Peer-reviewed version available at *Cancers* **2020**, *12*, 1453; doi:10.3390/cancers12061453

Article

DSCAM-AS1-driven proliferation of breast cancer cells involves regulation of alternative exon splicing and 3′-end usage.

Elhasnaoui J[1,2,†], Miano V[1,2,3,†], Ferrero G[1,2,4], Doria E[2], Leon AE[5], Fabricio ASC[5], Annaratone L[6], Castellano I[6], Sapino A[6,7], and De Bortoli M[1,2,*]

[1] Center for Molecular Systems Biology, University of Turin, Orbassano, 10043 Turin, Italy; jamal.elhasanoui@unito.it (J.E.); valentina.miano@unito.it (V.M.); giulio.ferrero@unito.it (G.F.);
[2] Department of Clinical and Biological Sciences, University of Turin, Orbassano, 10043 Turin, Italy;
[3] Division of Cellular and Molecular Pathology, Department of Pathology, University of Cambridge, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK;
[4] Department of Computer Science, University of Turin, 10149 Turin, Italy.
[5] Regional Center for Biomarkers, Department of Clinical Pathology, Azienda ULSS 3 Serenissima, Campo SS Giovanni e Paolo 6777, 30122 Venice, Italy.
[6] Department of Medical Sciences, University of Turin, Turin, Italy.
[7] Candiolo Cancer Institute, FPO-IRCCS, Candiolo, Italy.
* Correspondence: michele.debortoli@unito.it; Tel.: +39-0116-7050-58.
† These authors contributed equally to this work.

Received: date; Accepted: date; Published: date

Abstract: Background: *DSCAM-AS1* is a cancer-related long noncoding RNA with higher expression levels in Luminal A, B and HER2-positive Breast Cancer (BC), where its expression is strongly dependent on Estrogen Receptor Alpha (ERα).
Methods: To decipher its function, *DSCAM-AS1* expression was measured by qRT-PCR in tissue samples from 93 BC patients in addition to a meta-analysis of 30 gene expression datasets, together with the evaluation of its association with clinical data. By computational analyses of our RNA-Seq in MCF-7 cells, we investigated the *DSCAM-AS1* knock-down effects at both gene and isoform levels. Results: We confirmed *DSCAM-AS1* overexpression in high grade Luminal A, B and HER2+ BCs and found a significant correlation with disease relapse. 908 genes were regulated by *DSCAM-AS1*-silencing, primarily involved in cell cycle and inflammatory response. Noteworthy, the analysis of alternative splicing and isoform regulation revealed 2,085 splicing events regulated by *DSCAM-AS1*, enriched in differential polyadenylation sites and 3′UTR shortening events. Finally, the *DSCAM-AS1*-interacting splicing factor hnRNPL was predicted as the most enriched RBP for exon skipping and 3′UTR events. Conclusion: 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]. Many evidence on the oncogenic or oncosuppressor activities of non-coding RNA transcripts longer than 200 bp, defined as long noncoding RNAs (lncRNAs), were accumulated in the last years thanks to the diffusion of deep-sequencing technologies [2,3]. In the context of Breast Carcinoma (BC), lncRNAs 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 (ERα)-positive BC (Luminal A and B subtypes) represents the most frequent

Peer-reviewed version available at Cancers 2020, 12, 1453; doi:10.3390/cancers12061453

breast neoplasm with over 270,000 estimated new cases in the US population for 2020 [5]. Despite these tumors being characterized by 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 ERα regulated lncRNAs in the luminal BC model MCF-7 performed by our group, evidenced *DSCAM-AS1* as the most significantly ERα-regulated lncRNA in these cells [9]. In the same study, *DSCAM-AS1* gene was evidenced as overexpressed in ERα-positive tumors, particularly of the luminal B subtype. Furthermore, we showed that siRNA-mediated silencing of *DSCAM-AS1* induced a reduction of MCF-7 proliferation with an increase of 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]. Coherently with these results, our group showed that in BC cells grown in hormone-deprived medium ERα binding occurs in a super-enhancer region upstream of *DSCAM-AS1* locus which promotes the lncRNA overexpression in these cells [13]. These results, in addition to the evidence of *DSCAM-AS1* overexpression in poor outcome BC patients [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 the mRNA stability by the regulation of 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 composition and length 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 polyadenylation sites (APA), or may be the consequence of alternative splicing (AS) of internal exons [20]. AS is a regulatory mechanism which allows the fine-tuning of gene isoforms 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 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 ERα-positive BC patients. Then, to functionally investigate the activity of this lncRNA, we analysed the effects of *DSCAM-AS1* down-regulation by RNA-seq followed by gene- and isoform-level analyses. The knock-down of *DSCAM-AS1* strongly hampers cell growth and proliferation pathways in MCF-7 cells. Interestingly, *DSCAM-AS1* knock-down 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*

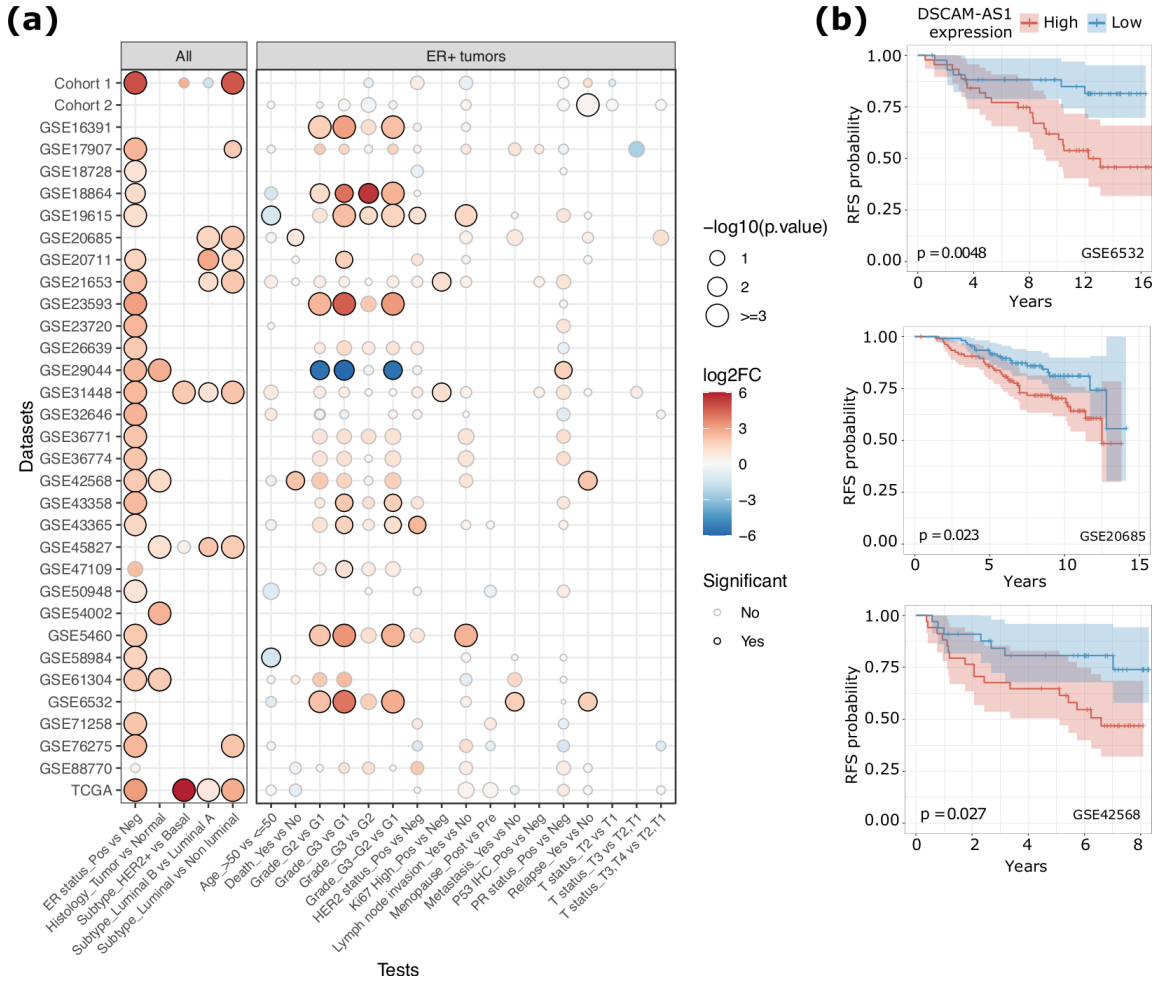
Peer-reviewed version available at Cancers 2020, 12, 1453; doi:10.3390/cancers12061453

in MCF-7 cells and shed light on the importance and diverse roles of this lncRNA in regulating isoforms expression.

2. Results

2.1. DSCAM-AS1 is overexpressed in more aggressive ERa positive BCs

To investigate the expression of *DSCAM-AS1* in tumor tissues, we analysed gene expression data from 30 public microarray datasets (**Supplementary Table 1a**) and RNA-Seq data from The Cancer Genome Atlas (TCGA). In addition to these public data, we quantified the *DSCAM-AS1* expression by qRT-PCR 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 **Supplementary Table 1b-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 log2FC ER+ vs ER- = 2.07) (**Figure 1a** and **Supplementary Table 1d**). This difference was statistically significant in 28 out of the 30 microarray datasets analysed (p-value < 0.05), in the TCGA RNA-Seq dataset (p-value < 2.2E-16) and in Cohort 1 (p-value = 8.01E-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+. Coherently with this result, *DSCAM-AS1* was observed as recurrently up-regulated in luminal BCs compared to non-luminal BCs (average log2FC = 2.37), with higher expression in luminal B tumors as compared to luminal A BCs (average log2FC = 1.20) as previously reported [9,12]. Furthermore, in non-luminal BCs, *DSCAM-AS1* was overexpressed in HER2+ BCs (**Figure 1a** and **Supplementary Figure 1a**).



Peer-reviewed version available at Cancers 2020, 12, 1453; doi:10.3390/cancers12061453

Figure 1. (a) Dot plot reporting the level of statistical significance of the differential *DSCAM-AS1* expression analyses between groups of 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 log2FC 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. ER, Estrogen Receptor; Pos, positive; Neg, negative; PR, Progesterone Receptor. (**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.

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* resulted as significantly overexpressed in BC with lymph node invasion (two datasets), high Ki67 level (two datasets), HER2 overexpression (two datasets) and 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 an 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 10 microarray datasets and in the BRCA TCGA cohort (**Supplementary Table 1e**). As reported in **Figure 1b**, the *DSCAM-AS1* overexpression was significantly associated with 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. A significant difference in relapse rate was observed also among subjects from the 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) (**Supplementary Figure 1b**).

2.2. DSCAM-AS1 knock-down induces a down-regulation 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 GapmeRs. The analysis of the RNA-Seq data evidenced 908 genes Differentially Expressed (DE, $|\log2FC| > 0.20$, adj. p-value < 0.05) upon *DSCAM-AS1* silencing (**Figure 2a** and **Supplementary Table 2**). Specifically, 420 genes showed a significant decrease in expression while 488 genes were up-regulated. Notably, the differential expression of top 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 down-regulated in MCF-7 by *DSCAM-AS1* silencing using siRNA [11]. Noteworthy, *DSCAM-AS1*-silenced cells showed a lower proliferation rate as compared to control (data not shown), even though *ESR1* expression was not perturbed (**Supplementary Table 2**), confirming previous results [9,10]. We also observed that *DSCAM-AS1*-silencing reduces the proliferation rate of HER2-amplified SK-BR-3 BC cells (**Supplementary Figure 2**), confirming that *DSCAM-AS1* is crucial for the survival of different BC cells .

Functional enrichment analysis of the DE genes showed distinct biological processes related to down-regulated and up-regulated genes. Specifically, down-regulated genes were mainly involved in cell cycle progression, cell growth and proliferation (**Figure 2c)**, while up-regulated genes were mainly enriched in terms related to the inflammatory response evoking type I and type II interferon signalling pathways, autophagy and ER-phagosome pathway, apoptosis, and regulation of cholesterol biosynthesis (**Figure 2d** and **Supplementary Table 3**).

Peer-reviewed version available at Cancers 2020, 12, 1453; doi:10.3390/cancers12061453

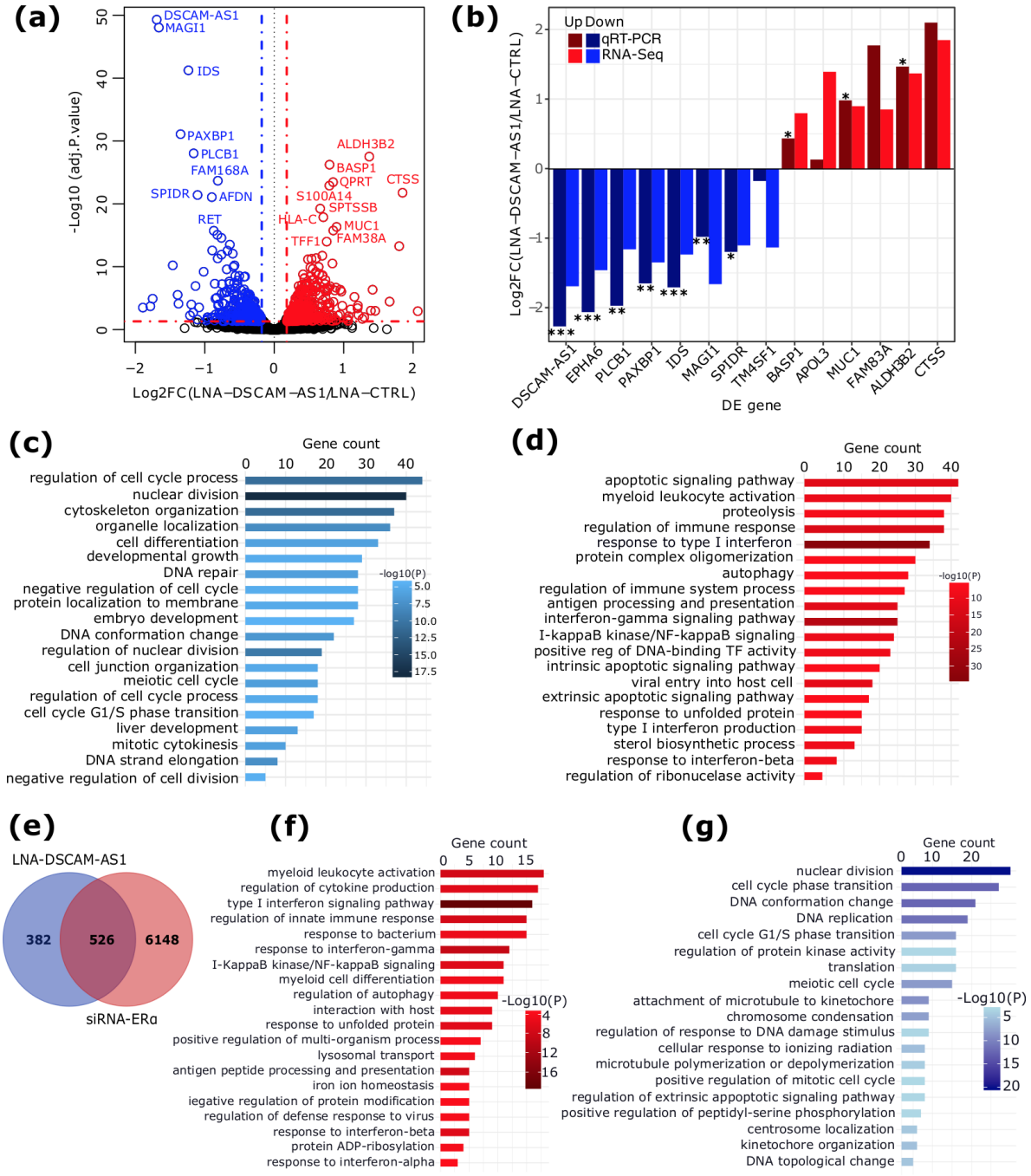


Figure 2. (a) Volcano plot showing the log2FC of gene expression and the statistical significance of the differential expression (DE) analysis performed between MCF-7 cells transfected with control or *DSCAM-AS1*-targeting LNA GapmeRs. In red are reported the up-regulated genes while in blue the down-regulated ones. The top first 20 significant DE genes are labeled. **(b)** Bar plot showing the expression log2FC 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 down-regulated and up-regulated genes, respectively. The number of genes per GO 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 ERα 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.

Peer-reviewed version available at Cancers 2020, 12, 1453; doi:10.3390/cancers12061453

We previously demonstrated that *DSCAM-AS1* is an ERα-regulated lncRNA [9,13]. We thus compared the list of DE genes obtained in this study with those DE upon ERα silencing [13]. Notably, 526 genes were detected as DE in both datasets (hypergeometric test p < 4.36e-48, representation factor = 1.7) (**Figure 2e** and **Supplementary Table 4**). 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 **(Supplementary Figure 3a**). The GO enrichment analysis of the four categories of overlapping genes showed that they are related to distinct biological processes (**Supplementary Table 5**). Notably, genes that showed an up-regulated expression in both datasets are involved in interferon signalling pathways and activation of the immune response as well as autophagy regulation (**Figure 2f**), while genes that were down-regulated in both datasets are involved in the regulation of cell cycle process (**Figure 2g**). Genes upregulated only in this study but downregulated in the siRNA-ERα experiment are related to metabolic processes such as sterol and cholesterol biosynthesis (**Supplementary Figure 3b**). Finally, genes downregulated in this study but upregulated in the siRNA-ERα experiment are related to developmental processes such as gland morphogenesis and neuron projection development pathway (**Supplementary Figure 3c**).

When considering differential expression at the isoform level, the silencing of *DSCAM-AS1* perturbed the expression of 1,035 isoforms of which 439 were down-regulated and 596 were up-regulated **(Supplementary Figure 4a** and **Supplementary Table 6)**. The DE isoforms were transcribed from 898 genes of which 381 (42%) genes were detected as DE in our gene-level analysis (**Supplementary Figure 4b)**. The GO enrichment analysis of the parent genes of these DE isoforms showed enrichment of terms similar to those obtained when considering the overall gene expression changes only **(Supplementary Figure 4c-d** and **Supplementary Table 7).**

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 confirmed this interaction also in our MCF-7 cellular model by hnRNPL Cross-linking ImmunoPrecipitation experiment (CLIP) (Supplementary Figure 5). Therefore, in order to test whether *DSCAM-AS1* silencing could display 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) the analysis of the isoform switching events, also called differential isoform usage, and (ii) the analysis of local AS changes occurring upon *DSCAM-AS1* silencing.

As shown in Figure 3a, using 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 this latter is the sum of the expression of all its isoforms. Next, changes in isoform fraction (dIF) upon *DSCAM-AS1* silencing were measured as $dIF = IF_{silencing} - IF_{control,}$ and significant changes were defined as isoform switching events (details are given in Materials and Methods section 4.6).

Peer-reviewed version available at Cancers 2020, 12, 1453; doi:10.3390/cancers12061453

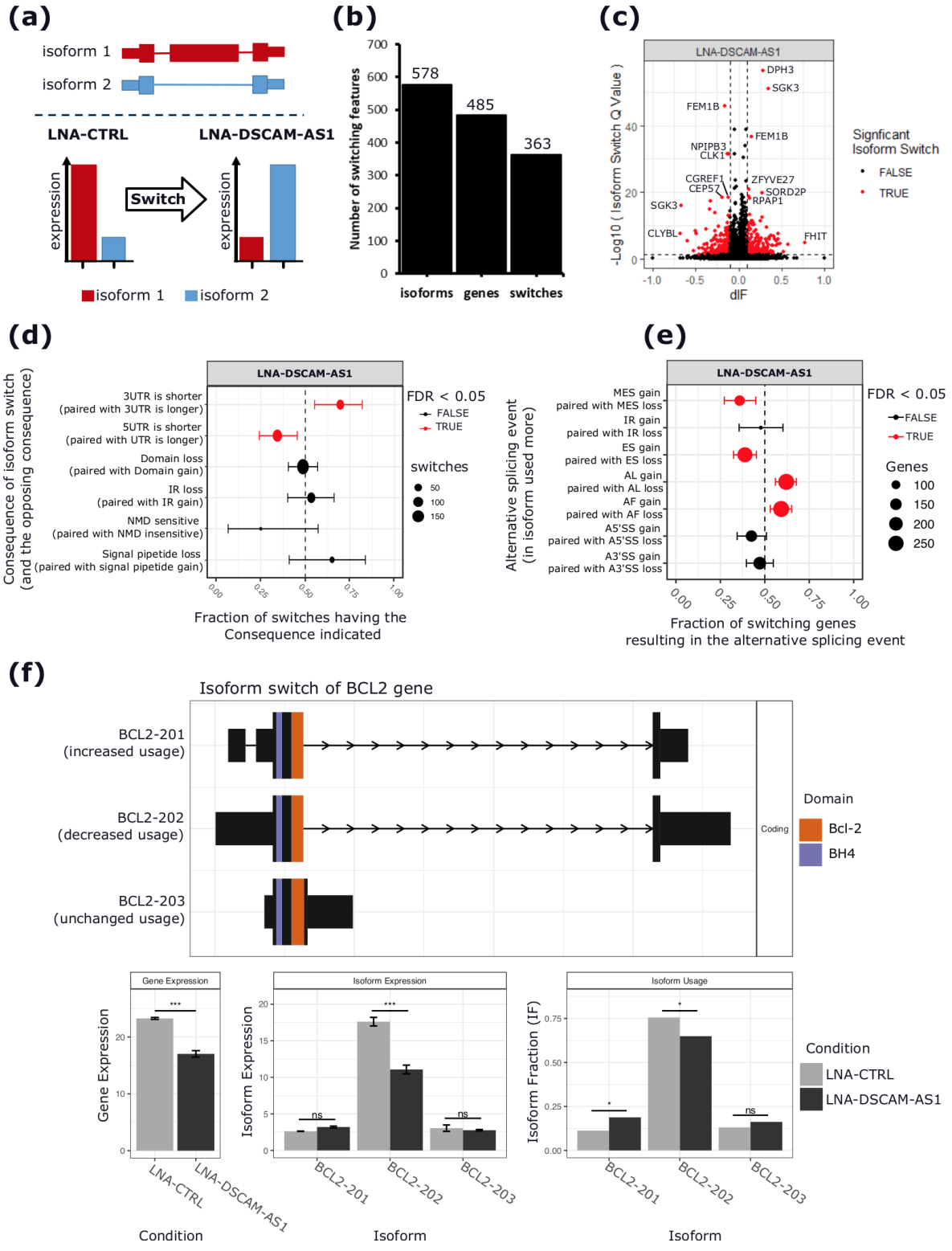


Figure 3. Isoform switching analysis. **(a)** Schematic overview of isoform switching concept used by IsoformSwitchAnalyzeR tool [27]. An isoform switching event is defined as a case where the relative contribution of the isoforms to the parent gene expression changes significantly between conditions. **(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 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

Peer-reviewed version available at Cancers 2020, 12, 1453; doi:10.3390/cancers12061453

means enriched upon *DSCAM-AS1* silencing. **(e)** AS events' 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 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 expression of *BCL2* gene, *BCL2* isoforms and their usage upon *DSCAM-AS1* silencing. **MES**, Multiple Exon Skipping; **IR**, Intron Retention; **ES**, Exon Skipping; **AL**, Alternative Last; **AF**, Alternative First; **A5'SS**, Alternative 5' Splicing Site; **A3'SS**, Alternative 3' Splicing Site.

As shown in **Figure 3b-c,** the isoform switching analysis revealed 485 genes showing 363 significant switching events involving 578 isoforms (**Supplementary Table 8a**). Interestingly, 254 genes (60%) have an isoform switching event with downstream consequences affecting the functional properties of 320 isoforms (**Supplementary Table 8a**). Notably, the annotation of the isoforms upregulated upon *DSCAM-AS1* silencing revealed an enrichment of 5'UTR lengthening and 3'UTR shortening events **(Figure 3d)**. Coherently, the analysis revealed that the switching events involved prevalently AF and AL compared to multiple or single exon skipping events (**Figure 3e**). The switching genes were involved in distinct pathways including 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 tightly regulated by hnRNPL [16], was characterized by a 3'UTR shortening upon *DSCAM-AS1* downregulation (**Figure 3f)**. The list of switching events affecting the 3'UTR length is reported in **Supplementary Table 8b**.

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 4a)**. The analysis identified 2,085 significant AS events (Posterior probability $P > 0.9$ and $|\Delta PSI| > 0.1$), differentially regulated upon *DSCAM-AS1* silencing, affecting the splicing pattern of 1,339 genes **(Figure 4b** and **Supplementary Table 9)**. Interestingly, the most frequent AS event in our data was tandem Alternative PolyA (APA) sites (1,326 events), resulting in shortening of isoforms from 497 genes **(Figure 4b)**. The biological processes involving the genes affected by AS events were mainly related to cell cycle progression, centrosome regulation, regulation of metabolism, and ncRNA processing (**Figure 4c** and **Supplementary Table 10**). Interestingly, 48 out of 76 genes related to the cell cycle progression process had an APA event, followed by 15 genes with a different alternative Transcriptional Start Sites (TS).

2.4. hnRNPL-binding motif is enriched at transcripts regulated by AS events upon 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 hnRNPL binding motif **(Figure 4d, 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 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 4d, right panel)**. Interestingly, the enrichment analysis revealed a significantly higher presence of 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=5.716e-08) (**Figure 4e**). Among the 3'UTR shortening events, hnRNPL motif was observed to be enriched in regions upstream both of the proximal or of the distal polyA site (**Figure 4f-g**). The list of RBPs predicted to be enriched for the 3'UTR lengthening events is shown in

Peer-reviewed version available at Cancers 2020, 12, 1453; doi:10.3390/cancers12061453

Supplementary Figure 6 and the full list of enriched RBPs enriched for different AS events is reported in **Supplementary table 11**.

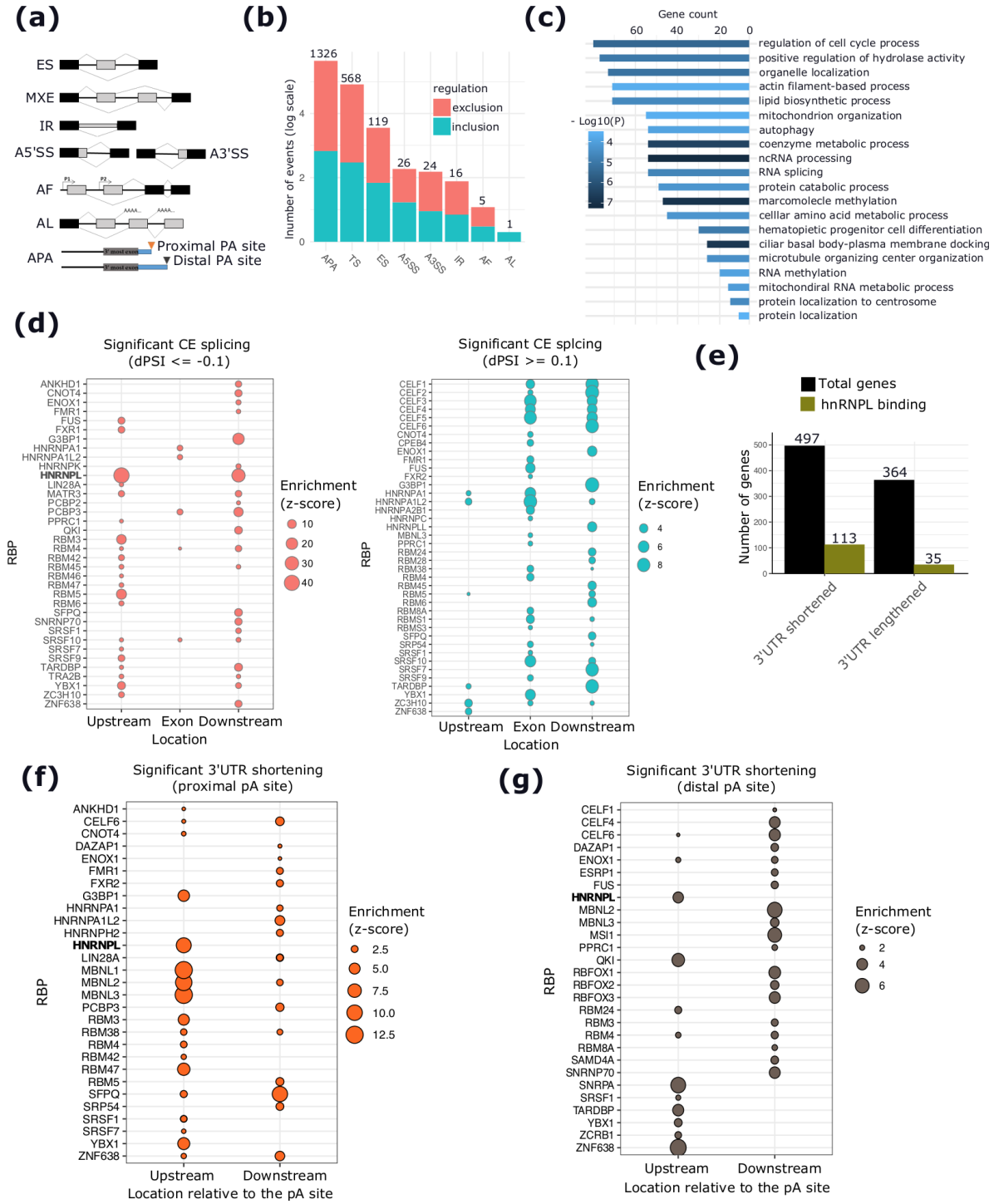


Figure 4. Alternative splicing (AS) changes upon DSCAM-AS1 silencing and prediction of binding motifs for RNA-binding proteins (RBPs). **(a)** Schematic depiction of the different classes of AS events analyzed using Whippet. **ES**, Exon Skipping; **MXE**, Mutually Exclusive Exons; **IR**, Intron Retention; **ASS**, Alternative 5′ (left) and 3′ (right) Splice Sites; **AF**, Alternative First Exons; **AL**, Alternative Last Exons; **APA**, Alternative PolyA site; **TS**, tandem Transcription Starts. **(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)** RBPs binding motif enrichment results for significant **ES** 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

Peer-reviewed version available at Cancers 2020, 12, 1453; doi:10.3390/cancers12061453

those having a predicted hnRNPL binding motif. **(f-g),** RBPs 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.

3. Discussion

In this study, by analysing *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 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 ERα positive BCs of the luminal subtype, as previously showed [9-10]. Coherently with the results from Niknafs and colleagues [10], in 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 of *DSCAM-AS1* expression in drug-resistant BCs and cell lines, including those resistant to Tamoxifen, a selective ERα 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, a clear evidence of the role of *DSCAM-AS1* as a reliable prognostic marker of BC relapse or drug resistance requires further investigation on 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 ERα-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 recruiting of different transcriptional regulators including the Transcription Factor AP2-γ (AP2-γ) 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 choose to use 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 saturation of silencing machineries, as like the siRNA/AGO complex [33]. Strikingly, by comparing the gene expression changes occurring upon *DSCAM-AS1* silencing to those occurring upon ERα silencing [34], both *DSCAM-AS1* and ERα 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 ERα 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].

The expression of many cell cycle related genes was strongly hampered by *DSCAM-AS1* silencing, including *MYC* (MYC proto-oncogene, bHLH transcription factor), a key regulator of cell growth, proliferation, and apoptosis [36], *RET* (ret proto-oncogene), a well known proto-oncogene

Peer-reviewed version available at Cancers 2020, 12, 1453; doi:10.3390/cancers12061453

which regulates cell proliferation and survival and a direct target of ERα signaling pathway [37], *TOP2A* (Topoisomerase II alpha), a proliferation marker whose higher expression in BC is associated with higher tumor grade and 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 minichromosome 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 analysed 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 **(Supplementary Figure 4b)**. 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 which 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 different AS events. Noteworthy, all the RBPs predicted to be enriched are highly expressed in our dataset. Specifically, 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. 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 CUGBP Elav-like family member 1 (CELF) and Splicing Regulator (SR) proteins, were predicted to be enriched in the case of exon inclusion events. 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 [40], Ras GTPase-activating protein-binding protein 1 (G3BP1), an essential splicing factor for the normal stress granules assembly and consequently the preservation of polyadenylated mRNAs [41], Splicing Factor proline/glutamine rich (SFPQ) known to facilitate miRNA-target binding [41,42], 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 [43].

DSCAM-AS1 silencing had a strong effect on alternative polyadenylation sites 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 [44], and in the latter it could involve both oncogenes and tumor suppressor genes [45–47]. Our findings of 3'UTR shortening events are in line with a study by Wang and colleagues where they show an enrichment of shortened 3'UTRs in samples derived from BC patients and characterized by a low proliferation rate [48]. Interestingly, in our data, 24 different genes are characterized by a 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* (**Supplementary Table 12**). 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]. In our data, the shortening of BCL2 isoform, as well as the

Peer-reviewed version available at Cancers 2020, 12, 1453; doi:10.3390/cancers12061453

gene down-regulation, is coherent with an impairment of hnRNPL-mediated stability on it which could be directly mediated by the interaction with *DSCAM-AS1*. Furthermore, in the study of Xue and colleagues, *CCNE1* 3'UTR lengthening was observed to be recurrently detected in six cancer types including BC [45]. However to our knowledge, no evidence of the interaction between *CCNE1* 3'UTR and hnRNPL were 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 a dedicated experimental validation. This is due to the heterogeneous and connected molecular pathways that can be affected by such AS event including a widespread alteration in the network of microRNA-target interactions [47], inhibition or induction of the NMD pathway [49], or generation of novel RNA fragments [50].

The hypothesis of a *DSCAM-AS1*-mediated regulation of hnRNPL interaction at gene 3'UTR is supported by previous evidence of lncRNA-RBP interaction in the post-transcriptional regulation of mRNAs, including the linc-RoR mediated increase of the mRNA stability of c-Myc gene by interaction with hnRNPI and AUF1 at gene 3'UTR [51,52]. Notworthly, while hnRNPL expression was not affected upon *DSCAM-AS1* silencing **(Supplementary Figure 7)**, hnRNPL binding motif resulted among the top enriched in the case of exon skipping and alternative polyadenylation sites events, suggesting that the observed AS changes might be caused by the disruption of the hnRNPL-*DSCAM-AS1* interaction upon the lncRNA-silencing.

Finally, we fully exploited our RNA-Seq dataset potential and analysed 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.

4.1.1. Cell culture and LNA GapmeR™ transfection

MCF-7 and SK-BR-3 cells were routinely grown in DMEM (Life Technologies, 31053–028) supplemented with 10% heat-inactivated FBS (Euroclone, ECS0180L) and 2 mM L-glutamine (ThermoFisher Scientific, 25030–024). Batches of human cell lines were purchased from ATCC. Cells were cultured at 37°C with 5% CO2. MCF-7 and SK-BR-3 cells were detached and transfected in suspension with LNAs (20 nM final concentration) using Lipofectamine3000 (ThermoFisher Scientific, L3000015), according to the manufacturer's protocol. Custom-designed LNA GapmeR™ from Exiqon were used to target all the DSCAM-AS1 lncRNA 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). Experiments were performed 48 hours after LNA transfection.

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), according to the manufacture's protocol. Total RNA used for downstream RNA-seq analysis was isolated from MCF-7 cells using the RNeasy Mini Kit, according to the manufacture's protocol (Qiagen, 74104). 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°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). qRT-PCR analysis was performed using the

Peer-reviewed version available at Cancers 2020, 12, 1453; doi:10.3390/cancers12061453

SYBR-green method (SensiFAST SYBR® Hi-ROX Kit, Bioline, BIO-92005). Real-time PCR primers were custom-designed or purchased from Qiagen (249900 QuantiTect Primer Assay) and are listed in Supplementary Table 13. 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 were analysed 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, Authorisation no. 9/2014 – General Authorisation 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 Supplementary Table 1b-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

2.5 million MCF-7 cells were seeded in 15cm diameter dishes and cultured for 48h before the crosslinking with 0,37% Formaldehyde (Merck-Millipore, F8775) for 10 minutes at 37°C. The crosslinking reaction was stopped with 125mM Glycine (Biorad, 1610717) for 5 minutes at room-temperature. Cells were then washed twice in 1X PBS supplemented with 1% proteases inhibitor cocktail (Merck-Millipore, P2714) and PMSF (Merck-Millipore, 93482), collected by scraping in 1ml of 1X PBS, pelleted by 5 minutes centrifugation 4,000 rpm at 4°C and lysed for 30 minutes 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% proteases inhibitor cocktail (Merck-Millipore, P2714), PMSF (Merck-Millipore, 93482) and RNase-inhibitor (RNaseOUT™ Recombinant Ribonuclease Inhibitor, ThermoFisher Scientific, 10777019). Total extracts were sonicated for 3 cycles (20" ON, 30" OFF) by using an immersion sonicator device. After spinning for 10 minutes, 13,000rpm at 4°C MCF-7 extracts were incubated for 2 hours with 30µl of BSA-coated beads (Protein G Sepharose 4 Fast Flow, GE Healthcare 17061801) for sample pre-clearing. The 5% of total volume was collected as 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) or of two different antibodies anti-hnRNPL (anti-hnRNPL 4D11 and D-5, Santa-cruz Biotechnology, sc-32317 and sc-48391) on a rotating platform. IP-samples were then incubated for 2 hours with 50µl of BSA-coated beads (Protein G Sepharose 4 Fast Flow, GE Healthcare 17061801). IP-samples were then washed 5 times in the RIPA lysis buffer. After wash steps, dried beads and input samples were resuspended in RIPA lysis buffer without SDS and de-crosslinked at 65°C for 1 hour at 650rpm by adding Proteinase K (Proteinase K Solution 20 mg/mL, ThermoFisher Scientific, AM2546). Immuno-precipitated RNA was extracted

Peer-reviewed version available at Cancers 2020, 12, 1453; doi:10.3390/cancers12061453

with PureZOL™ reagent (BioRad, 7326890), according to the manufacture's protocol, cDNA was synthetized and analysed by qRT-PCR as described above. hnRNPL-CLIP enrichment was normalized on input samples and expressed as enrichment of specific binding over the control nonspecific IgG binding.

4.1.5. Cell Proliferation Assay by Crystal-violet staining

SK-BR-3 cells were transfected with LNA-CTR or LNA-DSCAM-AS1 as described and $5*10^5$ cells/well were seeded in a 96MW microtiter plate in 5 technical replicates for each experimental condition. After 48 hours from transfection the culture medium was removed, and cells were washed twice in 1X PBS and fixed with methanol for 5 minutes. The staining was carried out with a 1% Crystal Violet solution for 10 minutes at room temperature (Merck-Millipore, V5265). Several washes with dH_2O were performed to remove excess of 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 50µg of total protein extract were loaded into an 8% Acrylamide gel. Antibodies used were designed against hnRNPL (Santa-cruz Biotechnology, sc-48391) and GAPDH (Santa-cruz Biotechnology, sc-32233). Densitometry readings were measured by the Image Lab Software (Biorad, version number 6.0.1.) and intensity ratio calculation are reports in **Supplementary Table 14.**

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 (**Supplementary Table 1a**) 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+ tumors subtype or associated with the information of the ER status were selected in order to perform separately 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 [53] in default settings. The analysis was performed only for clinical data associated with at least 3 samples per class analysed. *shinyGEO* was also used to perform the survival analyses on datasets associated with information on patient overall survival or relapse-free survival.

Analysis of *DSCAM-AS1* expression in TCGA datasets was performed by retrieving the processed RNA-Seq datasets from GDC data portal [54]. Specifically, the FPKM expression levels of 1,040 subjects were retrieved from this portal. *DSCAM-AS1* expression was evaluated with respect to subjects' clinical covariates. Clinical data were obtained from cBioPortal [55] 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. 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 data analysis:

A flow chart of the mRNA sequencing data analysis, including pre-processing and quality control, quantification at both gene and isoform levels, identification of differentially expressed genes and isoforms, identification of differentially regulated alternative splicing (AS) events is summarized in the **Figure S1** and **Table S1** in **Supplementary Methods**.

Raw reads were assessed for Phred quality scores using 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 bass

Peer-reviewed version available at Cancers 2020, 12, 1453; doi:10.3390/cancers12061453

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 [56]. STAR was run in two-pass mode allowing alignment to the transcriptome coordinates by setting the option --quantMode to TranscriptomeSAM; a summary statistics of read alignment per sample are given in **Table S2 Supplementary Methods**. Next, the expression levels in read counts, TPM, and FPKM units were then estimated at both genes and isoforms levels by running RSEM [57] on the alignment files in default parameters.

4.4. Differential expression analysis

Differentially expressed genes and isoforms upon DSCAM-AS1 silencing as compared to the control condition were identified using the DESeq2 R package (v1.26.0) in default parameters [58]. The expression at isoform level was summarized to gene-level using the *tx-import* bioconductor package [59] 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 (3 out of 6 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. All the data visualization plots including heat maps and volcano plots, MA plots were made using ggplot2 R package (v.3.2.1) [60]. A quality control check of the replicates is given in **Figure S2** in **Supplementary Material and Methods** file.

4.5. Gene Ontology Enrichment analysis

Gene Ontology terms enriched for up-regulated and down-regulated genes were obtained using the Gene Annotation and Analysis Resource Metascape program [61]. The list of up-regulated and down-regulated genes were analysed separately and using the Single List Analysis option. The statistically enriched GO terms related to each category of genes were obtained from GO Biological Processes branch. Only GO terms that were associated with an enrichment factor > 1.5 and an accumulative hypergeometric test adj. 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 similarities and each group or cluster was represented by the top 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 isoforms expression levels quantified in TPM (Transcript per million fragments mapped) 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 criteria was applied by selecting only those isoforms for which DSCAM-AS1 silencing induced a significant change (BH-corrected p-value <=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 [62], CPC2 [63] and Pfam [64] tools, respectively. The biological consequences of the observed switches, including intron retention, domain gain/loss, coding/non-coding potential, 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.

Peer-reviewed version available at Cancers 2020, 12, 1453; doi:10.3390/cancers12061453

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 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 polyadenylation (APA) sites. All the sequences and annotations used in this analysis were based on GRCh38 genome assembly and Gencode v27 annotation. To ensure quantification of expressed events, a prefiltering criteria was applied by only considering those splicing events whose supporting reads are at least 10 in at least two samples per condition. In addition, a splicing event with a ΔPSI value between the silencing and control conditions less than 10% ($|\Delta PSI| < 0.1$), or whose associated posterior probability is 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 ± 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 100 nucleotides on both sides from the APA site position. The RNA binding motifs for 105 different splicing factors collected from RNAcompete study [65] were used to perform binding motif enrichment analysis. For a number of RBPs, the motif from different species was confirmed in a previous study [66] 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 [66]. The tool FIMO (Find Individual Motif Occurrences) [67] 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, 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) [66]. The obtained z-scores per binding motif, region, and event were then visualized using ggplot2 Bioconductor package [60].

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

Peer-reviewed version available at Cancers 2020, 12, 1453; doi:10.3390/cancers12061453

express abnormal, and subtype-specific, variety of circRNAs [68]. 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 www.mdpi.com/xxx/s1, Supplementary Methods: Supplementary materials and method of the manuscript; Supplementary Figures: File reporting all the supplementary figures of the manuscripts. Supplementary Figure 1: Analysis of *DSCAM-AS1* expression in ER-negative BCs and results from survival analysis; Supplementary Figure 2: Effect of DSCAM-AS1 silencing in SK-BR3 cell line proliferation ; Supplementary Figure 3: Ontological analysis of DE genes not coherently expressed between this study and the experiment of ERα silencing; Supplementary Figure 4: Results from differential isoform expression analysis; Supplementary Figure 5: Result of the DSCAM-AS1-hnRNPL CLIP experiment; Supplementary Figure 6: Analysis of hnRNPL protein level upon DSCAM-AS1 silencing. Supplementary Table 1: Datasets information and results from the analysis of *DSCAM-AS1* expression with respect to different clinical data; Supplementary Table 2: Results from the gene-level differential expression analysis of *DSCAM-AS1* LNA experiment; Supplementary Table 3: Results from the gene set enrichment analysis of *DSCAM-AS1* LNA differentially expressed genes; Supplementary Table 4: Comparison of gene differentially expressed in the *DSCAM-AS1* LNA experiment and in MCF-7 cells treated with ER siRNA; Supplementary Table 5: Results from the gene set enrichment analysis of genes overlapped between *DSCAM-AS1* LNA and ER siRNA experiments; Supplementary Table 6: Results from the isoform-level differential expression analysis of *DSCAM-AS1* LNA experiment; Supplementary Table 7: Results from the gene set enrichment analysis of parent genes of *DSCAM-AS1* LNA differentially expressed isoforms; Supplementary Table 8: Isoform features from the isoform switching analysis; Supplementary Table 9: Results from differential AS analysis using Whippet; Supplementary Table 10: Results from the gene set enrichment analysis of genes characterized by differential alternative spliced events upon DSCAM-AS1 silencing. Supplementary Table 11: List of enriched RBPs binding motifs among different AS events; Supplementary Table 12: List of genes with 3'UTR shortening/lengthening events reported by Whippet; Supplementary Table 13: Custom-designed primer sequences and purchased primers catalog numbers; Supplementary Table 14: WB densitometry readings and intensity ratio calculation

Author Contributions: All authors have read and agree to the published version of the manuscript. Conceptualization, JE., GF, VM; bioinformatic analysis, JE; experimental analysis, VM; samples preparation and clinical data collection regarding Cohort_1, LA, IC, AS; samples selection and clinical data collection regarding Cohort_2, AEL, ASCF; writing—original draft preparation, JE, GF, VM, MDB; writing—review and editing, JE, GF, VM, ED, MDB.; visualization, JE and GF; supervision, MDB; project administration, MDB.; funding acquisition, MDB.

Funding: This research was funded by Associazione Italiana per la Ricerca sul Cancro [Grant IG 15600 to MDB]; by Fondazione CRT [grant 2017.0823 to MDB]; by University of Torino [2018 Local Research funding to MDB].

Acknowledgments: We thank Salvatore Oliviero from 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.
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. *Communications Biology* 2020, *3*.
3. Sahu, A.; Singhal, U.; Chinnaiyan, A.M. Long Noncoding RNAs in Cancer: From Function to Translation. *Trends in Cancer* 2015, *1*, 93–109.
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. *Science Advances* 2016, *2*, e1600220.
5. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2020. *CA Cancer J. Clin.* **2020**, *70*, 7–30.
6. Rani, A.; Stebbing, J.; Giamas, G.; Murphy, J. Endocrine Resistance in Hormone Receptor Positive Breast Cancer–From Mechanism to Therapy. *Frontiers in Endocrinology* 2019, *10*.
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? *Clinical Genetics* 2019, *95*, 558–568.

Peer-reviewed version available at Cancers 2020, 12, 1453; doi:10.3390/cancers12061453

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.
9. Miano, V.; Ferrero, G.; Reineri, S.; Caizzi, L.; Annaratone, L.; Ricci, L.; Cutrupi, S.; Castellano, I.; Cordero, F.; De Bortoli, M. Luminal long non-coding RNAs regulated by estrogen receptor alpha in a ligand-independent manner show functional roles in breast cancer. *Oncotarget* **2016**, *7*, 3201–3216.
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.
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. 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.
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.
13. Miano, V.; Ferrero, G.; Rosti, V.; Manitta, E.; Elhasnaoui, J.; Basile, G.; De Bortoli, M. Luminal lncRNAs Regulation by ERα-Controlled Enhancers in a Ligand-Independent Manner in Breast Cancer Cells. *Int. J. Mol. Sci.* **2018**, *19*.
14. Rossbach, O.; Hung, L.-H.; Khrameeva, E.; Schreiner, S.; König, J.; Curk, T.; Zupan, B.; Ule, J.; Gelfand, M.S.; Bindereif, A. Crosslinking-immunoprecipitation (iCLIP) analysis reveals global regulatory roles of hnRNP L. *RNA Biol.* **2014**, *11*, 146–155.
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. U. S. A.* **2017**, *114*, E5207–E5215.
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*.
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.
18. Gaudreau, M.-C.; Grapton, D.; Helness, A.; Vadnais, C.; Fraszczak, J.; Shooshtarizadeh, P.; Wilhelm, B.; Robert, F.; Heyd, F.; Möröy, T. Heterogeneous Nuclear Ribonucleoprotein L is required for the survival and functional integrity of murine hematopoietic stem cells. *Sci. Rep.* **2016**, *6*, 27379.
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.
20. Reyes, A.; Huber, W. Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. *Nucleic Acids Res.* **2018**, *46*, 582–592.
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*, 1689.
22. Matlin, A.J.; Clark, F.; Smith, C.W.J. Understanding alternative splicing: towards a cellular code. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 386–398.
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.H.; et al. Alternative transcription initiation leads to expression of a novel ALK isoform in cancer. *Nature* **2015**, *526*, 453–457.
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.
25. Rajan, P.; Elliott, D.J.; Robson, C.N.; Leung, H.Y. Alternative splicing and biological heterogeneity in prostate cancer. *Nat. Rev. Urol.* **2009**, *6*, 454–460.
26. Sveen, A.; Kilpinen, S.; Ruusulehto, A.; Lothe, R.A.; Skotheim, R.I. Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. *Oncogene* **2016**, *35*, 2413–2427.
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.
28. Sterne-Weiler, T.; Weatheritt, R.J.; 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.
29. Hui, J.; Hung, L.-H.; Heiner, M.; Schreiner, S.; Neumüller, N.; Reither, G.; Haas, S.A.; Bindereif, A. Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. *EMBO J.* **2005**, *24*, 1988–1998.

Peer-reviewed version available at Cancers 2020, 12, 1453; doi:10.3390/cancers12061453

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.* **2019**, *234*, 2880–2894.
31. Hong, S.P.; Chan, T.E.; Lombardo, Y.; Corleone, G.; Rotmensz, N.; Bravaccini, S.; Rocca, A.; Pruneri, G.; McEwen, K.R.; Coombes, R.C.; et al. Single-cell transcriptomics reveals multi-step adaptations to endocrine therapy. *Nat. Commun.* **2019**, *10*, 3840.
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.
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.
34. Caizzi, L.; Ferrero, G.; Cutrupi, S.; Cordero, F.; Ballaré, C.; Miano, V.; Reineri, S.; Ricci, L.; Friard, O.; Testori, A.; et al. Genome-wide activity of unliganded estrogen receptor-α in breast cancer cells. *Proceedings of the National Academy of Sciences* 2014, *111*, 4892–4897.
35. Xu, S.; Kong, D.; Chen, Q.; Ping, Y.; Pang, D. Oncogenic long noncoding RNA landscape in breast cancer. *Mol. Cancer* **2017**, *16*, 129.
36. Xu, J.; Chen, Y.; Olopade, O.I. MYC and Breast Cancer. *Genes & Cancer* 2010, *1*, 629–640.
37. Stine, Z.E.; McGaughey, D.M.; Bessling, S.L.; Li, S.; McCallion, A.S. Steroid hormone modulation of RET through two estrogen responsive enhancers in breast cancer. *Human Molecular Genetics* 2011, *20*, 3746–3756.
38. An, X.; Xu, F.; Luo, R.; Zheng, Q.; Lu, J.; Yang, Y.; Qin, T.; Yuan, Z.; Shi, Y.; Jiang, W.; et al. The prognostic significance of topoisomerase II alpha protein in early stage luminal breast cancer. *BMC Cancer* **2018**, *18*, 331.
39. Burd, C.G.; 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.
40. 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.
41. Aulas, A.; Caron, G.; Gkogkas, C.G.; Mohamed, N.-V.; Destroismaisons, L.; Sonenberg, N.; Leclerc, N.; Parker, J.A.; Vande Velde, C. G3BP1 promotes stress-induced RNA granule interactions to preserve polyadenylated mRNA. *J. Cell Biol.* **2015**, *209*, 73–84.
42. 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.
43. Vanharanta, S.; Marney, C.B.; 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*.
44. Gruber, A.J.; Zavolan, M. Alternative cleavage and polyadenylation in health and disease. *Nat. Rev. Genet.* **2019**, *20*, 599–614.
45. 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 Genomics* **2018**, *19*, 536.
46. Mayr, C.; Bartel, D.P. Widespread shortening of 3′UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **2009**, *138*, 673–684.
47. 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. *Nature Genetics* 2018, *50*, 783–789.
48. 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.* **2015**, *35*, 39–50.
49. Kurosaki, T.; Popp, M.W.; Maquat, L.E. Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nature Reviews Molecular Cell Biology* 2019, *20*, 406–420.
50. 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.

Peer-reviewed version available at Cancers 2020, 12, 1453; doi:10.3390/cancers12061453

51. 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.* **2016**, *44*, 3059–3069.
52. Lu, Y.; Liu, X.; Xie, M.; Liu, M.; Ye, M.; Li, M.; Chen, X.-M.; Li, X.; Zhou, R. The NF-κB–Responsive Long Noncoding RNA FIRRE Regulates Posttranscriptional Regulation of Inflammatory Gene Expression through Interacting with hnRNPU. *The Journal of Immunology* 2017, *199*, 3571–3582.
53. Dumas, J.; Gargano, M.A.; Dancik, G.M. shinyGEO: a web-based application for analyzing gene expression omnibus datasets. *Bioinformatics* **2016**, *32*, 3679–3681.
54. Grossman, R.L.; Heath, A.P.; Ferretti, V.; Varmus, H.E.; Lowy, D.R.; Kibbe, W.A.; Staudt, L.M. Toward a Shared Vision for Cancer Genomic Data. *New England Journal of Medicine* 2016, *375*, 1109–1112.
55. Gao, J.; Aksoy, B.A.; Dogrusoz, U.; Dresdner, G.; Gross, B.; Sumer, S.O.; Sun, Y.; Jacobsen, A.; Sinha, R.; Larsson, E.; et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **2013**, *6*, l1.
56. 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* **2013**, *29*, 15–21.
57. Li, B.; Dewey, C.N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011, *12*.
58. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550.
59. Soneson, C.; Love, M.I.; Robinson, M.D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res.* **2015**, *4*, 1521.
60. Wilkinson, L. ggplot2: Elegant Graphics for Data Analysis by WICKHAM, H. *Biometrics* 2011, *67*, 678–679.
61. Zhou, Y.; Zhou, B.; Pache, L.; Chang, M.; Khodabakhshi, A.H.; Tanaseichuk, O.; Benner, C.; Chanda, S.K. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **2019**, *10*, 1523.
62. Almagro Armenteros, J.J.; Tsirigos, K.D.; 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.
63. 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.
64. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A.; et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2019**, *47*, D427–D432.
65. Ray, D.; Ha, K.C.H.; Nie, K.; Zheng, H.; Hughes, T.R.; Morris, Q.D. RNAcompete methodology and application to determine sequence preferences of unconventional RNA-binding proteins. *Methods* **2017**, *118-119*, 3–15.
66. Sebestyén, E.; Singh, B.; Miñana, B.; Pagès, A.; Mateo, F.; Pujana, M.A.; Valcárcel, J.; Eyras, E. Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *Genome Res.* **2016**, *26*, 732–744.
67. Grant, C.E.; Bailey, T.L.; Noble, W.S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **2011**, *27*, 1017–1018.
68. 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.