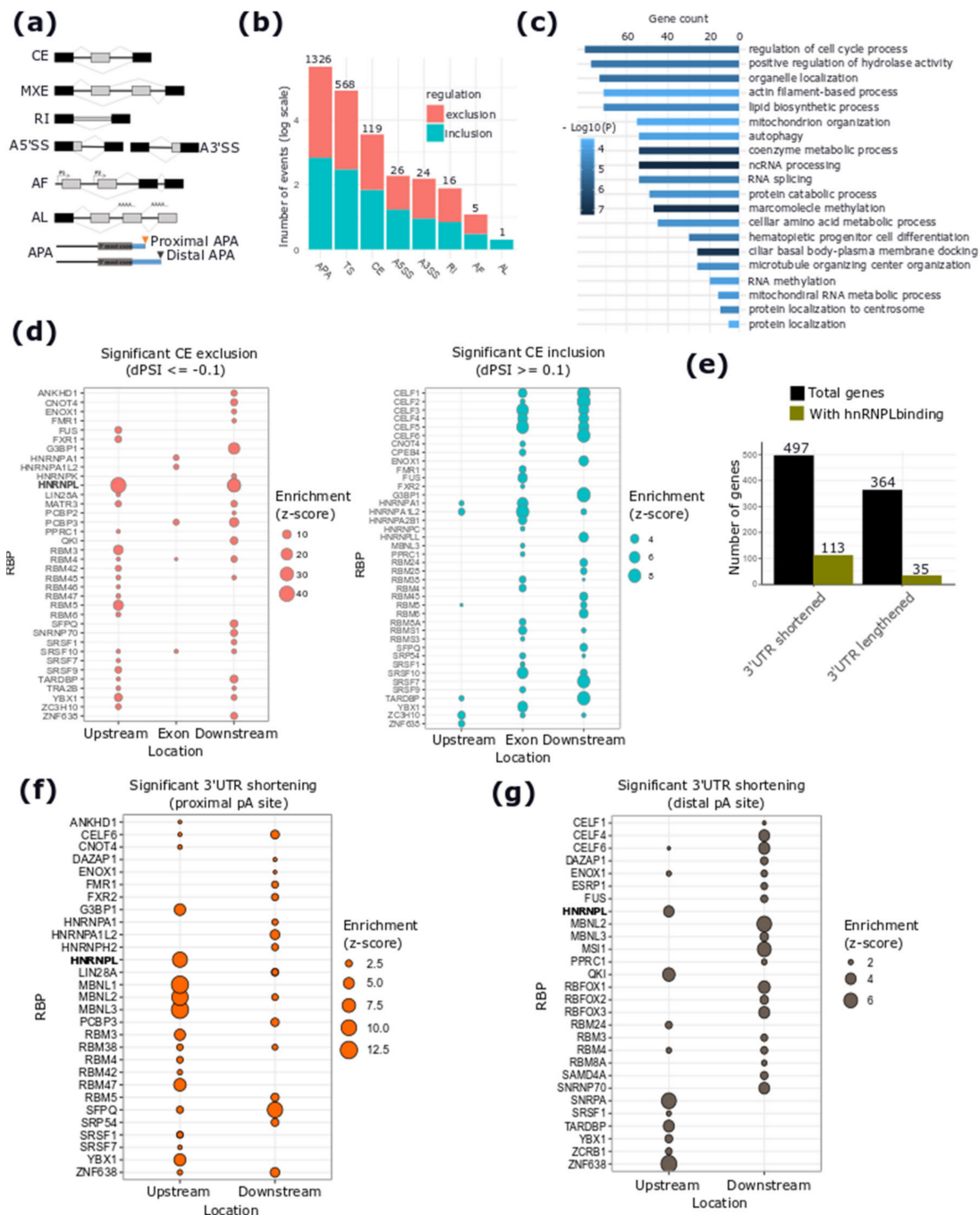


**Figure 4.** (a) Venn diagram showing the overlap between the genes with a differential polyA site usage and 3'UTR regulation in our dataset and those retrieved from The Cancer 3' UTR Atlas (TC3A) database with differential 3'UTR usage showing correlation with *DSCAM-AS1* expression in BCs. (b) Boxplot showing the level of *BCL2* proximal polyA site usage from TC3A database data considering luminal BCs characterized by a high or a low *DSCAM-AS1* expression; \*, Wilcoxon rank-sum test *p*-value < 0.05. (c) Scatterplot showing the relationship between *BCL2* proximal polyA site usage index and *DSCAM-AS1* expression (in FPKM) in BC samples retrieved from the TC3A database. Proximal to Distal polyA site Usage Index (PDU).

In addition to the isoform switching analysis, which detects occurring changes by considering the whole transcript sequence, we performed an analysis of changes involving local AS events upon *DSCAM-AS1* silencing using the Whippet tool [28], which quantifies the inclusion levels for seven AS event types (Figure 5a). The analysis identified 2085 significant AS events (Posterior probability  $p > 0.9$  and  $|\Delta\text{PSI}| > 0.1$ ), differentially regulated upon *DSCAM-AS1* silencing, affecting the splicing pattern of 1339 genes (Figure 5b and Table S10). Interestingly, the most frequent AS event in our data was tandem



Alternative PolyA (APA) sites (1326 events), resulting in the shortening of isoforms from 497 genes (Figure 5b). The biological processes involving the genes affected by AS events were mainly related to cell cycle progression, centrosome regulation, the regulation of metabolism, and ncRNA processing (Figure 5c and Table S11). Interestingly, 48 out of 76 genes related to the cell cycle progression process had an APA event, followed by 15 genes with different alternative Transcriptional Start Sites (TS).



**Figure 5.** Alternative splicing (AS) changes upon DSCAM-AS1 silencing and prediction of binding motifs for RNA-binding proteins (RBPs). (a) Schematic depiction of the different classes of AS events analyzed using Whippet. Exon Skipping (ES); Mutually Exclusive Exons (MXE); Intron Retention (IR); Alternative 5' (left) and 3' (right) Splice Sites (ASS); Alternative First Exons (AF); Alternative Last Exons (AL); Alternative PolyA (APA) site; tandem Transcriptional Start Sites (TS). (b) Number of significant AS events per type and direction of regulation (inclusion or exclusion). (c) GO-enriched terms related

to genes showing significant AS events. (d) RBP-binding motif enrichment results for significant CE (cassette exon) exclusion (left panel) and inclusion events (right panel), significance: z-score > 1.96. (e) Number of genes with significant 3'UTR splicing (shortening and lengthening) with indication of the number of those having a predicted heterogeneous nuclear ribonucleoprotein L (hnRNPL)-binding motif. (f,g), RBP-binding motif enrichment results for genes with significant 3'UTR shortening upon *DSCAM-AS1* silencing, considering a region upstream and downstream of the proximal (f) or the distal (g) APA sites, respectively, significance: z-score >1.96.

#### 2.4. hnRNPL-Binding Motif is Enriched Around Sites of AS Events Regulated by *DSCAM-AS1* Silencing

To identify putative RNA-binding proteins (RBPs) regulating the observed AS changes upon *DSCAM-AS1* silencing, we performed an RBP-binding motif enrichment analysis within the sequences of the exons and the flanking regions involved in AS events. For cassette exon skipping events, the analysis revealed a significant enrichment of the hnRNPL-binding motif (Figure 5d, left panel) within the 200 nucleotides upstream and downstream of the skipped exon. No enrichment was observed within exons, in agreement with the known intronic binding of hnRNPL [29]. This enrichment was observed only for exons characterized by a lower inclusion level upon *DSCAM-AS1* silencing. In contrast, a different subset of RBPs was identified to be enriched for exons showing more inclusion levels upon *DSCAM-AS1* silencing (Figure 5d, right panel). Interestingly, the enrichment analysis revealed a significantly higher presence of the hnRNPL-binding motif in 3'UTR-shortening than in 3'UTR-lengthening events when considering the gene region involved in APA events (Chi-squared test  $p$ -value =  $5.716 \times 10^{-08}$ ) (Figure 5e). Among the 3'UTR-shortening events, the hnRNPL motif was observed to be enriched in regions upstream both of the proximal or of the distal polyA site (Figure 5f,g). The list of RBPs predicted to be enriched for the 3'UTR-lengthening events is shown in Figure S6 and the full list of enriched RBPs enriched for different AS events is reported in Table S12.

### 3. Discussion

In this study, by analyzing *DSCAM-AS1* expression levels across multiple cohorts of BC patients' samples, we confirm *DSCAM-AS1* overexpression in ER+ tumors and its correlation with a worse prognosis. We clearly demonstrate the association of this overexpression with undifferentiated and more aggressive tumors. Furthermore, we show, in MCF-7 BC cells, that *DSCAM-AS1* silencing strongly impairs gene expression. More importantly, we show for the first time that *DSCAM-AS1* influences AS and the alternative 5' and 3' ends of mRNA transcripts, possibly through its interaction with hnRNPL. Since alteration of splicing patterns is emerging as an important hallmark of cancer, we believe that *DSCAM-AS1* may represent an important target for bringing to light novel aspects of BC biology.

The analysis of *DSCAM-AS1* expression in BC samples confirmed the prevalent expression of this lncRNA in ER $\alpha$ -positive BCs of the luminal subtype, as previously shown [9,10]. In coherence with the results from Niknafs and colleagues [10], in the TCGA data, *DSCAM-AS1* expression was not significantly related to particular clinical features or to a different patient outcome. Conversely, our meta-analysis of microarray data clearly highlights the *DSCAM-AS1* overexpression in more aggressive and less differentiated ER-positive BCs. Indeed, *DSCAM-AS1* expression was observed to be increasingly expressed from well to less differentiated BCs and, noteworthy, *DSCAM-AS1* expression was positively related to disease relapse in three datasets, as previously reported in an independent BC cohort from the study of Sun and colleagues [11]. These results are consistent with the previous observation of an increase in *DSCAM-AS1* expression in drug-resistant BCs and cell lines, including those resistant to Tamoxifen, a selective ER $\alpha$  modulator [10,30]. Interestingly, in a recent single-cell analysis of MCF-7 cells resistant to aromatase inhibitors, *DSCAM-AS1* emerged as increasingly expressed in a population of drug-resistant cells [31]. However, despite these promising results, clear evidence of the role of *DSCAM-AS1* as a reliable prognostic marker of BC relapse or drug resistance requires further investigation in a dedicated prospective study. It is interesting to

note that *DSCAM-AS1* was also detected as highly expressed in HER2-amplified, ER-negative tumors, suggesting an ER $\alpha$ -independent mechanism enhancing its expression in this BC subtype. As previously shown by our group, a super-enhancer region is localized upstream of the *DSCAM-AS1* locus and it is involved in the regulation of this lncRNA [13]. This genomic region can represent a platform for the recruitment of different transcriptional regulators, including the Transcription Factor AP2- $\gamma$  (AP2- $\gamma$ ), which was previously reported to lead overexpression of the *ERBB2* gene [32] and, in complex with other transcription factors, it could drive the *DSCAM-AS1* overexpression.

Silencing *DSCAM-AS1* by LNA GapmeRs transfection deeply altered the expression of hundreds of genes in MCF-7 cells. We chose LNA GapmeRs since, in comparison with siRNAs, they were demonstrated to be more stable, more efficient in targeting nuclear RNAs, and they rely on the activation of RNaseH, which cleaves the RNA:LNA hybrids, avoiding the saturation of silencing machineries, like the siRNA/AGO complex [33]. Strikingly, by comparing the gene expression changes occurring upon *DSCAM-AS1* silencing to those occurring upon ER $\alpha$  silencing [34], both *DSCAM-AS1* and ER $\alpha$  were predicted to be involved in the same pathway regulating cell growth and survival of MCF-7 BC cells, consistent with the role of *DSCAM-AS1* as a downstream effector of the ER $\alpha$  signaling pathway [13]. These results are in line with those from Sun and colleagues, reporting that siRNA-mediated *DSCAM-AS1* silencing induces cell cycle arrest in MCF-7 cells [11]. In addition, another study by Xu et al. confirmed the elevated expression of *DSCAM-AS1* in BC cells and showed that the silencing of *DSCAM-AS1* inhibited proliferation and cycle progression, as well as increased cell apoptosis in vitro [35]. Finally, overexpression of *DSCAM-AS1* in T-47D and ZR-75.1 BC cell lines, which show lower expression level of the lncRNA compared to MCF-7 cells, confers a proliferative advantage and a pronounced migratory phenotype [10], suggesting a *DSCAM-AS1* oncogenic role in BC progression.

The expression of many cell cycle-related genes was strongly hampered by *DSCAM-AS1* silencing, including MYC proto-oncogene, bHLH transcription factor (*MYC*), a key regulator of cell growth, proliferation, and apoptosis [36], ret proto-oncogene (*RET*), a well-known proto-oncogene which regulates cell proliferation and survival and a direct target of the ER $\alpha$  signaling pathway [37], Topoisomerase II alpha (*TOP2A*), a proliferation marker whose higher expression in BC is associated with higher tumor grade and the Ki67 index [38]. Furthermore, knocking down *DSCAM-AS1* also reduced the expression levels of genes involved in DNA replication, such as *POL2A*, as well as genes involved in DNA unwinding processes such as mini-chromosome maintenance complex component genes, *MCM4* and *MCM7*. Moreover, the RNA-seq analysis revealed a significant downregulation of apoptosis-related genes such as *BCL2*, an anti-apoptotic gene which has been previously reported to be downregulated upon knocking down *DSCAM-AS1* with siRNA [11].

Given the reported interaction between *DSCAM-AS1* and the splicing factor hnRNPL, also confirmed in this study, we analyzed the expression changes upon *DSCAM-AS1* silencing at isoform level, since these important events are undetectable when the analysis is related only to gene level. Indeed, the expression analysis at the isoform level reveals that almost more than half of the genes with at least one DE isoform in our RNA-seq dataset were not likewise classified as DE when considering gene level measurements only (Figure S4b). Although a direct GO analysis of isoforms is not applicable, we found that the genes with DE isoforms are involved in cellular processes similar to those observed for DE genes that were identified by gene-level analysis. This result demonstrates that, by different mechanisms, at both gene and isoform levels, *DSCAM-AS1* silencing affects the same pathways in MCF-7 cells. Furthermore, we applied an isoform switching analysis to identify cases where DE isoforms of the same gene were dysregulated in opposite directions, indicating a change in the expression of the gene that could not be appreciated by a classical gene-level analysis. Thus, we believe and foresee that genes showing isoforms dysregulated in opposite directions could be studied further to decipher their associated relevance with the levels of expression of *DSCAM-AS1* in the context of BC.

Interestingly, the RNA-binding motif enrichment analysis we performed revealed distinct subsets of RBPs as putative regulators of AS events. Noteworthy, all the RBPs predicted to be enriched are highly expressed in MCF-7 cells and five of them (CELF1, HNRNPA1, RBM24, MBNL3, SFPQ) were differentially expressed or spliced upon *DSCAM-AS1* silencing (Tables S2 and S10). Specifically, the hnRNPL-binding motif was predicted as the most frequent, among other RBPs motifs, in the case of exon skipping events, with almost all the predicted binding sites located in intronic regions upstream or downstream the regulated cassette exon. This is in line with previous evidence from literature showing that hnRNPL inhibits exon inclusion through its preferential binding to CA-rich splicing silencer elements located either upstream or downstream of the regulated exons [29]. The intronic binding motif enrichment of hnRNPL suggests that downregulation of *DSCAM-AS1* may reduce the fraction of hnRNPL engaged in the interaction, increasing the fraction available to regulate cassette exon splicing in other target RNAs. However, a more complex positional code has been proposed by recent studies to explain the activating or inhibiting activity of hnRNPL on target exons [14]. Thus, we believe that further experiments are strictly needed to decipher the mechanism or mechanisms by which *DSCAM-AS1* regulates the function or the binding of hnRNPL. We also found other RBPs known to inhibit exon inclusion such as HNRNPA1, a member of the heterogeneous nuclear ribonucleoprotein protein family [39]. In contrast, the binding motifs of a different set of RBPs, known to induce exon inclusion, such as CUG-Repeat Binding Protein (CUGBP) Elav-like family member 1 (CELF) and Splicing Regulator (SR) proteins, were predicted to be enriched in the case of exon inclusion events. Interestingly, among the enriched RBPs identified around *DSCAM-AS1*-mediated splicing events, we found PCBP2, which was previously reported to bind *DSCAM-AS1* in an RNA-pulldown coupled to a mass spectrometry experiment [10]. PCBP2 is a protein known to interact with hnRNPL [40], suggesting a complex network of splicing regulation related to *DSCAM-AS1*-hnRNPL interaction.

Furthermore, in addition to hnRNPL, we successfully identified a set of RBPs previously known to regulate APA site selection in the case of 3'UTR-shortening/lengthening events, including muscleblind-like proteins 1, 2 and 3 (MBNL1, MBNL2 and MBNL3), which are known to bind preferentially to 3'UTR regions [41], Ras guanosine triphosphate (GTP)-ase-activating protein-binding protein 1 (G3BP1), an essential splicing factor for normal stress granule assembly and, consequently, the preservation of polyadenylated mRNAs [42], Splicing Factor proline/glutamine rich (SFPQ) known to facilitate miRNA-target binding [42,43], as well as the RNA-binding motif 47 protein (RBM47), which was previously reported to inhibit the proliferation of different BC cell lines and whose binding was predominantly at 3'UTRs [44].

*DSCAM-AS1* silencing had a strong effect on alternative polyadenylation site selection, in accordance with the 3'UTR-shortening events identified by the isoform switching analysis. 3'UTR shortening driven by alternative polyadenylation was previously shown to be extensively present both in normal and in cancer cells [45] and, in the latter, it could involve both oncogenes and tumor suppressor genes [46–48]. Our findings of 3'UTR-shortening events are in line with a study by Wang and colleagues, where they show the enrichment of shortened 3'UTRs in samples derived from BC patients and characterized by a low proliferation rate [49]. Interestingly, in our data, 24 different genes are characterized by 3'UTR shortening upon *DSCAM-AS1* silencing and are annotated to cell cycle, DNA replication or apoptosis-related terms, including *BCL2*, *CASP2*, *CDKN2C*, *EGF*, *CEP290*, and *CCNE1* (Table S13). The stability of the antiapoptotic gene *BCL2* was previously reported to be regulated by hnRNPL, which prevents the trigger of the NMD by directly interacting with the longer 3'UTR of the gene [16]. However, the study from Lim and colleagues suggested that the 3'UTR splicing of *BCL2* is independent of hnRNPL in MCF-7, proposing the presence of an unknown factor interacting with the hnRNPL in the regulation of this splicing event [50]. In our data, the shortening of the *BCL2* isoform, as well as the gene downregulation, is coherent with the impairment of hnRNPL-mediated stability, which could be directly mediated by the interaction with *DSCAM-AS1*. Furthermore, the analysis of data of 3'UTR usage in tumor samples from TCGA confirmed differential *BCL2* 3'UTR usage in relation to the *DSCAM-AS1* expression. Despite the interaction between *DSCAM-AS1* and hnRNPL at

BCL2, 3'UTR deserves further experimental validations, and *DSCAM-AS1* represents a good candidate to regulate this process in luminal breast cancer cells.

Furthermore, in the study of Xue and colleagues, *CCNE1* 3'UTR lengthening was observed to be recurrently detected in six cancer types including BC [46]. However, to our knowledge, no evidence of the interaction between *CCNE1* 3'UTR and hnRNPL was previously described, and definitely, further investigations are needed to clarify the functional consequence of the 3'UTR alteration on the genes identified in our analysis. Indeed, the prediction of the functional consequence of a 3'UTR-shortening event is not trivial and requires dedicated experimental validation. This is due to the heterogeneous and connected molecular pathways that can be affected by such an AS event, including a widespread alteration in the network of microRNA–target interactions [48], inhibition or induction of the NMD pathway [51], or the generation of novel RNA fragments [52].

The hypothesis of a *DSCAM-AS1*-mediated regulation of hnRNPL interaction at gene 3'UTR is supported by previous evidence of lincRNA–RBP interaction in the post-transcriptional regulation of mRNAs, including the linc-RoR mediated increase in the mRNA stability of c-Myc gene by interaction with hnRNPI and AUF1 at gene 3'UTR [53]. Noteworthy, while hnRNPL expression was not affected upon *DSCAM-AS1* silencing (Figure S7), the hnRNPL-binding motif was among the most enriched in the case of exon skipping and alternative polyadenylation site events, suggesting that the observed AS changes might be caused by the disruption of the hnRNPL–*DSCAM-AS1* interaction upon lincRNA silencing. Our hypothesis about the direct activity of hnRNPL on the detected AS events upon *DSCAM-AS1*-silencing was further supported by data analysis of published studies on hnRNPL activity and RNA-binding profile in different cell lines [14,15,54]. Specifically, 47 out of 119 cassette exon events occurring upon LNA-mediated *DSCAM-AS1* silencing were observed also upon siRNA-mediated silencing of hnRNPL in LNCaP, an androgen receptor-positive prostate cancer cell line [15] (Figure S8 and Table S14a). Furthermore, by overlapping the coordinates of the genomic regions involved in the AS events with data of hnRNPL–RNA interactions from CLIP/RIP (RNA Immunoprecipitation)-based assays from these three studies, we observed 605 out of 2085 AS events identified in our analysis characterized by hnRNPL-binding events detected in at least one study (Table S14b). These results support a direct role of hnRNPL on the regulation of the AS pattern of these genes.

Finally, we fully exploited our RNA-Seq dataset potential and analyzed changes occurring upon *DSCAM-AS1* silencing at both gene and isoform levels. Importantly, by computational analysis, we successfully identified changes in AS, further supporting our *DSCAM-AS1*-hnRNPL interaction hypothesis. Further experiments, such as hnRNPL CLIP-seq, could shed light on the precise molecular mechanisms by which *DSCAM-AS1* interacts with hnRNPL to regulate alternative splicing.

## 4. Materials and Methods

### 4.1. Experimental Part

#### 4.1.1. Cell Culture and LNA GapmeR™ Transfection

MCF-7 and SK-BR-3 cells were routinely grown in DMEM (Dulbecco's Modified Eagle's Medium) (Life Technologies, 31053–028, Waltham, MA, USA) supplemented with 10% heat-inactivated FBS (fetal bovine serum) (Euroclone S.p.A, ECS0180L, Milan, Italy) and 2 mM L-glutamine (ThermoFisher Scientific, 25030–024, Waltham, MA, USA). Batches of human cell lines were purchased from ATCC (American Type Culture Collection). Cells were cultured at 37 °C with 5% CO<sub>2</sub>. The cell transfection in suspension was performed by seeding one million MCF-7 or SK-BR-3 cells along with a transfection mix composed of LNAs (20 nM final concentration) and Lipofectamine3000 (ThermoFisher Scientific, L3000015, Waltham, MA, USA), as transfecting reagents. Cells were left to attach overnight in the cell incubator and, subsequently, the medium was refreshed. Experiments were performed in three biological replicates 48 h after LNA transfection. Three custom-designed LNAs GapmeR™ from Exiqon were used to target all the *DSCAM-AS1* lincRNA isoforms (LNA\_1 5'-ATGGCAGTTGGAGGAG-3',

LNA\_2 5'-ACAGAGAAGGACATGG-3' and LNA\_3 5'-AAGTAGCTTCATCTTT-3'); negative control A-LNA longRNA GapmeR™ was used as a control LNA (5'-AACACGTCTATACGC-3'); Exiqon, 300611-00. Now Exiqon is part of Qiagen company, Hilden, Germany.

#### 4.1.2. RNA Isolation and Quantitative Real Time PCR (qRT-PCR) Analysis

Total RNA used for downstream qRT-PCR analysis was isolated from MCF-7 and SK-BR-3 cells using the PureZOL™ reagent (BioRad, 7326890, Hercules, CA, USA), according to the manufacturer's protocol. Total RNA used for downstream RNA-seq analysis was isolated from MCF-7 cells using the RNeasy Mini Kit, according to the manufacturer's protocol (Qiagen, 74104, Hilden, Germany). Frozen breast cancer tissues of Cohort\_1 [9] and breast cancer tissue powders of Cohort\_2 (supercooled biopsies pulverized using a micro-dismembrator (Braun, Melsungen, Germany) and stored at  $-80^{\circ}\text{C}$ ) were directly homogenized in PureZOL™ to extract total RNA. All total RNA samples purified with PureZOL™ reagent were subjected to DNase treatment to remove contaminating genomic DNA (Invitrogen™ ezDNase™ Enzyme, ThermoFisher Scientific, 11766051). First-strand cDNA synthesis was performed with the SensiFAST™ cDNA Synthesis Kit (Bioline, BIO-65054, London, UK). qRT-PCR analysis was performed using the SYBR®-green method (SensiFAST SYBR® Hi-ROX Kit, Bioline, BIO-92005, London, UK). Real-time PCR primers were custom-designed or purchased from Qiagen (249900 QuantiTect Primer Assay, Qiagen, Hilden, Germany) and are listed in Table S15. The expression of *18S* was used to normalize the expression level of specific targets.

#### 4.1.3. Quantification of *DSCAM-AS1* Expression in Primary Tumor Tissue Samples

The *DSCAM-AS1* expression was measured as described above in primary tissue samples from two different patient cohorts, namely Cohort\_1 and Cohort\_2.

Cohort\_1 consists of 42 samples and these were analyzed as described in [9] with updated follow-up information. The study was conducted in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans. The research protocol was approved by the Research Ethics Committee of the University of Turin. Considering that the present retrospective study did not modify the patients' treatment and was conducted anonymously, no specific written informed consent was required.

Cohort\_2 consists of 51 samples deriving from the Certified Biobank of the Regional Center for Biomarkers, Department of Clinical Pathology, Azienda ULSS 3 Serenissima, Venice, Italy. The samples were obtained from patients with primary BC previously collected for diagnostic purposes, whose data were anonymized by coding. For these samples, archived for more than twenty years, the need for informed consent was waived by the ethics committee in accordance with the requirements of Italian law (Italian Data Protection Authority—Garante Privacy, Authorization no. 9/2014—General Authorization to Process Personal Data for Scientific Research Purposes. Published in Italy's Official Journal No. 301 on 30th December 2014).

In Cohort\_2, *DSCAM-AS1* expression levels were considered high (1 = positive) if values of dCT from qRT-PCR analysis were  $<16$  or  $>16$  when *DSCAM-AS1* CT value was  $<28$ . Conversely, *DSCAM-AS1* expression levels were considered low (0 = negative) if values of dCT from qRT-PCR analysis were  $>16$  or  $<16$  when *DSCAM-AS1* CT value was  $>28$ . The complete clinical information of the subjects included in these cohorts is reported in Table S1b,c. Differences in *DSCAM-AS1* levels measured in samples from these subjects were statistically evaluated using the Wilcoxon rank-sum test and Chi-square test.

#### 4.1.4. Cross-Linking and Immunoprecipitation (CLIP) of hnRNPL

In total, 2.5 million MCF-7 cells were seeded in 15cm diameter dishes and cultured for 48 h before the crosslinking with 0.37% Formaldehyde (Merck Millipore, F8775, Burlington, MA, USA) for 10 min at  $37^{\circ}\text{C}$ . The crosslinking reaction was stopped with 125 mM Glycine (Biorad, 1610717, Hercules, CA, USA) for 5 min at room temperature. Cells were then washed twice in 1X PBS

supplemented with 1% protease inhibitor cocktail (Merck Millipore, P2714, Burlington, MA, USA) and PMSF (Phenylmethylsulfonyl Fluoride) (Merck Millipore, 93482, Burlington, MA, USA), collected by scraping in 1 ml of 1X PBS, pelleted by 5 min centrifugation 4000 rpm at 4 °C and lysed for 30 min on ice with RIPA lysis buffer (50 mM Tris HCl, 150 mM NaCl, 1.0% NP-40, 0.5% Sodium Deoxycholate, 1.0 mM EDTA, 0.1% SDS) supplemented with 1% protease inhibitor cocktail (Merck Millipore, P2714, Burlington, MA, USA), PMSF (Merck Millipore, 93482, Burlington, MA, USA) and RNase-inhibitor (RNaseOUT™ Recombinant Ribonuclease Inhibitor, ThermoFisher Scientific, 10777019, Waltham, MA, USA). Total extracts were sonicated for three cycles (20" ON, 30" OFF) by using an immersion sonicator device. After spinning for 10 min, 13,000 rpm at 4 °C, MCF-7 extracts were incubated for 2 h with 30 µL of BSA-coated beads (Protein G Sepharose 4 Fast Flow, GE Healthcare 17061801, Chicago, IL, USA) for sample pre-clearing. The 5% of total volume was collected as an input and stored at −20 °C until the end of the immuno-precipitation steps. Pre-cleared samples were incubated O/N at 4 °C with 10 µg of normal rabbit IgG (Merck Millipore, 12-370, Burlington, MA, USA) or of two different anti-hnRNPL antibodies (anti-hnRNPL 4D11 and D-5, Santa Cruz Biotechnology, sc-32317 and sc-48391, Dallas, TX, USA) on a rotating platform. IP (immunoprecipitation) samples were then incubated for 2 h with 50 µL of BSA-coated beads (Protein G Sepharose 4 Fast Flow, GE Healthcare 17061801, Chicago, IL, USA). IP samples were then washed five times in the RIPA lysis buffer. After the washing steps, dried beads and input samples were resuspended in RIPA lysis buffer without SDS and de-crosslinked at 65 °C for 1 h at 650 rpm by adding Proteinase K (Proteinase K Solution 20 mg/mL, ThermoFisher Scientific, AM2546, Waltham, MA, USA). Immuno-precipitated RNA was extracted with PureZOL™ reagent (BioRad, 7326890, Hercules, CA, USA), according to the manufacturer's protocol, cDNA was synthesized and analyzed by qRT-PCR as described above. hnRNPL-CLIP enrichment was normalized on input samples and expressed as the enrichment of specific binding over nonspecific IgG control binding.

#### 4.1.5. Cell Proliferation Assay by Crystal-Violet Staining

SK-BR-3 cells were transfected with LNA-CTRL or LNA-DSCAM-AS1 as described and  $5 \times 10^5$  cells/well were seeded in a 96-MW microtiter plate in five technical replicates for each experimental condition. After 48 h from transfection, the culture medium was removed, and cells were washed twice in 1X PBS and fixed with methanol for 5 min. The staining was carried out with a 1% Crystal Violet solution for 10 min at room temperature (Merck Millipore, V5265, Burlington, MA, USA). Several washes with dH<sub>2</sub>O were performed to remove excess staining. The incorporated dye was solubilized and eluted with 10% glacial acetic acid solution and the absorbance was measured at 595 nm.

#### 4.1.6. Western Blot

Whole-cell lysate was harvested in boiling lysis buffer (25 mM Tris-HCl pH 7.6, 1% SDS, 1 mM EDTA, 1 mM EGTA) and 50 µg of total protein extract was loaded into an 8% Acrylamide gel. The antibodies used were designed against hnRNPL (Santa Cruz Biotechnology, sc-48391, Dallas, TX, USA) and GAPDH (Santa Cruz Biotechnology, sc-32233, Dallas, TX, USA). Densitometry readings were measured by the Image Lab Software (Biorad, version number 6.0.1.) and intensity ratio calculations are reported in Table S16. Uncropped western blot images are reported in Figure S9.

#### 4.2. Analysis of DSCAM-AS1 Expression with Respect to Different Clinical Data

The analysis of the *DSCAM-AS1* in tumors characterized by different clinical parameters was performed by considering the data from 30 public microarray experiments (Table S1a) and from TCGA. Only microarray datasets characterized by at least 30 samples were selected. Furthermore, since *DSCAM-AS1* is a well-known ER- $\alpha$  target, only datasets composed of ER+ tumor subtypes or associated with the information of the ER status were selected in order to separately perform the analysis on ER+ and ER− tumors. The analysis of the *DSCAM-AS1* expression with respect to different

clinical data was performed using shinyGEO tools [55] in default settings. The analysis was performed only for clinical data associated with at least three samples per class analyzed. shinyGEO was also used to perform the survival analyses on datasets associated with information on patient overall survival or relapse-free survival.

The analysis of DSCAM-AS1 expression in TCGA datasets was performed by retrieving the processed RNA-Seq datasets from the GDC (Genomic Data Commons) data portal [56]. Specifically, the FPKM expression levels of 1040 subjects were retrieved from this portal. DSCAM-AS1 expression was evaluated with respect to subjects' clinical covariates. Clinical data were obtained from cBioPortal [57], considering the clinical data from the dataset named "Breast Invasive Carcinoma (TCGA, Firehose Legacy)". The analysis was performed separately for ER+ and ER- tumors separated based on the IHC level of ER. The significance of the differential expression between two sample classes was computed using the Wilcoxon rank-sum test. Survival analysis on TCGA data was performed using the survival v3.1 R package and ggsurvplot function of the survminer 0.4.6 R package.

#### 4.3. RNA Sequencing Libraries Preparation and Data Analysis

MCF-7 cells were transfected with control or DSCAM-AS1-targeting LNAs and RNA was extracted 48 h after transfection, as described above. An RNA quality check (RNA integrity number (RIN) > 8) was achieved with a Fragment Analyzer (Advanced Analytical Technologies, Inc., Ankeny, IA, USA) and quantified with Qubit (Qubit™ RNA HS Assay Kit, ThermoFisher Scientific, Waltham, MA, USA; Q32852). Two µg of RNA were polyA+ selected and RNA-Seq libraries were constructed using Illumina TruSeq RNA sample prep kit (TruSeq™ RNA Sample Prep Kit v2-Set B, Illumina, RS-122-2002, San Diego, CA, USA). Paired-end (PE) cluster generation was performed using cBot on Flow Cell v3 (TruSeq PE Cluster Kit v3-cBot-HS, Illumina, PE-401-3001, San Diego, CA, USA). The sequencing of libraries was performed on the HiSeq sequencing system (Illumina, San Diego, CA, USA). The raw RNA-Seq data were deposited Gene Expression Omnibus (GEO) with the identifier GSE150591.

A flow chart of the mRNA sequencing data analysis, including pre-processing and quality control, quantification at both gene and isoform levels, the identification of differentially expressed genes and isoforms and the identification of differentially regulated alternative splicing (AS) events is summarized in Figure S10 and Table S17.

Raw reads were assessed for Phred quality scores using the FASTQC program (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), and low bases and adaptor sequences were trimmed off using Fqtrim (<http://ccb.jhu.edu/software/fqtrim/>), retaining only reads of 75 base length. Then, clean reads were aligned against the human reference genome (GRCh38.p10 assembly) with Gencode v27 annotation (gencode.v27.annotation.gtf.gz) using STAR v2.5.1b [58]. STAR was run in two-pass mode, allowing alignment to the transcriptome coordinates by setting the option quantMode to TranscriptomeSAM; summary statistics of read alignment per sample are given in Table S18. Next, the expression levels in read counts, transcript per million fragments mapped (TPM), and FPKM units were then estimated at both gene and isoform levels by running RSEM (RNA-Seq by Expectation Maximization) [59] on the alignment files in default parameters.

#### 4.4. Differential Expression Analysis

Differentially expressed genes and isoforms upon DSCAM-AS1 silencing compared to the control condition were identified using the DESeq2 R package (v1.26.0) in default parameters [60]. The expression at isoform level was summarized to gene level using the *tx-import* bioconductor package [61] and the resulting count matrices were provided to DESeq2. Prior to DE analysis, lowly expressed genes and isoforms were discarded from the analysis and only genes or isoforms with more than 10 normalized read counts in at least one condition (three out of six samples) were considered for further downstream analysis. A gene or isoform was assigned as differentially expressed if its associated BH-adjusted *p*-value < 0.05. For all the data visualization plots, including heat maps and



volcano plots, MA (mean average) plots were made using the ggplot2 R package (v.3.2.1) [62]. A quality control check of the replicates is given in Figure S11.

#### 4.5. Gene Ontology Enrichment Analysis

Gene ontology terms enriched for upregulated and downregulated genes were obtained using the Gene Annotation and Analysis Resource Metascape program [63]. The list of upregulated and downregulated genes were analyzed separately using the Single List Analysis option. The statistically enriched GO terms related to each category of genes were obtained from the GO Biological Processes branch. Only GO terms that were associated with an enrichment factor > 1.5 and an accumulative hypergeometric test with an adjusted  $p$ -value < 0.05 were considered to be significant. To reduce redundancy, the GO terms showing a high number of overlapping genes and a large degree of redundancies were clustered into groups based on their degree of similarity and each group or cluster was represented by the most significant GO term. The top 20 significant clusters were selected for visualization purposes.

#### 4.6. Isoform Switching Analysis

To test for isoform switching events, IsoformSwitchAnalyzeR tool was applied [27]. Briefly, from the RNA-seq data, the tool takes, as inputs, isoform expression levels quantified in transcript per million fragments mapped (TPM) units normalized to transcript length and then calculates an isoform fraction (IF) ratio by dividing the isoform expression with the expression of the parent gene ( $TPM_{iso}/TPM_{gene}$ ). Lowly expressed genes with less than 1 TPM and lowly expressed isoforms not contributing to the expression of the gene ( $IF < 0.01$ ) were excluded from downstream analysis. The IF was then calculated per each of the remaining isoforms and per condition. For each isoform, a  $dIF$  ( $IF_{silencing} - IF_{control}$ ) representing the difference in isoform usage between the two conditions was calculated. A cut-off criterion was applied by selecting only those isoforms for which DSCAM-AS1 silencing induced a significant change (BH-corrected  $p$ -value  $\leq 0.05$ ) in IF by at least 10% (i.e.,  $|dIF| > 0.1$ ). Next, the sequences corresponding to those isoforms showing significant switching events upon DSCAM-AS1 silencing were extracted and then annotated for the presence of signal peptide sequences, coding potential and for their associated pfam protein domains using signalP [64], CPC2 [65] and Pfam [66] tools, respectively. The biological consequences of the observed switches, including intron retention, domain gain/loss, coding/non-coding potential and shortening/lengthening of the open reading frame were then evaluated for the switching isoforms from the same parent gene. Next, according to the applied annotation on the switching isoforms, genes were classified into genes with or without downstream functional consequences.

#### 4.7. Differential Alternative Splicing Analysis

The list of differentially regulated AS events upon DSCAM-AS1 silencing were identified using Whippet [28]. Whippet is a tool which takes, as inputs, a gene annotation model file together with a genome file and generates a contiguous splice graph index representation of each gene included in the annotation file. The reads are then directly mapped against the contiguous splice index. Splicing events are represented as nodes by Whippet, where each node corresponds to an exonic region of the gene and the incoming and outgoing edges to each node define the set of reads supporting its inclusion and exclusion, respectively. Whippet provides a splice index (PSI) as a measure of the inclusion level of each node by calculating the ratio of the paths supporting the inclusion of the node divided by the total number of the paths supporting both the inclusion and exclusion of that node. Whippet quantifies changes in different possible splicing events including Alternative First (AF) or Last (AL) exons, Single (SES) and Multiple Exon Skipping (MES) events, Alternative splice sites (A5'SS and A3'SS), Mutually Exclusive exons (MXE), Intron Retention (RI) events in addition to alternative Transcription Start (TS) and Alternative PolyA (APA) sites. All the sequences and annotations used in this analysis were based on GRCh38 genome assembly and Gencode v27 annotation. To ensure the quantification of expressed

events, a prefiltering criterion was applied by only considering those splicing events whose supporting reads were at least 10 in at least two samples per condition. In addition, a splicing event with a  $\Delta$ PSI value between the silencing and control conditions less than 10% ( $|\Delta$ PSI  $<$  0.1), or whose associated posterior probability was less than 0.90, were excluded from the downstream analysis.

#### 4.8. RBP-Binding Motif Enrichment Analysis

To identify RNA-binding proteins as putative regulators of the observed changes in each splicing event identified, the sequences of the regulated exon skipping events (ES), extended by  $\pm$  200 nucleotides on both sides, were scanned for the occurrence of RBP-binding motifs. For alternative polyadenylation sites, the APA nodes reported by Whippet were sorted by coordinates and classified as proximal or distal APA sites, both of which were extended by 100 nucleotides on both sides from the APA site position. The RNA-binding motifs for 105 different splicing factors collected from the RNAcompete study [67] were used to perform binding motif enrichment analysis. For a number of RBPs, the motif from different species was confirmed, in a previous study [68], to be conserved between the human and the other species. This includes RBM47 (chicken), SF1 (Drosophila), SRP4 (Drosophila), TRA2 (Drosophila), and PCBP3 (mouse). Next, the MoSEA (Motif Scan and Enrichment Analysis) package was used to search the sequence of the splicing events for the occurrence of RBP-binding motifs [68]. The tool Find Individual Motif Occurrences (FIMO) [69] was used to scan the sequences of the events for the presence of the RBP motifs using a  $p$ -value  $<$  0.001 as a cut-off. The binding motif enrichment was performed by comparing the number of occurrences of the binding motifs of the RBPs in the regulated events with that observed in a pool of 100 randomly selected sequences of the same size from equivalent regions in non-regulated events ( $|\Delta$ PSI  $<$  0.01 and  $p <$  0.5). Motif enrichment was performed separately for the two directions of splicing changes ( $\Delta$ PSI  $>$  0.1 or  $\Delta$ PSI  $<$  -0.1). An enrichment  $z$ -score per RNA-binding motif, the region, and direction of regulation was calculated by normalizing the observed frequency in the regulated events set with the mean and standard deviation of the 100 random control sets. The 100 random control sequences were sampled from non-regulated events for each region of regulation. An RBP was considered as enriched if associated with a  $z$ -score  $>$  1.96. The obtained  $z$ -scores per binding motif, region, and event were then visualized using the ggplot2 Bioconductor package [62].

#### 4.9. Analysis of Alternative 3'UTR Usage in TCGA Data

The analysis of alternative 3'UTR usage of breast cancer tissues from the TCGA was performed by considering the data from The Cancer 3' UTR Atlas (TC3A) [70]. In this project, the APA usage in each tumor is quantified as the Percentage of Distal polyA site Usage Index (PDUI), which indicates a fraction of a specific polyA site used among those of a specific gene. The PDUI values of 10,267 3'UTR sites were downloaded from the project website (<http://tc3a.org/>) by selecting those from the BRCA cohort. Then, data of 638 luminal breast cancer tumors associated with DSCAM-AS1 FPKM expression values were extracted and considered for the analysis. To test the relationship between DSCAM-AS1 and the 3'UTR usage, a Pearson correlation analysis and a Wilcoxon rank-sum test analysis (separating tumors in two classes based on the median expression of DSCAM-AS1) was performed. Only the events characterized by a significant result in both analyses were considered significant.

#### 4.10. Overlap with Public hnRNPL RNA-Binding Experiments

The overlap between the genomic regions involved in the AS events upon DSCAM-AS1 silencing and the hnRNPL RNA-binding sites was performed by considering the AS region provided by Whippet extended upstream and downstream of 200 bp. The hnRNPL-binding sites considered were retrieved from data of three different studies [14,15,54]. Specifically, the three replicates of the hnRNPL iCLIP experiment from [14] were retrieved from GEO (Gene Expression Omnibus) page GSE37562, while the RIP-Seq from [15] were retrieved from GEO page GSE72841. Finally, the HITS-CLIP (High-Throughput Sequencing of RNA isolated by CrossLinking Immunoprecipitation) data from [54] were retrieved

from the Dorina v2.0 database [71], considering the results from untreated CD4+ and Jurkat cell lines. The coordinates of the hnRNPL–RNA-binding peaks were converted in hg38 using the LiftOver webtool (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>) and the overlap with the regions involved in AS events was performed using the coverageBed function of the BEDTool [72]. An overlap was considered valid if associated with a coverage greater than zero.

## 5. Conclusions

The data presented here represent a definite advancement in understanding the role of the lncRNA *DSCAM-AS1* in BC cell growth. Transcriptome analyses clearly indicate *DSCAM-AS1* involvement in the regulation of splicing and 3′-end usage, an aspect that is considered increasingly important in cancer biology, and that also correlates strongly with the observation that BC cells express an abnormal, and subtype-specific, variety of circRNAs [73]. Finally, we believe that *DSCAM-AS1* has potential as a marker of luminal BC that can be exploited for studies on the resistance to endocrine treatments in advanced BC.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2072-6694/12/6/1453/s1>, Figure S1: Analysis of *DSCAM-AS1* expression in ER-negative BCs and results from survival analysis; Figure S2: Effect of *DSCAM-AS1* silencing in SK-BR3 cell line proliferation; Figure S3: Ontological analysis of DE genes not coherently expressed between this study and the experiment of ER $\alpha$  silencing; Figure S4: Results from differential isoform expression analysis; Figure S5: Result of the *DSCAM-AS1*-hnRNPL CLIP experiment; Figure S6: The list of RBPs enriched in the 3′UTR-lengthening events upstream and downstream of proximal and distal APA sites. Figure S7: Analysis of hnRNPL protein level upon *DSCAM-AS1* silencing. Figure S8: Number of overlapping cassette exons significantly regulated in this study and upon siRNA-mediated silencing of hnRNPL in LNCaP prostate cancer cell line. Figure S9: Uncropped western blot image related to the results showed in Figure S7. Figure S10: Flow chart showing the different bioinformatic tools and steps used in this study. Figure S11: Quality control check of the RNA-seq dataset. Table S1: Datasets information and results from the analysis of *DSCAM-AS1* expression with respect to different clinical data; Table S2: Results from the gene-level differential expression analysis of *DSCAM-AS1* LNA experiment; Table S3: Results from the gene set enrichment analysis of *DSCAM-AS1* LNA differentially expressed genes; Table S4: Comparison of gene differentially expressed in the *DSCAM-AS1* LNA experiment and in MCF-7 cells treated with ER siRNA; Table S5: Results from the gene set enrichment analysis of genes overlapped between *DSCAM-AS1* LNA and ER siRNA experiments; Table S6: Results from the isoform-level differential expression analysis of *DSCAM-AS1* LNA experiment; Table S7: Results from the gene set enrichment analysis of parent genes of *DSCAM-AS1* LNA differentially expressed isoforms; Table S8: Isoform features from the isoform switching analysis; Table S9: Results of the correlation and statistical analysis between *DSCAM-AS1* expression and the 3′UTR usage data in luminal breast cancer from the TC3A project; Table S10: Results from differential AS analysis using Whippet; Table S11: Results from the gene set enrichment analysis of genes characterized by differential alternative spliced events upon *DSCAM-AS1* silencing. Table S12: List of enriched RBPs binding motifs among different AS events; Table S13: List of genes with 3′UTR-shortening/lengthening events reported by Whippet; Table S14: a, Results from the overlap between the list of cassette exons events upon *DSCAM-AS1* silencing and those from hnRNPL silencing experiment in LNCaP prostate cancer cell line. b, Results from the overlap between the genomic coordinates of the region involved in the AS events and the hnRNPL–RNA-binding regions obtained by public CLIP/RIP-based experiments; Table S15: Custom-designed primer sequences and purchased primers catalog numbers; Table S16: WB densitometry readings and intensity ratio calculation. Table S17: Bioinformatic programs used in this study. Their versions and references are reported. Table S18: Alignment summary statistics of RNA-seq reads reported by STAR. The total number of reads and the mapping rates are given for each sample.

**Author Contributions:** Conceptualization, J.E., G.F., V.M.; bioinformatic analysis, J.E., G.F.; experimental analysis, V.M.; samples preparation and clinical data collection regarding Cohort\_1, L.A., I.C., A.S.; samples selection and clinical data collection regarding Cohort\_2, A.E.L., A.S.C.F.; writing—Original draft preparation, J.E., G.F., V.M., M.D.B.; writing—Review and editing, J.E., G.F., V.M., E.D., M.D.B.; visualization, J.E. and G.F.; supervision, M.D.B.; project administration, M.D.B.; funding acquisition, M.D.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Associazione Italiana per la Ricerca sul Cancro (Grant IG 15600 to M.D.B.); by Fondazione CRT (grant 2017.0823 to M.D.B.); and by the University of Torino (2018 Local Research funding to M.D.B.).

**Acknowledgments:** We thank Salvatore Oliviero from the Italian Institute for Genomic Medicine (IIGM, Torino, Italy) for the RNA-seq experiment carried out in his laboratory.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Slack, F.J.; Chinnaiyan, A.M. The role of non-coding RNAs in oncology. *Cell* **2019**, *179*, 1033–1055. [[CrossRef](#)] [[PubMed](#)]
2. Carlevaro-Fita, J.; PCAWG Drivers and Functional Interpretation Group; Lanzós, A.; Feuerbach, L.; Hong, C.; Mas-Ponte, D.; Pedersen, J.S.; Johnson, R.; PCAWG Consortium. Cancer lncRNA census reveals evidence for deep functional conservation of long noncoding RNAs in tumorigenesis. *Commun. Biol.* **2020**, *3*, 56. [[CrossRef](#)] [[PubMed](#)]
3. Sahu, A.; Singhal, U.; Chinnaiyan, A.M. Long noncoding RNAs in cancer: From function to translation. *Trends Cancer* **2015**, *1*, 93–109. [[CrossRef](#)] [[PubMed](#)]
4. Van Grembergen, O.; Bizet, M.; De Bony, E.J.; Calonne, E.; Putmans, P.; Brohée, S.; Olsen, C.; Guo, M.; Bontempi, G.; Sotiriou, C.; et al. Portraying breast cancers with long noncoding RNAs. *Sci. Adv.* **2016**, *2*, e1600220. [[CrossRef](#)]
5. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2020. *CA Cancer J. Clin.* **2020**, *70*, 7–30. [[CrossRef](#)] [[PubMed](#)]
6. Rani, A.; Stebbing, J.; Giamas, G.; Murphy, J. Endocrine resistance in hormone receptor positive breast cancer—from mechanism to therapy. *Front. Endocrinol.* **2019**, *10*, 245. [[CrossRef](#)]
7. Mathias, C.; Zambalde, E.P.; Rask, P.; Gradia, D.F.; De Oliveira, J.C. Long non-coding RNAs differential expression in breast cancer subtypes: What do we know? *Clin. Genet.* **2019**, *95*, 558–568. [[CrossRef](#)]
8. Liu, D.; Rudland, P.S.; Sibson, D.R.; Barraclough, R. Identification of mRNAs differentially-expressed between benign and malignant breast tumour cells. *Br. J. Cancer* **2002**, *87*, 423–431. [[CrossRef](#)]
9. Miano, V.; Ferrero, G.; Reineri, S.; Caizzi, L.; Annaratone, L.; Ricci, L.; Cutrupi, S.; Castellano, I.; Cordero, F.; De Bortoli, M.; et al. Luminal long non-coding RNAs regulated by estrogen receptor alpha in a ligand-independent manner show functional roles in breast cancer. *Oncotarget* **2015**, *7*, 3201–3216. [[CrossRef](#)]
10. Niknafs, Y.S.; Han, S.; Ma, T.; Speers, C.; Zhang, C.; Wilder-Romans, K.; Iyer, M.K.; Pitchiaya, S.; Malik, R.; Hosono, Y.; et al. The lncRNA landscape of breast cancer reveals a role for DSCAM-AS1 in breast cancer progression. *Nat. Commun.* **2016**, *7*, 12791. [[CrossRef](#)]
11. Sun, W.; Li, A.-Q.; Zhou, P.; Jiang, Y.-Z.; Jin, X.; Liu, Y.-R.; Guo, Y.-J.; Yang, W.-T.; Shao, Z.-M.; Xu, X.-E.; et al. DSCAM-AS1 regulates the G/S cell cycle transition and is an independent prognostic factor of poor survival in luminal breast cancer patients treated with endocrine therapy. *Cancer Med.* **2018**, *7*, 6137–6146. [[CrossRef](#)] [[PubMed](#)]
12. Vu, T.N.; Pramana, S.; Calza, S.; Suo, C.; Lee, D.; Pawitan, Y. Comprehensive landscape of subtype-specific coding and non-coding RNA transcripts in breast cancer. *Oncotarget* **2016**, *7*, 68851–68863. [[CrossRef](#)] [[PubMed](#)]
13. Miano, V.; Ferrero, G.; Rosti, V.; Manitta, E.; Elhasnaoui, J.; Basile, G.; De Bortoli, M. Luminal lncRNAs regulation by ER $\alpha$ -controlled enhancers in a ligand-independent manner in breast cancer cells. *Int. J. Mol. Sci.* **2018**, *19*, 593. [[CrossRef](#)]
14. Rossbach, O.; Hung, L.-H.; Khrameeva, E.E.; Schreiner, S.; König, J.; Curk, T.; Zupan, B.; Ule, J.; Gelfand, M.; Bindereif, A.; et al. Crosslinking-immunoprecipitation (iCLIP) analysis reveals global regulatory roles of hnRNP L. *RNA Biol.* **2014**, *11*, 146–155. [[CrossRef](#)] [[PubMed](#)]
15. Fei, T.; Chen, Y.; Xiao, T.; Li, W.; Cato, L.; Zhang, P.; Cotter, M.B.; Bowden, M.; Lis, R.T.; Zhao, S.G.; et al. Genome-wide CRISPR screen identifies HNRNPL as a prostate cancer dependency regulating RNA splicing. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E5207–E5215. [[CrossRef](#)] [[PubMed](#)]
16. Kishor, A.; Ge, Z.; Hogg, J.R. HnRNP L-dependent protection of normal mRNAs from NMD subverts quality control in B cell lymphoma. *EMBO J.* **2019**, *38*, 38. [[CrossRef](#)]
17. Jia, R.; Zhang, S.; Liu, M.; Zhang, Y.; Liu, Y.; Fan, M.; Guo, J. HnRNP L is important for the expression of oncogene SRSF3 and oncogenic potential of oral squamous cell carcinoma cells. *Sci. Rep.* **2016**, *6*, 35976. [[CrossRef](#)]
18. Gaudreau, M.-C.; Grapton, D.; Helness, A.; Vadnais, C.; Fraszczak, J.; Shooshtarizadeh, P.; Wilhelm, B.; Robert, F.; Heyd, F.; Mörröy, T.; et al. Heterogeneous Nuclear Ribonucleoprotein L is required for the survival and functional integrity of murine hematopoietic stem cells. *Sci. Rep.* **2016**, *6*, 27379. [[CrossRef](#)]
19. Barash, Y.; Calarco, J.A.; Gao, W.; Pan, Q.; Wang, X.; Shai, O.; Blencowe, B.J.; Frey, B.J. Deciphering the splicing code. *Nature* **2010**, *465*, 53–59. [[CrossRef](#)]

20. Reyes, A.; Huber, W. Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. *Nucleic Acids Res.* **2017**, *46*, 582–592. [[CrossRef](#)]
21. Eswaran, J.; Horvath, A.; Godbole, S.; Reddy, S.D.; Mudvari, P.; Ohshiro, K.; Cyanam, D.; Nair, S.; Fuqua, S.A.W.; Polyak, K.; et al. RNA sequencing of cancer reveals novel splicing alterations. *Sci. Rep.* **2013**, *3*, 3. [[CrossRef](#)] [[PubMed](#)]
22. Matlin, A.J.; Clark, F.; Smith, C.W. Understanding alternative splicing: Towards a cellular code. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 386–398. [[CrossRef](#)] [[PubMed](#)]
23. Wiesner, T.; Lee, W.; Obenauf, A.C.; Ran, L.; Murali, R.; Zhang, Q.F.; Wong, E.W.P.; Hu, W.; Scott, S.N.; Shah, R.; et al. Alternative transcription initiation leads to expression of a novel ALK isoform in cancer. *Nature* **2015**, *526*, 453–457. [[CrossRef](#)] [[PubMed](#)]
24. Davuluri, R.V.; Suzuki, Y.; Sugano, S.; Plass, C.; Huang, T.H.-M. The functional consequences of alternative promoter use in mammalian genomes. *Trends Genet.* **2008**, *24*, 167–177. [[CrossRef](#)]
25. Rajan, P.; Elliott, D.; Robson, C.N.; Leung, H.Y. Alternative splicing and biological heterogeneity in prostate cancer. *Nat. Rev. Urol.* **2009**, *6*, 454–460. [[CrossRef](#)]
26. Sveen, A.; Kilpinen, S.; Ruusulehto, A.; A Lothe, R.; I Skotheim, R. Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. *Oncogene* **2015**, *35*, 2413–2427. [[CrossRef](#)]
27. Vitting-Seerup, K.; Sandelin, A. IsoformSwitchAnalyzeR: Analysis of changes in genome-wide patterns of alternative splicing and its functional consequences. *Bioinformatics* **2019**, *35*, 4469–4471. [[CrossRef](#)]
28. Sterne-Weiler, T.; Weatheritt, R.; Best, A.; Ha, K.C.H.; Blencowe, B.J. Whippet: An efficient method for the detection and quantification of alternative splicing reveals extensive transcriptomic complexity. *bioRxiv* **2017**, 158519. [[CrossRef](#)]
29. Hui, J.; Hung, L.-H.; Heiner, M.; Schreiner, S.; Neumüller, N.; Reither, G.; Haas, S.; Bindereif, A. Intronic CA-repeat and CA-rich elements: A new class of regulators of mammalian alternative splicing. *EMBO J.* **2005**, *24*, 1988–1998. [[CrossRef](#)]
30. Ma, Y.; Bu, D.; Long, J.; Chai, W.; Dong, J. LncRNA DSCAM-AS1 acts as a sponge of miR-137 to enhance Tamoxifen resistance in breast cancer. *J. Cell. Physiol.* **2018**, *234*, 2880–2894. [[CrossRef](#)]
31. Hong, S.P.; Chan, T.E.; Lombardo, Y.; Corleone, G.; Rotmensz, N.; Bravaccini, S.; Rocca, A.; Pruneri, G.; McEwen, K.; Coombes, R.C.; et al. Single-cell transcriptomics reveals multi-step adaptations to endocrine therapy. *Nat. Commun.* **2019**, *10*, 3814–3840. [[CrossRef](#)] [[PubMed](#)]
32. Perissi, V.; Menini, N.; Cottone, E.; Capello, D.; Sacco, M.; Montaldo, F.; De Bortoli, M. AP-2 transcription factors in the regulation of ERBB2 gene transcription by oestrogen. *Oncogene* **2000**, *19*, 280–288. [[CrossRef](#)] [[PubMed](#)]
33. Hagedorn, P.H.; Persson, R.; Funder, E.D.; Albæk, N.; Diemer, S.L.; Hansen, D.J.; Møller, M.R.; Papargyri, N.; Christiansen, H.; Hansen, B.R.; et al. Locked nucleic acid: Modality, diversity, and drug discovery. *Drug Discov. Today* **2018**, *23*, 101–114. [[CrossRef](#)] [[PubMed](#)]
34. Caizzi, L.; Ferrero, G.; Cutrupi, S.; Cordero, F.; Ballare, C.; Miano, V.; Reineri, S.; Ricci, L.; Friard, O.; Testori, A.; et al. Genome-wide activity of unliganded estrogen receptor- $\alpha$  in breast cancer cells. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 4892–4897. [[CrossRef](#)]
35. Xu, S.; Kong, D.; Chen, Q.; Ping, Y.; Pang, D. Oncogenic long noncoding RNA landscape in breast cancer. *Mol. Cancer* **2017**, *16*, 129. [[CrossRef](#)]
36. Xu, J.; Chen, Y.; Olopade, O.I. MYC and breast cancer. *Genes Cancer* **2010**, *1*, 629–640. [[CrossRef](#)]
37. Stine, Z.E.; McGaughey, D.; Bessling, S.L.; Li, S.; McCallion, A.S. Steroid hormone modulation of RET through two estrogen responsive enhancers in breast cancer. *Hum. Mol. Genet.* **2011**, *20*, 3746–3756. [[CrossRef](#)]
38. An, X.; Xu, F.; Luo, R.; Zheng, Q.; Lu, J.; Yang, Y.; Qin, T.; Yuan, Z.; Shi, Y.-X.; Jiang, W.-Q.; et al. The prognostic significance of topoisomerase II alpha protein in early stage luminal breast cancer. *BMC Cancer* **2018**, *18*, 331. [[CrossRef](#)]
39. Burd, C.; Dreyfuss, G. RNA binding specificity of hnRNP A1: Significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J.* **1994**, *13*, 1197–1204. [[CrossRef](#)]
40. Kim, J.H.; Hahm, B.; Kim, Y.K.; Choi, M.; Jang, S.K. Protein-protein interaction among hnRNPs shuttling between nucleus and cytoplasm. *J. Mol. Biol.* **2000**, *298*, 395–405. [[CrossRef](#)]
41. Batra, R.; Charizanis, K.; Manchanda, M.; Mohan, A.; Li, M.; Finn, D.J.; Goodwin, M.; Zhang, C.; Sobczak, K.; Thornton, C.A.; et al. Loss of MBNL leads to disruption of developmentally regulated alternative polyadenylation in RNA-mediated disease. *Mol. Cell* **2014**, *56*, 311–322. [[CrossRef](#)] [[PubMed](#)]

42. Aulas, A.; Caron, G.; Gkogkas, C.G.; Mohamed, N.-V.; Destroismaisons, L.; Sonenberg, N.; Leclerc, N.; Parker, A.; Velde, C.V. G3BP1 promotes stress-induced RNA granule interactions to preserve polyadenylated mRNA. *J. Cell Biol.* **2015**, *209*, 73–84. [[CrossRef](#)] [[PubMed](#)]
43. Bottini, S.; Hamouda-Tekaya, N.; Mategot, R.; Zaragosi, L.-E.; Audebert, S.; Pisano, S.; Grandjean, V.; Mauduit, C.; Benahmed, M.; Barbry, P.; et al. Post-transcriptional gene silencing mediated by microRNAs is controlled by nucleoplasmic Sfpq. *Nat. Commun.* **2017**, *8*, 1189. [[CrossRef](#)] [[PubMed](#)]
44. Vanharanta, S.; Marney, C.; Shu, W.; Valiente, M.; Zou, Y.; Mele, A.; Darnell, R.B.; Massagué, J. Loss of the multifunctional RNA-binding protein RBM47 as a source of selectable metastatic traits in breast cancer. *eLife* **2014**, *3*, 3. [[CrossRef](#)]
45. Gruber, A.J.; Zavolan, M. Alternative cleavage and polyadenylation in health and disease. *Nat. Rev. Genet.* **2019**, *20*, 599–614. [[CrossRef](#)]
46. Xue, Z.; Warren, R.L.; Gibb, E.A.; Macmillan, D.; Wong, J.; Chiu, R.; Hammond, S.A.; Yang, C.; Nip, K.M.; Ennis, C.A.; et al. Recurrent tumor-specific regulation of alternative polyadenylation of cancer-related genes. *BMC Genom.* **2018**, *19*, 536. [[CrossRef](#)]
47. Mayr, C.; Bartel, B. Widespread shortening of 3'UTRs by Alternative cleavage and Polyadenylation activates oncogenes in cancer cells. *Cell* **2009**, *138*, 673–684. [[CrossRef](#)]
48. Park, H.J.; Ji, P.; Kim, S.; Xia, Z.; Rodriguez, B.; Li, L.; Su, J.; Chen, K.; Masamha, C.P.; Baillat, D.; et al. 3' UTR shortening represses tumor-suppressor genes in trans by disrupting ceRNA crosstalk. *Nat. Genet.* **2018**, *50*, 783–789. [[CrossRef](#)]
49. Wang, X.; Li, M.; Yin, Y.; Li, L.; Tao, Y.; Chen, D.; Li, J.; Han, H.; Hou, Z.; Zhang, B.; et al. Profiling of alternative polyadenylation sites in luminal B breast cancer using the SAPAS method. *Int. J. Mol. Med.* **2014**, *35*, 39–50. [[CrossRef](#)]
50. Lim, M.-H.; Lee, D.-H.; Jung, S.E.; Youn, D.-Y.; Park, C.S.; Lee, J.-H. Effect of Modulation of hnRNP L Levels on the Decay of bcl-2 mRNA in MCF-7 Cells. *Korean J. Physiol. Pharmacol.* **2010**, *14*, 15–20. [[CrossRef](#)]
51. Kurosaki, T.; Popp, M.W.; Maquat, L.E. Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 406–420. [[CrossRef](#)] [[PubMed](#)]
52. Malka, Y.; Steiman-Shimony, A.; Rosenthal, E.; Argaman, L.; Cohen-Daniel, L.; Arbib, E.; Margalit, H.; Kaplan, T.; Berger, M. Post-transcriptional 3'-UTR cleavage of mRNA transcripts generates thousands of stable uncapped autonomous RNA fragments. *Nat. Commun.* **2017**, *8*, 2029. [[CrossRef](#)] [[PubMed](#)]
53. Huang, J.; Zhang, A.; Ho, T.-T.; Zhang, Z.; Zhou, N.; Ding, X.; Zhang, X.; Xu, M.; Mo, Y.-Y. Linc-RoR promotes c-Myc expression through hnRNP I and AUF1. *Nucleic Acids Res.* **2015**, *44*, 3059–3069. [[CrossRef](#)] [[PubMed](#)]
54. Shankarling, G.; Cole, B.S.; Mallory, M.J.; Lynch, K.W. Transcriptome-wide RNA interaction profiling reveals physical and functional targets of hnRNP L in human T cells. *Mol. Cell. Biol.* **2013**, *34*, 71–83. [[CrossRef](#)]
55. Dumas, J.; Gargano, M.A.; Dancik, G.M. shinyGEO: A web-based application for analyzing gene expression omnibus datasets. *Bioinformatics* **2016**, *32*, 3679–3681. [[CrossRef](#)]
56. Grossman, R.L.; Heath, A.P.; Ferretti, V.; Varmus, H.E.; Lowy, U.R.; Kibbe, W.A.; Staudt, L.M. Toward a shared vision for cancer genomic data. *N. Engl. J. Med.* **2016**, *375*, 1109–1112. [[CrossRef](#)]
57. Gao, J.; Aksoy, B.A.; Dogrusoz, U.; Dresdner, G.; Gross, B.; Sumer, S.O.; Sun, Y.; Skanderup, A.J.; Sinha, R.; Larsson, E.; et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **2013**, *6*, 11. [[CrossRef](#)]
58. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2012**, *29*, 15–21. [[CrossRef](#)]
59. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **2011**, *12*, 323. [[CrossRef](#)]
60. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 002832. [[CrossRef](#)]
61. Sonesson, C.; I Love, M.; Robinson, M.D. Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000Research* **2015**, *4*, 1521. [[CrossRef](#)] [[PubMed](#)]
62. Wilkinson, L. ggplot2: Elegant graphics for data analysis by WICKHAM, H. *Biometrics* **2011**, *67*, 678–679. [[CrossRef](#)]
63. Zhou, Y.; Zhou, B.; Pache, L.; Chang, M.; Khodabakhshi, A.H.; Tanaseichuk, O.; Benner, C.; Chanda, S. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **2019**, *10*, 1523. [[CrossRef](#)] [[PubMed](#)]

64. Armenteros, J.J.A.; Tsirigos, K.; Sønderby, C.K.; Petersen, T.N.; Winther, O.; Brunak, S.; Von Heijne, G.; Nielsen, H. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **2019**, *37*, 420–423. [[CrossRef](#)]
65. Kang, Y.-J.; Yang, D.-C.; Kong, L.; Hou, M.; Meng, Y.-Q.; Wei, L.; Gao, G. CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* **2017**, *45*, W12–W16. [[CrossRef](#)] [[PubMed](#)]
66. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.; Qureshi, M.; Richardson, L.J.; A Salazar, G.; Smart, A.; et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2019**, *47*, D427–D432. [[CrossRef](#)]
67. Ray, D.; Ha, K.C.; Nie, K.; Zheng, H.; Hughes, T.R.; Morris, Q. RNAcompete methodology and application to determine sequence preferences of unconventional RNA-binding proteins. *Methods* **2016**, 3–15. [[CrossRef](#)]
68. Sebestyén, E.; Singh, B.; Miñana, B.; Pagès, A.; Mateo, F.; Pujana, M.A.; Valcárcel, J.; Eyra, E. Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *Genome Res.* **2016**, *26*, 732–744. [[CrossRef](#)]
69. Grant, C.E.; Bailey, T.L.; Noble, W.S. FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **2011**, *27*, 1017–1018. [[CrossRef](#)] [[PubMed](#)]
70. Feng, X.; Li, L.; Wagner, E.J.; Li, W. TC3A: The cancer 3' UTR atlas. *Nucleic Acids Res.* **2018**, *46*, D1027–D1030. [[CrossRef](#)]
71. Blin, K.; Dieterich, C.; Wurmus, R.; Rajewsky, N.; Landthaler, M.; Akalin, A. DoRiNA 2.0—Upgrading the doRiNA database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res.* **2014**, *43*, D160–D167. [[CrossRef](#)] [[PubMed](#)]
72. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842. [[CrossRef](#)] [[PubMed](#)]
73. Coscujuela Tarrero, L.; Ferrero, G.; Miano, V.; De Intinis, C.; Ricci, L.; Arigoni, M.; Riccardo, F.; Annaratone, L.; Castellano, I.; Calogero, R.A.; et al. Luminal breast cancer-specific circular RNAs uncovered by a novel tool for data analysis. *Oncotarget* **2018**, *9*, 14580–14596. [[CrossRef](#)] [[PubMed](#)]



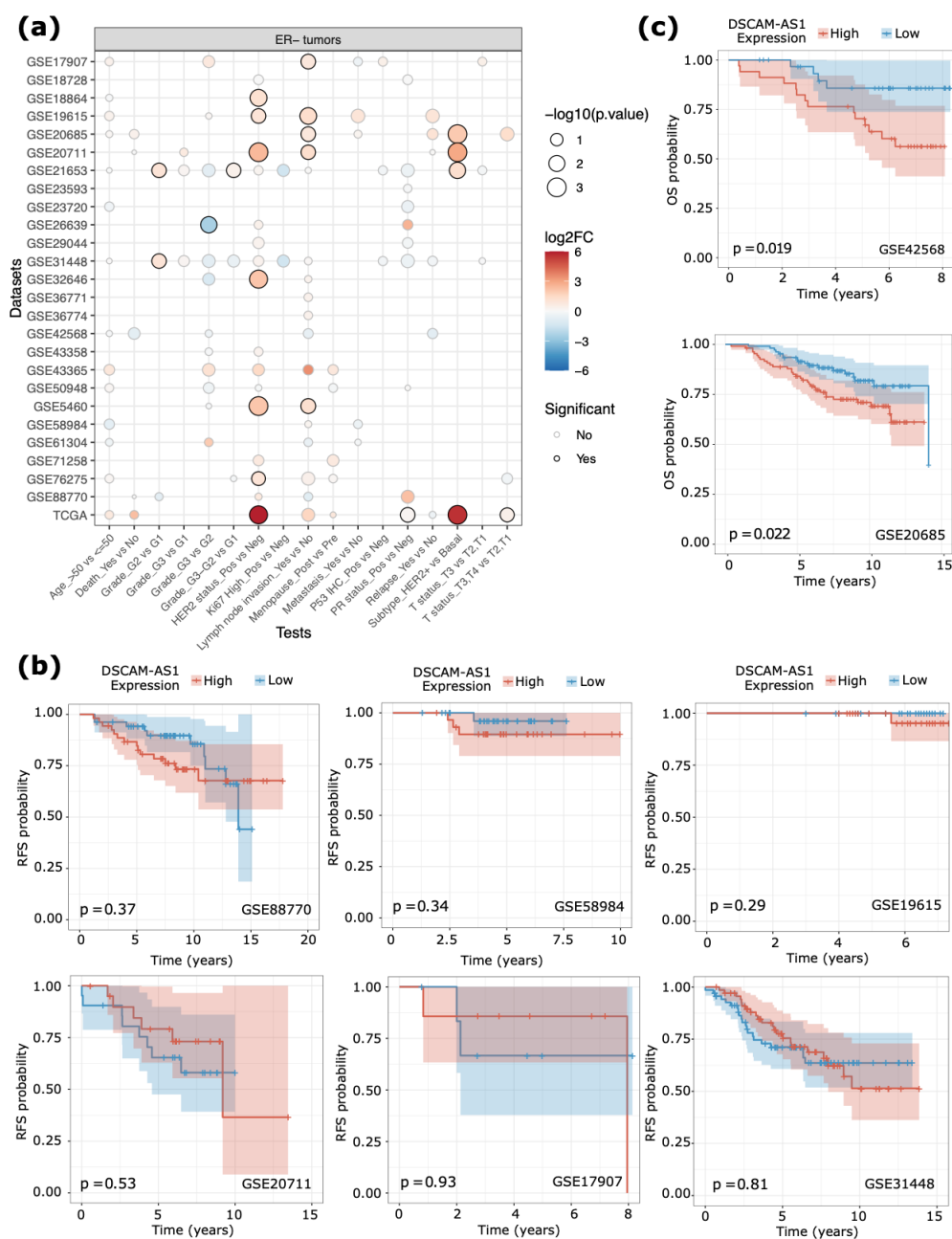
© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Article

# DSCAM-AS1-Driven Proliferation of Breast Cancer Cells Involves Regulation of Alternative Exon Splicing and 3'-End Usage

Jamal Elhasnaoui, Valentina Miano, Giulio Ferrero, Elena Doria, Antonette E. Leon, Aline S.C. Fabricio, Laura Annaratone, Isabella Castellano, Anna Sapino and Michele De Bortoli

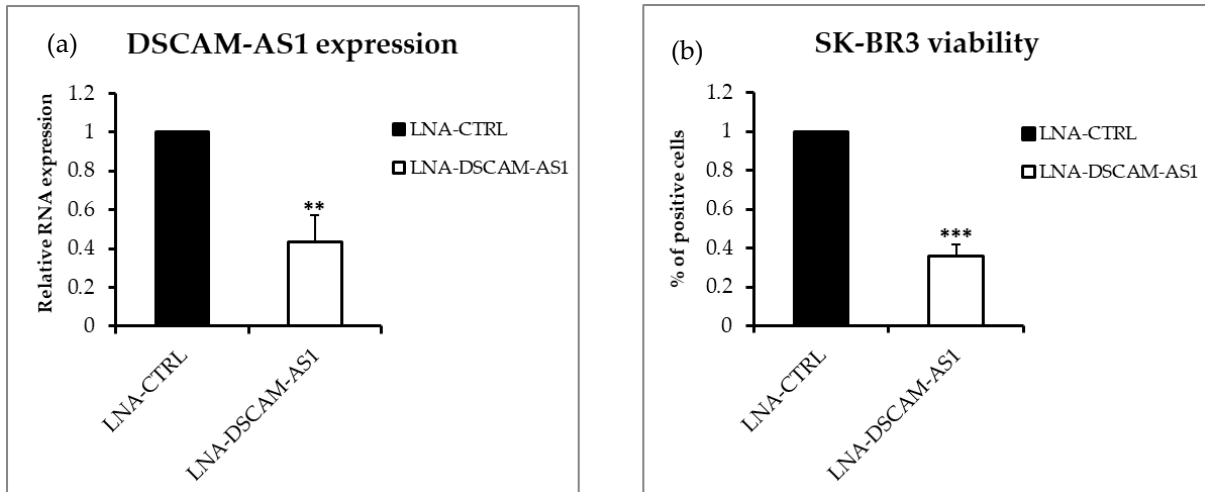
## Supplementary Figures



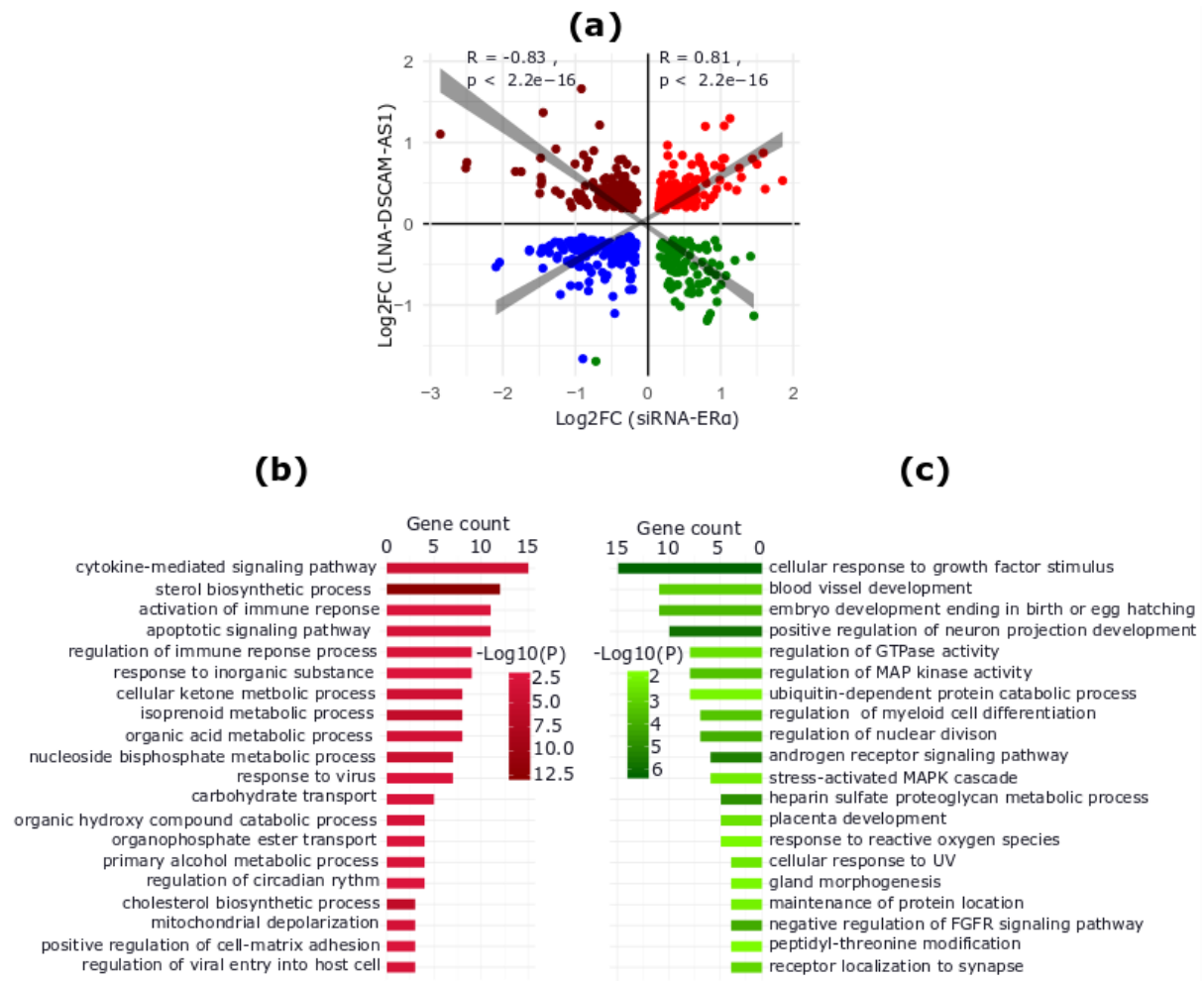
**Figure S1.** (a) Dot plot reporting the level of statistical significance of the differential *DSCAM-AS1* expression analyses between groups of ER-negative BC patients separated with respect to specific



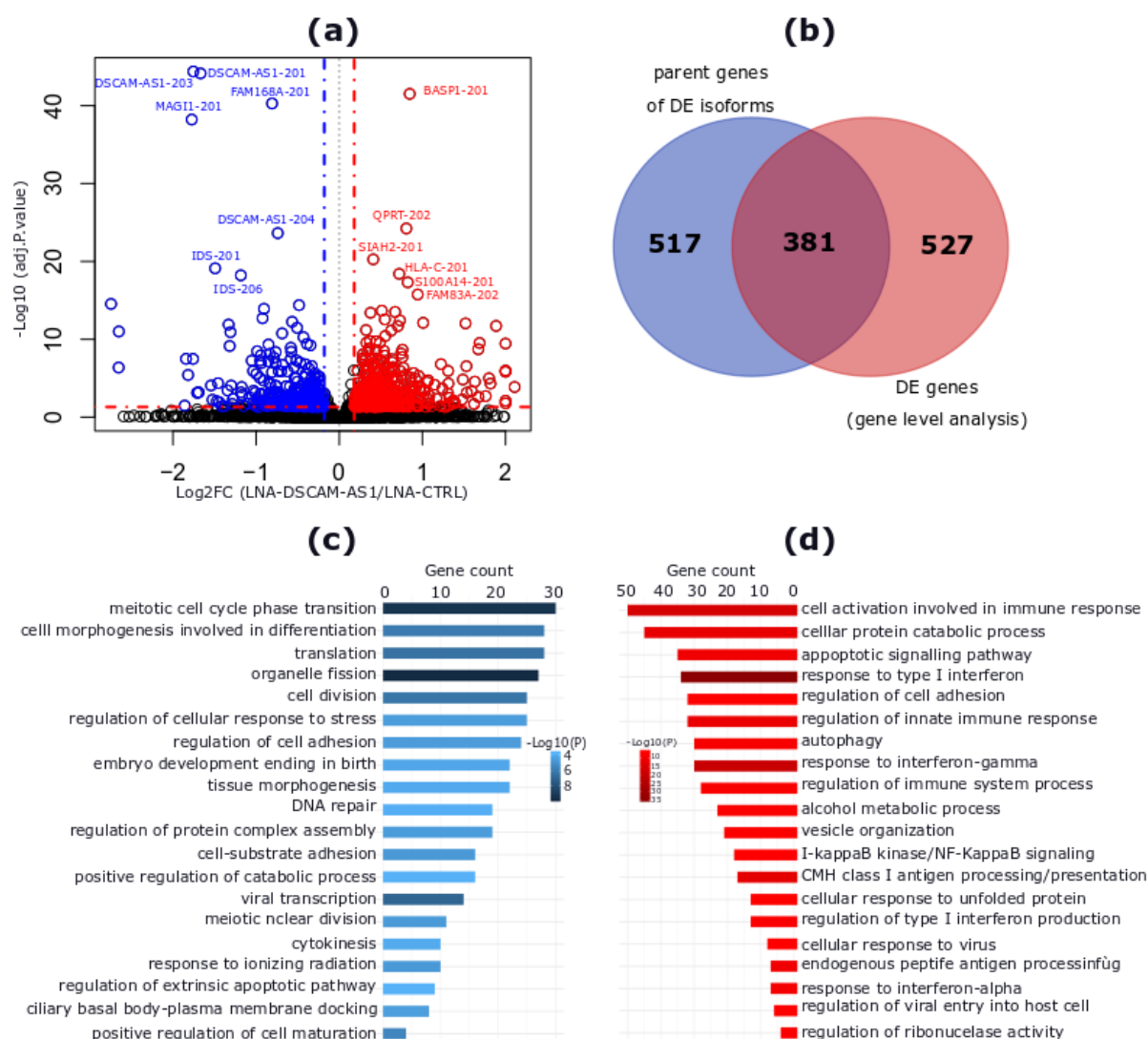
clinical data. The size of the dot is proportional to the significance of the results while the color code represents the log<sub>2</sub>FC of expression. ER, Estrogen Receptor; Pos, positive; Neg, negative; PR, Progesterone Receptor. (b) Kaplan-Meier curves representing the datasets with nonsignificant difference in Relapse Free Survival (RFS) of BC patients based on the median level of DSCAM-AS1 expression. *p*-value by log-rank test. (c) Kaplan-Meier curves representing the Overall Survival (OS) of BC patients based on the median level of DSCAM-AS1 expression. *p*-value by log-rank test.



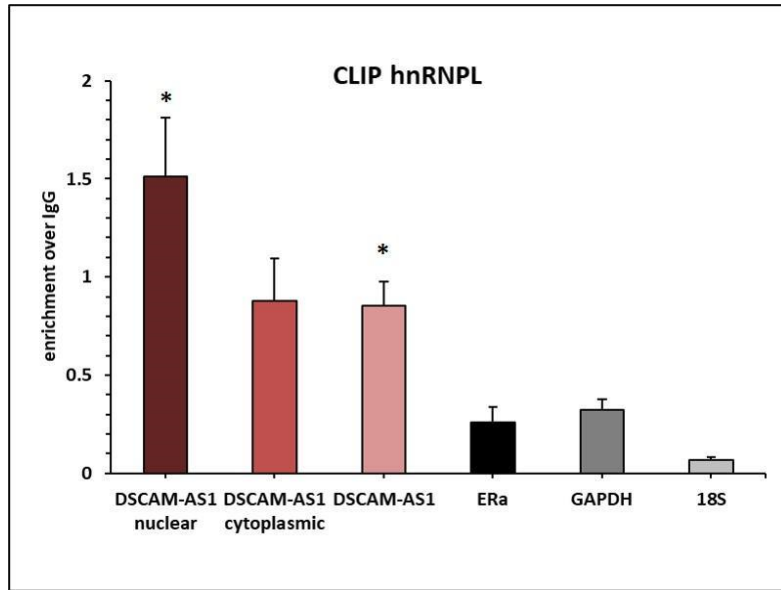
**Figure S2.** (a) Expression levels of *DSCAM-AS1* and (b) viability measure by Crystal Violet Assay in SK-BR-3 cells upon transection of *DSCAM-AS1*-targeting or control LNA. Error bars represent the standard deviation of three biological replicates. Significance from T-test: \*\*, *p*-value < 0.01; \*\*\*, *p*-value < 0.001.



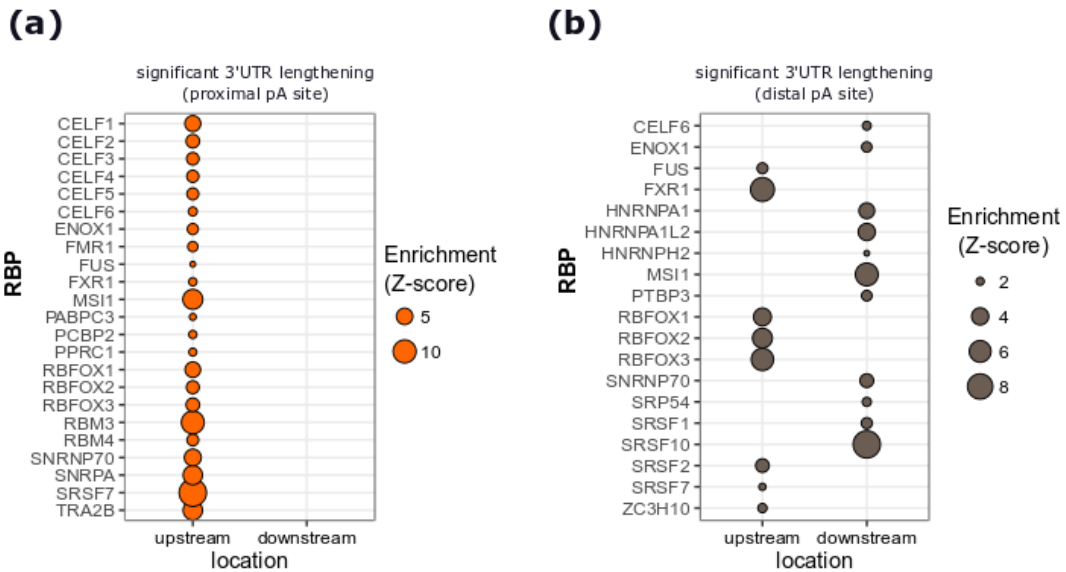
**Figure S3.** (a) The scatter plot shows the  $\text{log}_2$  fold change of DE genes between this study and those obtained upon ER $\alpha$  silencing (siER $\alpha$ ) [13]. Dark red and green dots represent genes upregulated in this study while downregulated in the siER $\alpha$  experiment, and those downregulated in this study and upregulated in the siER $\alpha$  experiment, respectively. Blue and red dots represent those genes downregulated or upregulated in both studies, respectively. (b) GO terms enriched for genes upregulated in this study and downregulated in the siER $\alpha$  experiment. (c) GO terms enriched for genes downregulated in this study while upregulated in the siER $\alpha$  experiment.



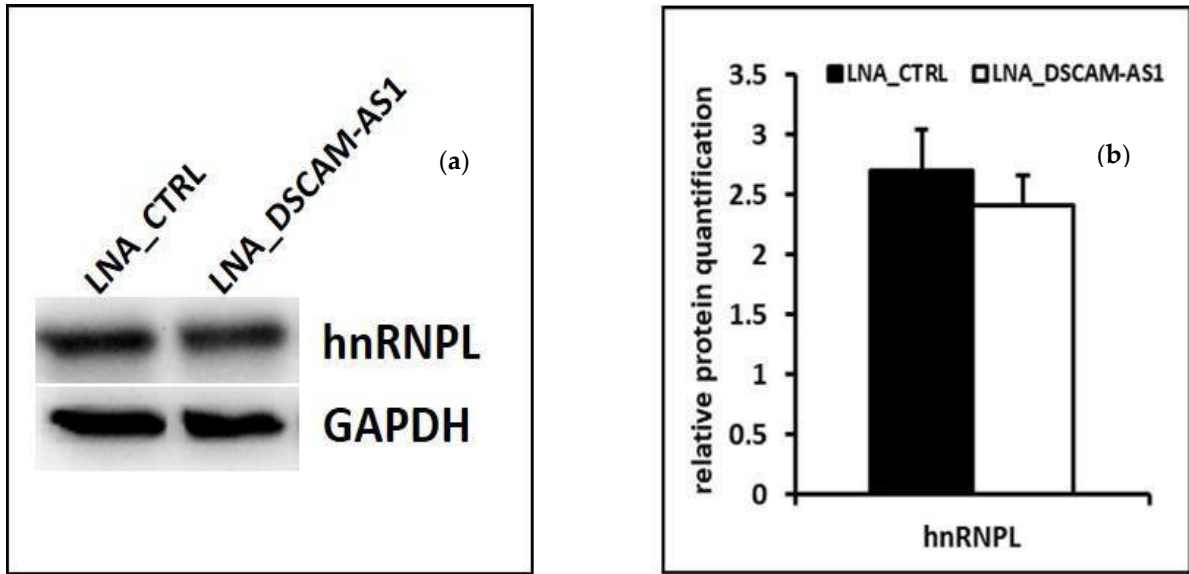
**Figure S4.** (a) Volcano plot showing the log<sub>2</sub>FC of gene expression and the statistical significance of the differential expression (DE) analysis at isoform level performed between MCF-7 cells transfected with control or *DSCAM-AS1*-targeting LNA GapmeRs. In red are reported the up-regulated isoforms while in blue the down-regulated ones. (b) Venn diagram showing the overlap between DE genes and parent genes of DE isoforms. (c–d) Enriched GO biological processes related to parent genes of downregulated and upregulated isoforms, respectively.



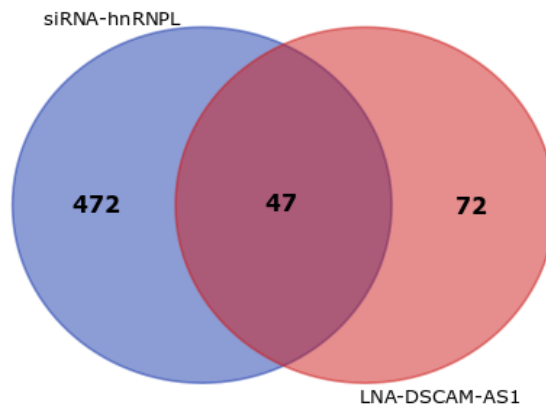
**Figure S5.** Cross-linking ImmunoPrecipitation (CLIP) of hnRNPL in MCF-7 cells shows evidence for physical interaction with *DSCAM-AS1* transcripts, considering separately the nuclear and the cytoplasmic isoforms. ERα, GAPDH and 18S were used as the negative control. Error bars represent the standard error of three biological replicates. Significance from T-test (*DSCAM-AS1* versus 18S enrichment): \*,  $p$ -value < 0.05.



**Figure S6:** The list of RBPs predicted to have an enrichment of their binding motifs in the 3'UTR lengthening events upon *DSCAM-AS1* silencing. The enrichment is shown for a selected region upstream and downstream of proximal (a) and distal (b) APA sites, respectively. Significance: z-score > 1.96.

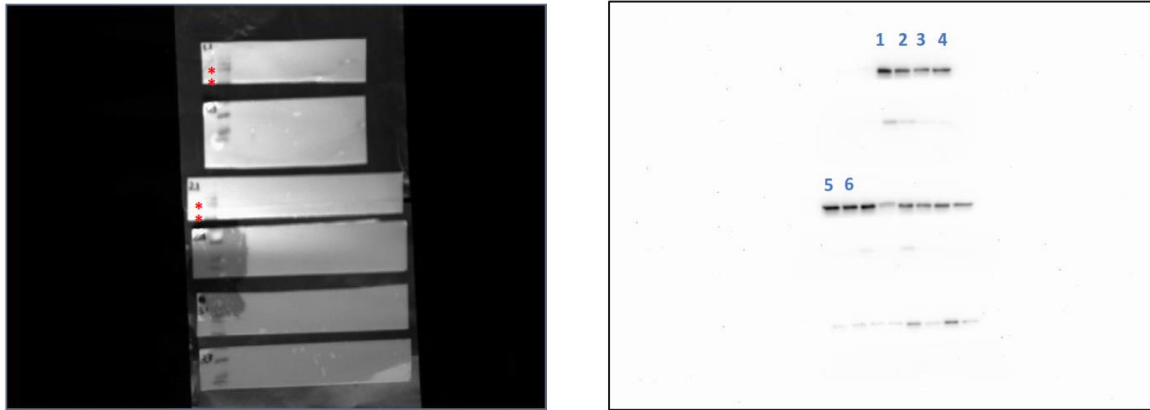


**Figure S7.** (a) Representative image of Western blot analysis of hnRNPL protein, GAPDH: loading control. (b) The histogram shows the protein level of hnRNPL in MCF-7 cells transfected with *DSCAM-AS1* or control LNA. hnRNPL values are relative to GAPDH. Error bars represent the standard error of three biological replicates.

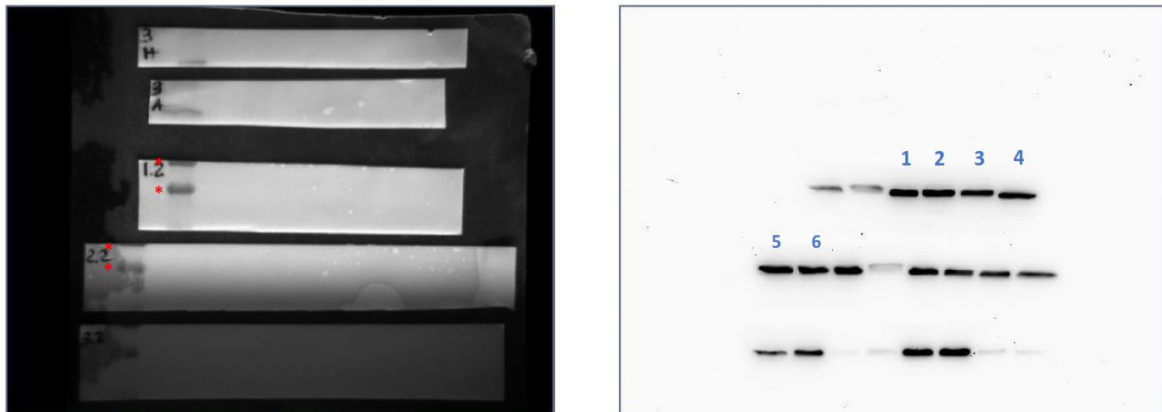


**Figure S8.** Venn diagram showing the number of overlapping exons significantly regulated in our study and upon siRNA-mediated silencing of hnRNPL in LNCaP prostate cancer cells [15].

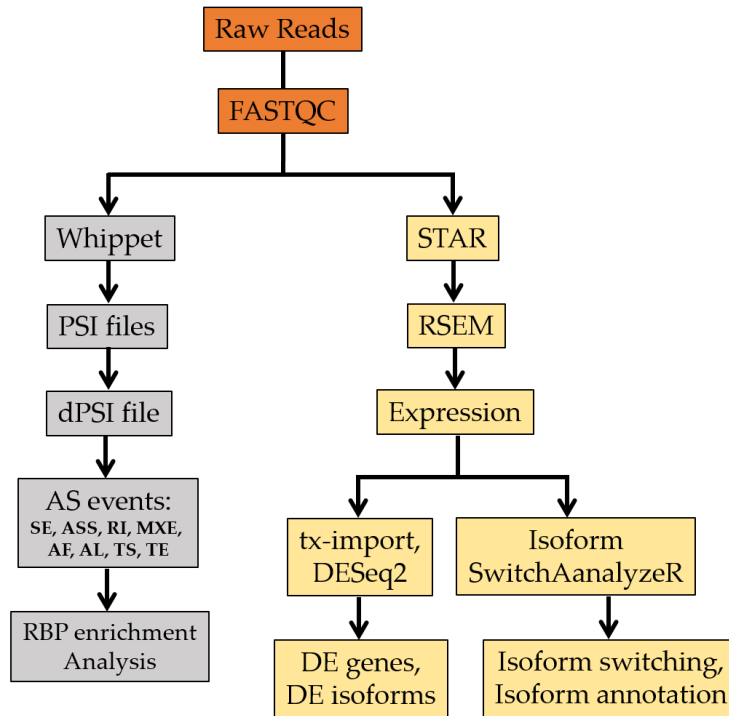
A)



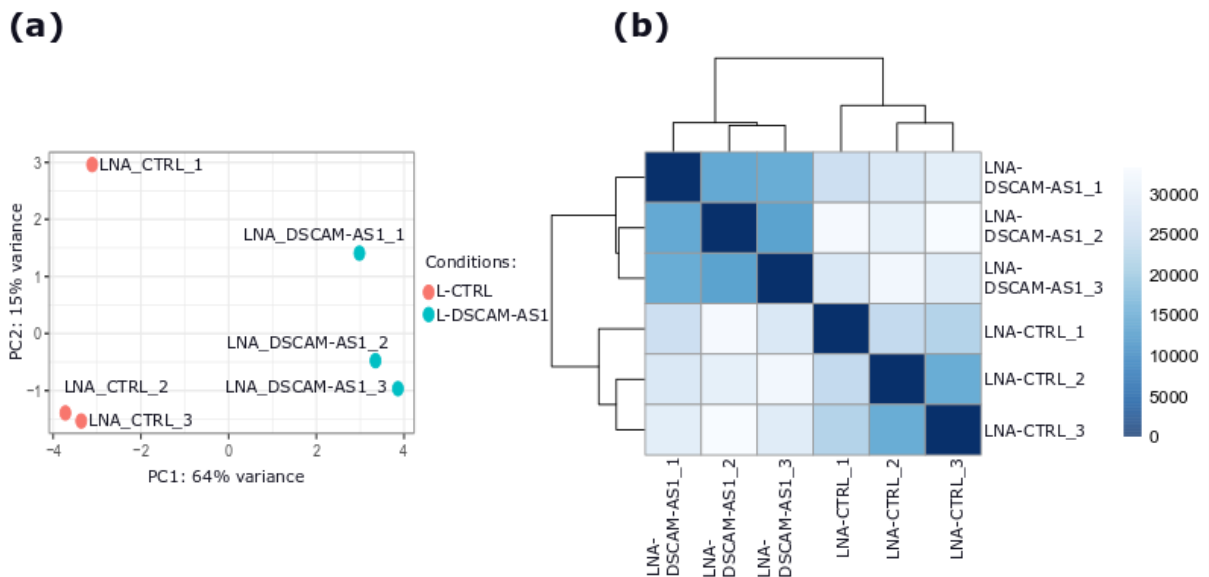
B)



**Figure 9.** Uncropped western blot image relative to the results showed in Figure S7. **(A)** hnRNPL (Left) Uncropped image of markers. Red stars indicate the bands of the molecular marker at 75 KDa (upper) and 50 KDa (lower). (Right) Uncropped image of hnRNPL. Blue numbers indicate the analyzed bands for hnRNPL protein quantification (predicted MW 68 KDa). **(B)** GAPDH. (Left) Uncropped image of markers. Red stars indicate the bands of the molecular marker at 50 KDa (upper) and 37 KDa (lower). (Right) Uncropped image of GAPDH. Blue numbers indicate the analyzed bands for GAPDH protein quantification (predicted MW 37 KDa). Samples' Legend: 1 = LNA\_CTRL replicate 1. 2 = LNA\_DSCAM-AS1 replicate 1. 3 = LNA\_CTRL replicate 2. 4 = LNA\_DSCAM-AS1 replicate 2. 5 = LNA\_CTRL replicate3. 6 = LNA\_DSCAM-AS1 replicate 3.



**Figure S10.** Flow chart showing the different bioinformatic tools and steps used in this study. The FASTQC utility was used for quality check of raw RNA-seq reads. Two pipelines were then applied: (i) shown on the left is the pipeline used for the analysis of alternative splicing changes upon DSCAM-AS1 silencing, using whippet and (ii) on the right is shown the pipeline and tools used for performing differential expression analysis at both gene and isoform levels, as well as isoform switching analysis.



**Figure S11.** Quality control check of the RNA-seq dataset. (a), a PCA plot showing the separation of replicates based on the gene normalized read counts. (b), heat map reporting the dissimilarity matrix between replicates using the gene normalized read counts.