Accepted Preprint first posted on 30 May 2018 as Manuscript ERC-18-0068

Splicing dysregulation as a driver of breast cancer

Abigail Read[1,2] and Rachael Natrajan[1,2*]

[1]The Breast Cancer Now Toby Robins Research Centre, The Institute of Cancer Research, London, UK

[2]Division of Molecular Pathology, The Institute of Cancer Research, London, UK

*Correspondence to:

Rachael. Natrajan@icr.ac.uk

Running title: alternative splicing in breast cancer

Abstract

Breast cancer is known to be a heterogeneous disease driven by a large repertoire of molecular abnormalities, which contribute to its diverse clinical behavior. Despite the success of targeted therapy approaches for breast cancer patient management, there is still a lack of the molecular understanding of aggressive forms of the disease and clinical management of these patients remains difficult. The advent of high throughput sequencing technologies, have paved the way for a more complete understanding of the molecular make-up of the breast cancer genome. As such, it is becoming apparent that disruption of canonical splicing within breast cancer governs its clinical progression. In this review, we discuss the role of dysregulation of spliceosomal component genes and associated factors in the progression of breast cancer, their role in therapy resistance and the use of quantitative isoform expression as potential prognostic and predictive biomarkers with a particular focus on oestrogen receptor positive breast cancer.

Introduction

Dysregulation of alternative splicing (AS) is widely considered a new hallmark of cancer and its products are being acknowledged as potentially useful biomarkers (Ladomery 2013). Canonical RNA splicing takes place in all mammalian cells and during this process pre-mRNA becomes mature mRNA via the excision of introns and pasting together of exons (Figure 1). Alternative splicing affects about 90% of human genes resulting in a diverse selection of isoforms from one gene, each having different structural and functional properties that leads to a larger and more diverse cellular proteome. Indeed throughout evolution alternative splicing has been used to propel species development evidenced by an increase in AS in higher eukaryotes compared to lower (Keren, et al. 2010).

Splicing is performed by the spliceosome which is a multi-protein complex called a "metalloribozyme" that is made up of 5 small nuclear riboproteins (snRNPs) that contain snRNAs and a large number of accessory proteins to recognize the pre-mRNA being spliced. Assembly of this complex takes place during transcription suggesting that transcription and splicing machineries are space restricted as they happen closely in time (Herzel, et al. 2017). The most commonly occurring spliceosome is the U2-dependent spliceosome that is assembled from the U1, U2, U5 and U4/U6 snRNPs and is responsible for the splicing of 99% of human introns as reviewed in Dvinge et al. (Dvinge, et al. 2016). Splicing is a two-step reaction involving transesterification occurring between two RNA nucleotides. The spliceosome recognizes introns containing the consecutive nucleotides GU at the 5' splice site (SS) by U1 snRNP binding and an AG sequence at the 3 'splice site by U2AF1 binding. In order to properly position the splicing machinery a key adenine (also referred to as the branch point (BP)) must be recognized by the Splicing Factor 1 protein (SF1) as well as recognition of the polypyrimidine tract (poly-Y) by the U2 small nuclear RNA auxiliary factor 2 (U2AF2) (Pandya-Jones 2011). When the spliceosome complex is correctly bound to the mRNA it can carry out intron excision (Figure 1A). Exons are subsequently joined together and release a lariat intron, which is then degraded. The

spliceosome components are then released and recycled for use in subsequent rounds of splicing. Splicing factors such as serine/arginine rich (SR) proteins (SRSF) and the splicing factor 3b complex (SF3B) work in association with the splicing core complex to coordinate canonical and alternative splicing. The expression levels and binding affinities of the different splicing factors plays a stoichiometric role in determining the final isoform of the protein that is to be expressed (da Luz, et al. 2017).

Dysregulation of the normal splicing process governs many aspects of cancer cell biology such as managing cellular proliferation, angiogenesis, resisting apoptosis, adapting cell metabolism, enhancing the ability to invade and metastasize, and plays a role in resistance to cancer therapy (David and Manley 2010; Lee and Abdel-Wahab 2016). The role of alternative splicing in disease can result from aberrant splicing of a gene due to incorrect 5' or 3' splice site recognition leading to intron retention, exon skipping or exon inclusion (Figure 1A). AS may then lead to a premature stop codon resulting from a frame-shift, whereby these transcripts are subsequently degraded by nonsense mediated decay (NMD) (Figure 1B). There are multiple ways in which a cancer cell can induce aberrant splicing including: 1) when there is a mutation in the exon or surrounding introns that compromises the canonical splicing signal thereby allowing an alternative signal to dominate and an aberrant mRNA to be made; 2) a mutation in one of the splicing regulators interrupts splice site selection and results in a pattern of alternative splicing in multiple genes; 3) changes in histone acetylation of alternative exons (Khan, et al. 2014) and 4) alterations in other RNA-binding proteins, splicing enhancers and suppressors, or lncRNAs. Such splicing errors can lead to alterations in relative isoform expression of a particular mRNA or lead to an aberrant protein that has a change of function. A more detailed discussion on points 1 and 3 are detailed elsewhere (Martinez-Montiel, et al. 2017). Aberrantly spliced apoptotic genes such as the RNA binding protein *RBM5* have been implicated in breast cancers as having an opposing role because the resulting isoform is more anti-apoptotic (Fushimi, et al. 2008). Another example is the B-cell lymphoma gene, *Bcl-x,* which can be spliced into two different

isoforms, long and short. *Bcl-x(L)* has anti-apoptotic properties where as *Bcl-x(s)* has pro-apoptotic properties. High levels of *Bcl-x(L*) are seen in various types of cancer (Boise, et al. 1993; Fushimi et al. 2008; Takehara, et al. 2001). A similar situation is seen with the myeloid cell leukemia-1 gene and its' two isoforms *Mcl-1(s)* and *Mcl-1(L)*. The long isoform is anti-apoptotic and seen frequently increased compared to the short isoform in breast and ovarian cancer cells and is linked to gene amplification of *MCL-1* itself (Bae, et al. 2000; Bingle, et al. 2000; Gautrey and Tyson-Capper 2012). The choice between the long and short isoform is influenced by the splicing factors SRSF1 and SRSF5, which are also frequently upregulated in breast cancer (Gautrey and Tyson-Capper 2012).

Managing key cellular processes such as epithelial to mesenchymal differentiation (EMT) is a clear advantage of being able to manipulate the expression of different isoforms of a certain gene (Shapiro, et al. 2011). As such, acquiring the ability to hijack these processes is critical in the evolution of a cancer cell in order to provide a fitness advantage. Given this, it is reasonable to postulate that the characterisation of the splicing program of a cancer cell could predict its genomic and mutational status and potentially treatment outcome (Danan-Gotthold, et al. 2015). Indeed, differential expression of AS transcripts in specific subtypes of breast cancer may add additional prognostic information in addition to canonical gene expression or protein expression biomarkers.

Evidence of splicing dysregulation in breast cancer

Since the seminal studies from Perou and colleagues describing the intrinsic subtypes of breast cancer (Perou, et al. 2000), it is now widely accepted that the molecular make-up of breast cancer is heterogenous and governed by differences in transcriptional make up. Inevitably this also applies to the degree of isoform usage in cancer cells as well. For instance, well known driver oncogenes and tumour suppressor genes such as *ERBB2* and

BRCA1 are known to be differentially spliced in different subtypes of breast cancer (as reviewed by Martinez-Montiel et al. (Martinez-Montiel, et al. 2017)). BRCA1 which is involved in homologous recombination DNA repair, is alternatively spliced in breast cancer to exclude exon 11 which contains the nuclear localization signal (Thakur, et al. 1997). The Δ11q isoform produces a protein that is absent from the nucleus and is therefore unable to assist in DNA damage repair. Studies have shown that down regulation of the full length nuclear *BRCA1* isoform and overexpression of the cytoplasmic Δ11q isoform is evident in subsets of breast cancer and is potentially mediated through the presence of a non-functional TRA2β splicing factor (Raponi, et al. 2014; Wiener, et al. 2015). Another example is the ERBB2 tyrosine kinase signalling receptor, which is often found as alternatively spliced in breast cancer as the *Δ16HER2* isoform. *Δ16HER2* is constitutively active as a homodimer and promotes transformation in the mammary gland (Marchini, et al. 2011). *BRCA1* and *ERBB2* splicing, as well as splicing of *Bcl-x* and *Mcl-1* as described above, are examples of common driver oncogenes and tumour suppressor genes that can be aberrantly spliced in breast cancer. Alternative splicing has also been show to regulate protein diversity of the oestrogen receptor itself. In particular, previous studies have shown the ERαΔ5 splice variant has a positive effect on activation of transcription in the absence of oestrogen leading to constitutive transcriptional activation (Bollig and Miksicek 2000; Fuqua, et al. 1991). ESR1 aberrant splicing events have also been identified in circulating tumour cells from metastatic breast cancer patients that have progressed on endocrine therapy, suggesting a role in mediating resistance (Beije, et al. 2018). Current data sets describing alternative splicing events in the context of spliceosomal gene mutations however, do not show changes in splicing of the oestrogen receptor itself (Darman, et al. 2015; Maguire, et al. 2015). Alternative spliced isoforms of genes known to be transcriptionally regulated by the oestrogen receptor such as CyclinD1 (cyclin D1b) and FGFR1 (FGFR1-beta) are also associated with poor prognosis in ER+ breast cancer (Wei, et al. 2011; Wendt, et al. 2014).

Alternative spliced transcripts as prognostic and predictive biomarkers in breast cancer

The recent advent of RNA-sequencing technologies has revolutionised our view of the molecular make up of breast cancer. These advances now allow accurate global quantification of the transcriptional isoform make-up in individual tumours rather than relative quantification that is based on microarray probe design. Indeed, a number of studies have shown that alternative isoform usage can be specific to different breast cancer molecular subtypes (Gracio, et al. 2017; Menon, et al. 2014; Sebestyen, et al. 2015; Stricker, et al. 2017; Zhao, et al. 2016). For instance, Sebestyen et al. identified a specific 7 gene isoform signature that accurately identified basal-like breast cancers, including a number of known driver genes such as *CTNND1* (Sebestyen et al. 2015). Analysis of the splicing balance (relative ratios of isoforms produced) in breast tumours revealed changes in isoform usage in oncogenic and tumour suppressive pathways that was not apparent when looking solely at gene expression data (Gracio et al. 2017). Importantly, it was found that the balance of different transcript isoforms was associated with patient prognosis. A subset of genes including the proto-oncogene *MYB* were identified to correlate with basal-like breast cancer patient survival based on varying isoform levels but not on whole gene expression analyses (Gracio et al. 2017). Additionally, splicing but not gene expression levels of immune related genes *CCR7* and *FCRL3* were found to determine the immune control of the tumour. This has potential relevance given the role of lymphocytic infiltration in prognosis in breast cancer. Differential isoform usage can also stratify between different molecular subtypes of breast cancer. Indeed, global dysregulation of splicing specific to individual subtypes may drive the heterogeneous nature of breast cancer due to variation in the cellular proteome. Stricker et al (Stricker et al. 2017) looked at the global isoform differences between ER+ and triple-negative breast cancer (TNBC) and identified a signature of subtype specific alternatively spliced transcripts. Interestingly around 63% of the genes that were found to be differentially expressed, between subtypes were also alternatively spliced. The particular type of splicing that occurred between the subtypes (exon skipping, intron retention,

alternative acceptor or donor) however was not significantly different indicating the unique splicing programs of each intrinsic subtype is not necessarily due to the activity of one general splicing mechanism but more likely due to target gene selection (Stricker et al. 2017). Interestingly, this study also identified a significant difference in the total expression of some spliceosomal component genes themselves, such as *YBX1* and *MAGOH* suggesting dysregulation of spliceosomal component proteins governs splicing dysregulation.

Although clear differences in transcript isoforms have been identified in different molecular subtypes of breast cancer, to date no study has assessed the value of alternatively spliced transcripts as prognostic and predictive clinical biomarkers for patient stratification and of treatment response to both standard chemotherapy and targeted endocrine therapy. Assessment of differences in transcript isoform expression could add much needed biomarkers for patients that are most likely relapse on standard of care therapy. Ideally this would need to be tested in the context of randomised clinical trial cohorts, where good quality RNA-sequencing data at sufficient depth is acquired.

Dysregulation of spliceosomal factors in Breast Cancer

Molecular alterations affecting spliceosomal component genes themselves are also known to be involved in breast cancer tumorigenesis. There is evidence that mutations, copy number alterations and differential expression of spliceosomal component genes and their interacting proteins are associated with specific molecular and histological subtypes of breast cancer as well as being associated with aggressive disease and resistance to therapy in multiple tumour types (Ng, et al. 2012; Siegfried and Karni 2017; Sotillo, et al. 2015; Stark, et al. 2009). These alterations are thought to drive breast cancer progression through specific or novel isoform selectivity of key genes (Anczukow, et al. 2015; da Luz et al. 2017; Gokmen-Polar, et al. 2015; Maguire et al. 2015; Martinez-Montiel et al. 2017; Silipo, et al. 2015; Vanharanta, et al. 2014).

196

197 **Mutations in spliceosomal component genes**

198 Mutations affecting different components of the spliceosome have been identified in a range
199 of solid and non-solid malignancies (Biankin, et al. 2012; Furney, et al. 2013;
200 Papaemmanuil, et al. 2011; Quesada, et al. 2011; Yoshida and Ogawa 2014). Mutations in
201 the splicing factor SF3B1 are the most common across multiple tumour types, and at
202 particular high frequencies in myelodysplastic syndrome (MDS), chronic myeloid leukaemia
203 (CLL), uveal melanoma, (UV), pancreatic cancer and breast cancer. Mutations generally
204 cluster at hotspot amino acid residues K700, R625, K666, and H662 (Cerami, et al. 2012;
205 Gao, et al. 2013). However, each cancer type harbours a different variation of hotspot
206 mutations. For example, K700E mutations are invariably found in breast cancer, pancreatic
207 cancer and CLL, whereas uveal melanoma and endometrial cancers harbour the R625,
208 R666 and R662 hotspots, suggesting some tissue specificity of the mutations. SF3B1
209 hotspot mutations in CLL are associated with a poor prognosis however in uveal melanoma
210 and MDS the prognosis is better with the presence of an SF3B1 mutation (Furney et al.
211 2013; Quesada et al. 2011). Interestingly, additional spliceosomal component genes are
212 also recurrently mutated at high frequencies particularly in myleodysplastic syndromes,
213 including U2AF1, which has a distinct S34F/Y hotspot mutation and mutations in SRSF2 that
214 are associated with a poor outcome in MDS (Thol, et al. 2012); and ZRSR2. Both SRSF2
215 and ZRSR2 both harbour mutations spread throughout the gene, suggestive of a tumour
216 suppressive function (Yoshida, et al. 2011). Mutations in these genes including SF3B1 occur
217 in a mutually exclusive manner in MDS, suggesting that cells may tolerate only a partial
218 deviation from normal splicing activity. Indeed, these genes are all involved in the 3′-splice
219 site recognition during pre-mRNA processing, inducing abnormal RNA splicing and
220 compromised haematopoiesis (Yoshida et al. 2011), implicating splicing dysregulation as a
221 major driving force behind the development of MDS.

222

Our group has explored the mutational repertoire of spliceosomal component genes in breast cancer from a meta-analysis of whole genome and exome sequencing data (Maguire et al. 2015) (Figure 2). This analysis identified that around 5.6% of unselected breast cancers have mutations in spliceosome component genes at low frequencies. The most common spliceosomal gene mutation is *SF3B1,* which is associated with ER+ breast cancer and seen in around 3% of ER+ tumours (Pereira, et al. 2016), whereas mutations in *SON* and *SAP130* appear to be associated with ER- disease (Maguire et al. 2015). Interestingly, we identified SF3B1 K700E mutations at higher frequencies in some rarer histological subtypes of breast cancer including 16% of papillary carcinomas and 8% of mucinous carcinomas of the breast, suggesting they may underpin their biology (Maguire et al. 2015). SF3B1 K700E mutations were also found to associate with losses of 16q11-q13 and gains of 16q12-q13 indicating a distinct mechanism of breast cancer progression independent of the canonical early event of 1q gain and 16q loss (Maguire et al. 2015).

The association of *SF3B1* mutations and breast cancer clinical prognosis however is unclear, although mutations are being increasingly seen in metastatic disease (Lefebvre, et al. 2016; Pereira, et al. 2016). Further studies however are needed in order to truly assess the effect *SF3B1* hotspot mutations on outcome. Of note, *SF3B1* mutations have been observed in adenoid cystic carcinomas of the breast (an ER-negative special histological subtype) that has an excellent clinical outcome and at increased frequency in ER+ mucinous and papillary carcinomas of the breast, These data perhaps suggest that *SF3B1* mutations maybe associated with a good prognosis (Maguire et al. 2015; Martelotto, et al. 2015).

SF3B is a complex that is part of the U2 spliceosome and controls 3' SS recognition. Its core is required for alignment of the branch site proteins, which allows for correct branch site selection during the splicing process (Cretu, et al. 2016). *SF3B1 (SF3B155)* is the largest component of the SF3B complex and contains the HEAT superhelix domain consisting of 20 tandem repeats of two alpha helices joined by a short loop. Mutations in the HEAT domains,

which are responsible for interacting with pre-mRNA and other pre-mRNA binding proteins, result in a change in the tertiary structure that causes the selection of an alternative branch site (Alsafadi, et al. 2016; Darman et al. 2015; Kesarwani, et al. 2017). It is not known however, whether the SF3B1 mutant protein has a stronger affinity for the newly exposed branch point sequence or if it is coping with a disruption in binding to the canonical branch point sequence (Darman et al. 2015). Indeed, mutations in SF3B1 lead to alternative branchpoint usage and subsequent usage of a 3' cryptic SS. This leads to aberrant transcript expression and subsequent nonsense mediated decay of around half the aberrant transcripts and hence leads to protein downregulation (Alsafadi et al. 2016; Darman et al. 2015; Kesarwani et al. 2017).

Although present as hotspot single amino acid changes, SF3B1 mutations are thought to lead to a change in function. This is because knockdown or overexpression of the mutant protein does not recapitulate the aberrant splice pattern seen in mutant versus wildtype patients (Alsafadi et al. 2016). Additional evidence suggests that these mutations may actually be loss of canonical function. For instance, using the Degron-knock in approach to inactivate mutant or wild-type alleles specifically, Zhou et al. found that degradation of only the mutant SF3B1 allele in heterozygous SF3B1 mutant cells had no effect on growth whereas degradation of only the wildtype allele resulted in a decrease in viability of the cells (Zhou, et al. 2015). This suggests that SF3B1 is not likely to be haploinsufficient given the cells are solely relying on the wildtype copy of the gene to survive. This observation helps explain why SF3B1 mutations are uniformly heterozygous, as two copies of the mutant allele would likely be lethal.

The most common SF3B1 mutation in breast cancer is the K700E variant akin to CLL but K666Q and K666E are also observed, albeit at much lower frequencies (Maguire et al. 2015). Gene expression analysis in ER-positive disease shows that SF3B1 mutations affect regulators of the cell cycle, metabolism, and motility as well as protein degradation and

apoptosis, and splicing regulation itself (Maguire et al. 2015). Commonly differentially spliced mRNAs have been associated with SF3B1 mutations across tumour types including uveal melanoma, chronic lymphocytic leukaemia, pancreatic cancer, and breast cancer. Although a large number of transcripts have been identified to be aberrantly spliced and some are cancer specific (e.g. ABCB7 alternative splicing is only observed in MDS and gives rise to increased mitochondrial iron accumulation found in MDS patients with ring sideroblasts (Dolatshad, et al. 2016), the overlap is rather strikingly consistent between tumour types, suggesting that there is a distinct signature of genes that are alternatively spliced and furthermore can be used as markers of the mutation status (Biankin et al. 2012; Dolatshad et al. 2016; Furney et al. 2013; Maguire et al. 2015; Obeng, et al. 2016; Quesada et al. 2011; Wang, et al. 2016). However, it has not yet been identified which of the many differentially spliced genes is/are responsible for the tumorigenic phenotype and if these are different between different cancer types. In our study, we used siRNA to silence different genes that had been identified as alternatively spliced in our data set as well as across multiple cancer types. Silencing eight different genes (*ABCC5, ANKHD1, DYNLL1, F8, RPL31, TMEM14C, UQCC, and CRNDE)* did not show any changes in viability (Maguire et al. 2015). Given around half of all aberrantly expressed transcripts are subjected to NMD, they could be acting as tumour suppressors rather than in an oncogenic manner and will need to be explored in the future.

Spliceosomal component genes as oncoproteins in breast cancer

As well as mutations, alterations in components of the spliceosome, such as deletions or amplifications, are commonly seen across breast cancer (Figure 2, Table 1). In a similar vein to spliceosomal component mutations, they may lead to dysregulation of canonical splicing. SF3B3 (SF3B130) a component of the SF3B complex has been found to be significantly overexpressed in ER+ breast cancers and is associated with aggressive disease and resistance to tamoxifen therapy (Gokmen-Polar et al. 2015). SF3B3 is positioned closely to SF3B1 in the U2 complex and helps maintain the HEAT domain's structural plasticity and

has the ability to alter pre-mRNA splicing hence affecting gene expression in the cell (Garcia-Blanco, et al. 2004). Over-expression of SF3B3 has thus been postulated to contribute to spicing aberrations in cancer cells. In clear cell renal cell carcinoma, SF3B3 overexpression was found to increase the expression of the pro-proliferative full length isoform of *EZH2* and not the commonly expressed *EZH2Δ14* that is found in normal tissue (Chen, et al. 2017), thus promoting tumorigenicity *in vivo*. It could be that EZH2 AS plays a role in mediating the aggressive behaviour in endocrine resistant ER+ breast cancer as well, however this has yet to be elucidated. SF3B3 has also been found to be amplified and highly expressed at the transcript level in basal-like breast cancers (Srihari, et al. 2016). Overall the level is actually higher in ER- than ER+ disease, perhaps highlighting the higher proliferative rate of these tumours.

The SRSF family of proteins are serine-arginine rich splicing factors that are commonly found to be mutated or dysregulated in cancer (Das and Krainer 2014). These proteins contain RNA recognition motif (RPM) domains that contact the mRNA and also interact with other splicing machinery (Das and Krainer 2014). *SRSF1* also referred to as SF2/ASF is the most common protein of this family to play a role in breast cancer and overexpression is associated with a poor prognosis in ER+ breast cancers (Anczukow, et al. 2012). Overexpressing SRSF1 in 3D mammary organotypic assays is associated with larger acini structures indicating its oncogenic phenotype (Anczukow et al. 2012). This study also highlighted specific isoform dysregulation of the tumour suppressors BIM and BIN1, which resulted in loss of their pro-apoptotic functions (Anczukow et al. 2012; Karni, et al. 2007). SRSF1 upregulation is thought to play a role in EMT through alternative splicing modulation of its transcriptional target genes (Valacca, et al. 2010). Mechanistically this is linked back to the splicing regulator Sam68, which modulates levels of SRSF1 (Valacca et al. 2010). It was found that SRSF1 is more likely to facilitate exon inclusion when it binds closer to the 5' site of the splice junction and promotes exon skipping or inclusion when it binds to the 3' end (Anczukow et al. 2015). SRSF1 was found to alternatively splice *CASC4* by including exon

9, resulting in a longer protein. When tested alone, overexpression of this isoform of *CASC4* phenocopied the tumorigenic abilities of SRSF1 overexpression by increasing proliferation and acinar size and decreasing apoptosis (Anczukow et al. 2015). These data highlight promising targets for therapeutic development in patients with SRSF1 overexpression.

Other members of the SRSF family have also been implicated in breast cancer. For instance, SRSF2 gene amplification at 17q25 has been observed in 6% of breast cancers, although it is uncertain whether this plays an oncogenic role, given evidence that mutations are loss of function in this gene (Chung, et al. 2017). Finally, SRSF4 overexpression has been identified in a small subset of breast cancer and its expression has been found responsible for cisplatin induced alternative splicing that leads to apoptosis. Experiments where SRSF4 was silenced showed a decrease in apoptosis upon treatment with cisplatin and highlight the possibility of modulating splicing to regulate chemotherapy sensitivity (Gabriel, et al. 2015).

Dysregulation of spliceosomal accessory proteins

Along with the major components of the spliceosome that were described above, there are also other regulators of splicing that have been found to be mutated or dysregulated in breast cancer. LIN28A has been identified specifically in HER2–positive breast cancer as being a regulator of alternative splicing through interactions with hnRNPA1 (Xiong, et al. 2017; Yang, et al. 2015). Loss of LIN28A in breast cancer results in isoform switching of the ENAH gene, which is overexpressed in some primary breast tumours (Xiong et al. 2017; Yang et al. 2015). It has also been identified as a feature of the malignant phenotype in a model of breast cancer progression and has been correlated with an unfavourable outcome in HER2-positive breast cancer (Du, et al. 2012). Another example are the epithelial splicing regulatory proteins (ESRP1 and ESRP2) which are splicing factors that have been found to regulate the alternative splicing that governs the epithelial to mesenchymal transition and are amplified in breast cancers (Bebee, et al. 2015; Brown, et al. 2011; Warzecha, et al.

2009) and regulate EMT in breast tumours via activating AKT signalling (Brown et al. 2011). The RNA binding protein RBFOX2 is also involved in cellular transition, whose upregulation can perturb splicing events in breast cancer (Du et al. 2012; Lapuk, et al. 2010). During EMT, RBFOX2-regulated splicing shifts from epithelial-to mesenchymal-specific events, subsequently leading to a higher degree of tissue invasiveness (Braeutigam, et al. 2014). Another RNA binding protein, RBM47, has the ability to alter splicing by binding to introns and 3' UTRs and loss of expression has been shown to prevent breast cancer progression and metastasis (Vanharanta et al. 2014). Taken together these lines of evidence point to a fundamental role triggered by splicing dysregulation in breast cancer cells that can cause detrimental effects and lead to the progression of disease.

Evidence of oncogene induced dependency on the spliceosome

Aside from alterations in spliceosomal component genes themselves, there is emerging evidence that oncogene activation imparts a functional dependency on SF3B1 and other components in breast cancer. A number of spliceosomal component proteins are known transcriptional targets of the oncoprotein MYC (including *SF3B1* and *SRSF1*), and have been shown to both contribute to and co-operate with MYC in malignant transformation (Das, et al. 2012; Koh, et al. 2015). For instance, MYC addicted triple-negative breast cancers cells have been shown to impart a specific dependency on the spliceosome via *BUD31* and *SF3B1* (Hsu, et al. 2015) and impaired tumorigenesis was observed when SF3B1 was knocked down or pharmacologically inhibited in breast cancer cells hyper expressing MYC (Hsu et al. 2015). This could be explained due to the increased burden put on the spliceosome when the rate of transcription is increased due to MYC signalling. Recently, knockdown of SF3B1 was found to result in apoptosis in TNBC with MCL-1 inactivation being a likely mechanistic explanation, given MCL-1 is and SF3B1 splicing target (Gao and Koide 2013; Sridhar P, et al. 2017). Interestingly, MYC and MCL-1 have been shown to cooperate in chemo-resistant TNBCs (Lee, et al. 2017). This could be further support for the intricate co-operation of MYC with the spliceosome and the resulting changes

391 in isoform dominance that allow the manipulation of cancer cells. In addition, SRSF1 is a
392 known direct target of MYC. MYC induction leads to SRSF1-mediated alternative splicing of
393 key protein isoforms involved in proliferation and anchorage-independent growth such as
394 MKNK2 and TEAD1 (Anczukow et al. 2012; Das et al. 2012), which is in part through
395 potentiating eIF4E activation (Anczukow et al. 2012; Das et al. 2012). Together these
396 studies suggest that multiple spliceosomal proteins are critical MYC targets that contribute to
397 its oncogenic potential by enabling MYC to regulate the expression of specific protein
398 isoforms via alternative splicing.
399
400 **Therapeutic targeting of the spliceosome**
401 There is emerging evidence that disruption of spliceosomal proteins induces selectivity to
402 inhibitors that target the spliceosome. Indeed a number of these inhibitors have been
403 developed including Spliceostatin A, Pladienolides (including E7107), and meyamycin
404 analogues that are all specific SF3B inhibitors as reviewed in (Lee and Abdel-Wahab 2016)
405 that inhibit canonical splicing (Kaida, et al. 2007). We, and others, have shown that *SF3B1*
406 mutant cells selectively sensitive to spliceosomal inhibitors (Maguire et al. 2015; Obeng et
407 al. 2016). Moreover, SF3b inhibition in *SF3B1* mutant cells resulted in a change in the
408 reversal of the conserved splicing signature, suggesting that *SF3B1* mutations are change of
409 function rather than loss of function and that these alterations in aberrant isoforms could be
410 used as biomarkers of therapeutic response (Maguire et al. 2015). There is additional
411 evidence that other spliceosomal gene mutations can be therapeutically targeted with
412 spliceosomal inhibitors. These include SRSF2 mutations, whereby genetically modified mice
413 expressing the Srsf2(P95H) mutation were sensitive to treatment with the spliceosome
414 inhibitor E7107, which decreased leukemic burden (Lee, et al. 2016). Similar selective
415 sensitivity in mutant U2AF1 cells to sudemycins has also been reported in *in vitro* and *in vivo*
416 (Shirai C, et al. 2015). In addition, MYC addicted TNBC's have been shown to be more
417 sensitive to inhibition with the spliceosome inhibitor SD6 than MYC non-addicted cells are
418 (Hsu et al. 2015), a mechanism that is likely due to the increased stress and dependency on

SF3B1 (as discussed above). Further functional studies in the context of clear cell renal carcinoma show that knockdown of SF3B3 in SF3B3 overexpressing cells *in vivo* reduced tumour growth, highlighting the potential utility of SF3b inhibitors as a therapeutic agent for patients with SF3B3 amplification and/or overexpression (Chen et al. 2017). These lines of evidence raise the possible clinical utility of SF3b inhibitors in patients with additional spliceosomal gene mutations as well as other indirect reliance on the spliceosome. Further studies are warranted to ascertain if overexpression of spliceosomal genes also confers sensitivity to these compounds in breast cancer.

Phase one clinical trials have been performed for E7107 in patients with solid tumours and although the drug has been shown to be on target in patients (i.e. perturbs splicing), the US and European trials were suspended due to an unexpected toxicity involving bilateral optic neuritis (Eskens, et al. 2013; Hong, et al. 2014). Further studies to understand causes of toxicity as well as new clinical trials will be necessary to take advantage of splicing's therapeutic vulnerability in cancer. Currently, H3 biomedicine is testing the compound H3B-8800 which inhibits the SF3b complex and was successful in preclinical studies treating a range of spliceosomal mutant cancers (Buonamici, et al. 2016). The compound is now in phase one studies (NCT02841540) for Myelodysplastic Sydromes, Acute Myeloid Leukemia, and Chronic Myelomonocytic Leukemia.

Conclusions

Mutations and changes in expression of splicing factors that lead to aberrant splicing is a hallmark of cancer that is also relevant to breast cancer. Development of prognostic and predictive aberrant splicing signatures specifically to predict patients that will respond to endocrine (or indeed CDK4/6 inhibitor) therapy and could be used particularly useful going forward. The increasing technical advances in sequencing methodologies, particularly those that aim to increase RNA read lengths, will undoubtedly enhance the ability to detect these

events in the future and further increase our understanding of aberrant transcript expression on breast cancer tumorigenesis and therapy resistance. There is increasing evidence that spliceosomal component genes themselves are dysregulated in breast cancer, through mutations in *SF3B1* that are also observed in metastatic disease and upregulation of SF3B3 and SRSF1 in particular, which are associated with resistance to endocrine therapy. Dissecting the function of the expression of the consequent alternatively spliced transcripts would give insight into the mechanism of these alterations and the role they play in therapy resistance. Indeed with the development of spliceosome inhibitors themselves, and exciting preclinical data in other tumour types highlights a potential novel treatment strategy in combination with endocrine therapy and CDK4/6 inhibitors for patients with metastatic disease with spliceosomal gene alterations.

Declaration of interest, Funding and Acknowledgements

Declaration of interest

The authors declare no conflict of interest

Funding

We thank Breast Cancer Now for funding this work as part of Programme Funding to the Breast Cancer Now Toby Robins Research Centre

References

Alsafadi S, Houy A, Battistella A, Popova T, Wassef M, Henry E, Tirode F, Constantinou A, Piperno-Neumann S, Roman-Roman S, et al. 2016 Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nat Commun* **7** 10615.
Anczukow O, Akerman M, Clery A, Wu J, Shen C, Shirole NH, Raimer A, Sun S, Jensen MA, Hua Y, et al. 2015 SRSF1-Regulated Alternative Splicing in Breast Cancer. *Mol Cell* **60** 105-117.
Anczukow O, Rosenberg AZ, Akerman M, Das S, Zhan L, Karni R, Muthuswamy SK & Krainer AR 2012 The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. *Nat Struct Mol Biol* **19** 220-228.
Bae J, Leo CP, Hsu SY & Hsueh AJ 2000 MCL-1S, a splicing variant of the antiapoptotic BCL-2 family member MCL-1, encodes a proapoptotic protein possessing only the BH3 domain. *J Biol Chem* **275** 25255-25261.
Bebee TW, Park JW, Sheridan KI, Warzecha CC, Cieply BW, Rohacek AM, Xing Y & Carstens RP 2015 The splicing regulators Esrp1 and Esrp2 direct an epithelial splicing program essential for mammalian development. *Elife* **4**.
Beije N, Sieuwerts AM, Kraan J, Van NM, Onstenk W, Vitale SR, van der Vlugt-Daane M, Dirix LY, Brouwer A, Hamberg P, et al. 2018 Estrogen receptor mutations and splice variants determined in liquid biopsies from metastatic breast cancer patients. *Mol Oncol* **12** 48-57.
Biankin AV, Waddell N, Kassahn KS, Gingras MC, Muthuswamy LB, Johns AL, Miller DK, Wilson PJ, Patch AM, Wu J, et al. 2012 Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* **491** 399-405.
Bingle CD, Craig RW, Swales BM, Singleton V, Zhou P & Whyte MK 2000 Exon skipping in Mcl-1 results in a bcl-2 homology domain 3 only gene product that promotes cell death. *J Biol Chem* **275** 22136-22146.
Boise LH, Gonzalez-Garcia M, Postema CE, Ding L, Lindsten T, Turka LA, Mao X, Nunez G & Thompson CB 1993 bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death. *Cell* **74** 597-608.
Bollig A & Miksicek RJ 2000 An estrogen receptor-alpha splicing variant mediates both positive and negative effects on gene transcription. *Mol Endocrinol* **14** 634-649.
Braeutigam C, Rago L, Rolke A, Waldmeier L, Christofori G & Winter J 2014 The RNA-binding protein Rbfox2: an essential regulator of EMT-driven alternative splicing and a mediator of cellular invasion. *Oncogene* **33** 1082-1092.
Brown RL, Reinke LM, Damerow MS, Perez D, Chodosh LA, Yang J & Cheng C 2011 CD44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression. *J Clin Invest* **121** 1064-1074.
Buonamici S, Thomas M, Seiler M, Chan B, Caleb B, Darman R, Fekkes P, Karr C, Keaney C, Klimek V, et al. 2016 H3B-8800, an Orally Bioavailable Modulator of the SF3b Complex, Shows Efficacy in Spliceosome-Mutan.
Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, et al. 2012 The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2** 401-404.
Chen K, Xiao H, Zeng J, Yu G, Zhou H, Huang C, Yao W, Xiao W, Hu J, Guan W, et al. 2017 Alternative Splicing of EZH2 pre-mRNA by SF3B3 Contributes to the Tumorigenic Potential of Renal Cancer. *Clin Cancer Res* **23** 3428-3441.

Chung FF, Tan PF, Raja VJ, Tan BS, Lim KH, Kam TS, Hii LW, Tan SH, See SJ, Tan YF, et al. 2017 Jerantinine A induces tumor-specific cell death through modulation of splicing factor 3b subunit 1 (SF3B1). *Sci Rep* **7** 42504.
Cretu C, Schmitzova J, Ponce-Salvatierra A, Dybkov O, De Laurentiis EI, Sharma K, Will CL, Urlaub H, Luhrmann R & Pena V 2016 Molecular Architecture of SF3b and Structural Consequences of Its Cancer-Related Mutations. *Mol Cell* **64** 307-319.
da Luz FAC, Brigido PC, Moraes AS, Araujo RA & Silva MJB 2017 Splicing Factors in Breast Cancer: Drivers of the Breast Tumor Fate.
Danan-Gotthold M, Golan-Gerstl R, Eisenberg E, Meir K, Karni R & Levanon EY 2015 Identification of recurrent regulated alternative splicing events across human solid tumors. *Nucleic Acids Res* **43** 5130-5144.
Darman RB, Seiler M, Agrawal AA, Lim KH, Peng S, Aird D, Bailey SL, Bhavsar EB, Chan B, Colla S, et al. 2015 Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3' Splice Site Selection through Use of a Different Branch Point. *Cell Rep* **13** 1033-1045.
Das S, Anczukow O, Akerman M & Krainer AR 2012 Oncogenic splicing factor SRSF1 is a critical transcriptional target of MYC. *Cell Rep* **1** 110-117.
Das S & Krainer AR 2014 Emerging functions of SRSF1, splicing factor and oncoprotein, in RNA metabolism and cancer. *Mol Cancer Res* **12** 1195-1204.
David CJ & Manley JL 2010 Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev* **24** 2343-2364.
Dolatshad H, Pellagatti A, Liberante FG, Llorian M, Repapi E, Steeples V, Roy S, Scifo L, Armstrong RN, Shaw J, et al. 2016 Cryptic splicing events in the iron transporter ABCB7 and other key target genes in SF3B1-mutant myelodysplastic syndromes. *Leukemia* **30** 2322-2331.
Du JW, Xu KY, Fang LY & Qi XL 2012 Clinical significance of Mena and Her-2 expression in breast cancer. *Eur J Gynaecol Oncol* **33** 455-458.
Dvinge H, Kim E, Abdel-Wahab O & Bradley RK 2016 RNA splicing factors as oncoproteins and tumour suppressors. *Nat Rev Cancer* **16** 413-430.
Eskens FA, Ramos FJ, Burger H, O'Brien JP, Piera A, de Jonge MJ, Mizui Y, Wiemer EA, Carreras MJ, Baselga J, et al. 2013 Phase I pharmacokinetic and pharmacodynamic study of the first-in-class spliceosome inhibitor E7107 in patients with advanced solid tumors. *Clin Cancer Res* **19** 6296-6304.
Fuqua SA, Fitzgerald SD, Chamness GC, Tandon AK, McDonnell DP, Nawaz Z, O'Malley BW & McGuire WL 1991 Variant human breast tumor estrogen receptor with constitutive transcriptional activity. *Cancer Res* **51** 105-109.
Furney SJ, Pedersen M, Gentien D, Dumont AG, Rapinat A, Desjardins L, Turajlic S, Piperno-Neumann S, de la Grange P, Roman-Roman S, et al. 2013 SF3B1 mutations are associated with alternative splicing in uveal melanoma. *Cancer Discov* **3** 1122-1129.
Fushimi K, Ray P, Kar A, Wang L, Sutherland LC & Wu JY 2008 Up-regulation of the proapoptotic caspase 2 splicing isoform by a candidate tumor suppressor, RBM5. *Proc Natl Acad Sci U S A* **105** 15708-15713.
Gabriel M, Delforge Y, Deward A, Habraken Y, Hennuy B, Piette J, Klinck R, Chabot B, Colige A & Lambert C 2015 Role of the splicing factor SRSF4 in cisplatin-induced modifications of pre-mRNA splicing and apoptosis. *BMC Cancer* **15** 227.
Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, et al. 2013 Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6** pl1.

Gao Y & Koide K 2013 Chemical perturbation of Mcl-1 pre-mRNA splicing to induce apoptosis in cancer cells. *ACS Chem Biol* **8** 895-900.
Garcia-Blanco MA, Baraniak AP & Lasda EL 2004 Alternative splicing in disease and therapy. *Nat Biotechnol* **22** 535-546.
Gautrey HL & Tyson-Capper AJ 2012 Regulation of Mcl-1 by SRSF1 and SRSF5 in cancer cells. *PLoS One* **7** e51497.
Gokmen-Polar Y, Neelamraju Y, Goswami CP, Gu X, Nallamothu G, Janga SC & Badve S 2015 Expression levels of SF3B3 correlate with prognosis and endocrine resistance in estrogen receptor-positive breast cancer. *Mod Pathol* **28** 677-685.
Gracio F, Burford B, Gazinska P, Mera A, Mohd Noor A, Marra P, Gillett C, Grigoriadis A, Pinder S, Tutt A, et al. 2017 Splicing imbalances in basal-like breast cancer underpin perturbation of cell surface and oncogenic pathways and are associated with patients' survival. *Sci Rep* **7** 40177.
Herzel L, Ottoz DSM, Alpert T & Neugebauer KM 2017 Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat Rev Mol Cell Biol* **18** 637-650.
Hong DS, Kurzrock R, Naing A, Wheler JJ, Falchook GS, Schiffman JS, Faulkner N, Pilat MJ, O'Brien J & LoRusso P 2014 A phase I, open-label, single-arm, dose-escalation study of E7107, a precursor messenger ribonucleic acid (pre-mRNA) splicesome inhibitor administered intravenously on days 1 and 8 every 21 days to patients with solid tumors. *Invest New Drugs* **32** 436-444.
Hsu TY, Simon LM, Neill NJ, Marcotte R, Sayad A, Bland CS, Echeverria GV, Sun T, Kurley SJ, Tyagi S, et al. 2015 The spliceosome is a therapeutic vulnerability in MYC-driven cancer. *Nature* **525** 384-388.
Kaida D, Motoyoshi H, Tashiro E, Nojima T, Hagiwara M, Ishigami K, Watanabe H, Kitahara T, Yoshida T, Nakajima H, et al. 2007 Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nat Chem Biol* **3** 576-583.
Karni R, de Stanchina E, Lowe SW, Sinha R, Mu D & Krainer AR 2007 The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat Struct Mol Biol* **14** 185-193.
Keren H, Lev-Maor G & Ast G 2010 Alternative splicing and evolution: diversification, exon definition and function. *Nat Rev Genet* **11** 345-355.
Kesarwani AK, Ramirez O, Gupta AK, Yang X, Murthy T, Minella AC & Pillai MM 2017 Cancer-associated SF3B1 mutants recognize otherwise inaccessible cryptic 3' splice sites within RNA secondary structures. *Oncogene* **36** 1123-1133.
Khan DH, Gonzalez C, Cooper C, Sun JM, Chen HY, Healy S, Xu W, Smith KT, Workman JL, Leygue E, et al. 2014 RNA-dependent dynamic histone acetylation regulates MCL1 alternative splicing. *Nucleic Acids Res* **42** 1656-1670.
Koh CM, Bezzi M, Low DH, Ang WX, Teo SX, Gay FP, Al-Haddawi M, Tan SY, Osato M, Sabo A, et al. 2015 MYC regulates the core pre-mRNA splicing machinery as an essential step in lymphomagenesis. *Nature* **523** 96-100.
Ladomery M 2013 Aberrant alternative splicing is another hallmark of cancer. *Int J Cell Biol* **2013** 463786.
Lapuk A, Marr H, Jakkula L, Pedro H, Bhattacharya S, Purdom E, Hu Z, Simpson K, Pachter L, Durinck S, et al. 2010 Exon-level microarray analyses identify alternative splicing programs in breast cancer. *Mol Cancer Res* **8** 961-974.
Lee KM, Giltnane JM, Balko JM, Schwarz LJ, Guerrero-Zotano AL, Hutchinson KE, Nixon MJ, Estrada MV, Sanchez V, Sanders ME, et al. 2017 MYC and MCL1 Cooperatively Promote

Chemotherapy-Resistant Breast Cancer Stem Cells via Regulation of Mitochondrial Oxidative Phosphorylation. *Cell Metab* **26** 633-647 e637.
Lee SC & Abdel-Wahab O 2016 Therapeutic targeting of splicing in cancer. *Nat Med* **22** 976-986.
Lee SC, Dvinge H, Kim E, Cho H, Micol JB, Chung YR, Durham BH, Yoshimi A, Kim YJ, Thomas M, et al. 2016 Modulation of splicing catalysis for therapeutic targeting of leukemia with mutations in genes encoding spliceosomal proteins. *Nat Med* **22** 672-678.
Lefebvre C, Bachelot T, Filleron T, Pedrero M, Campone M, Soria JC, Massard C, Levy C, Arnedos M, Lacroix-Triki M, et al. 2016 Mutational Profile of Metastatic Breast Cancers: A Retrospective Analysis. *PLoS Med* **13** e1002201.
Maguire SL, Leonidou A, Wai P, Marchio C, Ng CK, Sapino A, Salomon AV, Reis-Filho JS, Weigelt B & Natrajan RC 2015 SF3B1 mutations constitute a novel therapeutic target in breast cancer. *J Pathol* **235** 571-580.
Marchini C, Gabrielli F, Iezzi M, Zenobi S, Montani M, Pietrella L, Kalogris C, Rossini A, Ciravolo V, Castagnoli L, et al. 2011 The human splice variant Delta16HER2 induces rapid tumor onset in a reporter transgenic mouse. *PLoS One* **6** e18727.
Martelotto LG, De Filippo MR, Ng CK, Natrajan R, Fuhrmann L, Cyrta J, Piscuoglio S, Wen HC, Lim RS, Shen R, et al. 2015 Genomic landscape of adenoid cystic carcinoma of the breast. *J Pathol* **237** 179-189.
Martinez-Montiel N, Anaya-Ruiz M, Perez-Santos M & Martinez-Contreras RD 2017 Alternative Splicing in Breast Cancer and the Potential Development of Therapeutic Tools. *Genes (Basel)* **8**.
Menon R, Im H, Zhang EY, Wu SL, Chen R, Snyder M, Hancock WS & Omenn GS 2014 Distinct splice variants and pathway enrichment in the cell-line models of aggressive human breast cancer subtypes. *J Proteome Res* **13** 212-227.
Ng KP, Hillmer AM, Chuah CT, Juan WC, Ko TK, Teo AS, Ariyaratne PN, Takahashi N, Sawada K, Fei Y, et al. 2012 A common BIM deletion polymorphism mediates intrinsic resistance and inferior responses to tyrosine kinase inhibitors in cancer. *Nat Med* **18** 521-528.
Obeng EA, Chappell RJ, Seiler M, Chen MC, Campagna DR, Schmidt PJ, Schneider RK, Lord AM, Wang L, Gambe RG, et al. 2016 Physiologic Expression of Sf3b1(K700E) Causes Impaired Erythropoiesis, Aberrant Splicing, and Sensitivity to Therapeutic Spliceosome Modulation. *Cancer Cell* **30** 404-417.
Pandya-Jones A 2011 Pre-mRNA splicing during transcription in the mammalian system. *Wiley Interdiscip Rev RNA* **2** 700-717.
Papaemmanuil E, Cazzola M, Boultwood J, Malcovati L, Vyas P, Bowen D, Pellagatti A, Wainscoat JS, Hellstrom-Lindberg E, Gambacorti-Passerini C, et al. 2011 Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N Engl J Med* **365** 1384-1395.
Pereira B, Chin SF, Rueda OM, Vollan HK, Provenzano E, Bardwell HA, Pugh M, Jones L, Russell R, Sammut SJ, et al. 2016 The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun* **7** 11479.
Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, et al. 2000 Molecular portraits of human breast tumours. *Nature* **406** 747-752.
Quesada V, Conde L, Villamor N, Ordonez GR, Jares P, Bassaganyas L, Ramsay AJ, Bea S, Pinyol M, Martinez-Trillos A, et al. 2011 Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat Genet* **44** 47-52.

Raponi M, Smith LD, Silipo M, Stuani C, Buratti E & Baralle D 2014 BRCA1 exon 11 a model of long exon splicing regulation. *RNA Biol* **11** 351-359.
Sebestyen E, Zawisza M & Eyras E 2015 Detection of recurrent alternative splicing switches in tumor samples reveals novel signatures of cancer. *Nucleic Acids Res* **43** 1345-1356.
Shapiro IM, Cheng AW, Flytzanis NC, Balsamo M, Condeelis JS, Oktay MH, Burge CB & Gertler FB 2011 An EMT-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype. *PLoS Genet* **7** e1002218.
Shirai C, Tripathi M, Ley J, Ndonwi M, White BS, Tapia R, Saez B, Bertino A, Shao J, Kim S, et al. 2015 Preclinical Activity of Splicing Modulators in U2AF1 Mutant MDS/AML. *Blood* **126:1653**.
Siegfried Z & Karni R 2017 The role of alternative splicing in cancer drug resistance. *Curr Opin Genet Dev* **48** 16-21.
Silipo M, Gautrey H & Tyson-Capper A 2015 Deregulation of splicing factors and breast cancer development. *J Mol Cell Biol* **7** 388-401.
Sotillo E, Barrett DM, Black KL, Bagashev A, Oldridge D, Wu G, Sussman R, Lanauze C, Ruella M, Gazzara MR, et al. 2015 Convergence of Acquired Mutations and Alternative Splicing of CD19 Enables Resistance to CART-19 Immunotherapy. *Cancer Discov* **5** 1282-1295.
Sridhar P, Chan S, Lock YJ & F P 2017 Preclinical evaluation of the SF3B1 inhibitor E7107 in triple negative breast cancer. *Cancer Research*.
Srihari S, Kalimutho M, Lal S, Singla J, Patel D, Simpson PT, Khanna KK & Ragan MA 2016 Understanding the functional impact of copy number alterations in breast cancer using a network modeling approach. *Mol Biosyst* **12** 963-972.
Stark M, Wichman C, Avivi I & Assaraf YG 2009 Aberrant splicing of folylpolyglutamate synthetase as a novel mechanism of antifolate resistance in leukemia. *Blood* **113** 4362-4369.
Stricker TP, Brown CD, Bandlamudi C, McNerney M, Kittler R, Montoya V, Peterson A, Grossman R & White KP 2017 Robust stratification of breast cancer subtypes using differential patterns of transcript isoform expression. *PLoS Genet* **13** e1006589.
Takehara T, Liu X, Fujimoto J, Friedman SL & Takahashi H 2001 Expression and role of Bcl-xL in human hepatocellular carcinomas. *Hepatology* **34** 55-61.
Thakur S, Zhang HB, Peng Y, Le H, Carroll B, Ward T, Yao J, Farid LM, Couch FJ, Wilson RB, et al. 1997 Localization of BRCA1 and a splice variant identifies the nuclear localization signal. *Mol Cell Biol* **17** 444-452.
Thol F, Kade S, Schlarmann C, Loffeld P, Morgan M, Krauter J, Wlodarski MW, Kolking B, Wichmann M, Gorlich K, et al. 2012 Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. *Blood* **119** 3578-3584.
Valacca C, Bonomi S, Buratti E, Pedrotti S, Baralle FE, Sette C, Ghigna C & Biamonti G 2010 Sam68 regulates EMT through alternative splicing-activated nonsense-mediated mRNA decay of the SF2/ASF proto-oncogene. *J Cell Biol* **191** 87-99.
Vanharanta S, Marney CB, Shu W, Valiente M, Zou Y, Mele A, Darnell RB & Massague J 2014 Loss of the multifunctional RNA-binding protein RBM47 as a source of selectable metastatic traits in breast cancer. *Elife* **3**.
Wang L, Brooks AN, Fan J, Wan Y, Gambe R, Li S, Hergert S, Yin S, Freeman SS, Levin JZ, et al. 2016 Transcriptomic Characterization of SF3B1 Mutation Reveals Its Pleiotropic Effects in Chronic Lymphocytic Leukemia. *Cancer Cell* **30** 750-763.
Warzecha CC, Shen S, Xing Y & Carstens RP 2009 The epithelial splicing factors ESRP1 and ESRP2 positively and negatively regulate diverse types of alternative splicing events. *RNA Biol* **6** 546-562.

Wei M, Zhu L, Li Y, Chen W, Han B, Wang Z, He J, Yao H, Yang Z, Zhang Q, et al. 2011 Knocking down cyclin D1b inhibits breast cancer cell growth and suppresses tumor development in a breast cancer model. *Cancer Sci* **102** 1537-1544.
Wendt MK, Taylor MA, Schiemann BJ, Sossey-Alaoui K & Schiemann WP 2014 Fibroblast growth factor receptor splice variants are stable markers of oncogenic transforming growth factor beta1 signaling in metastatic breast cancers. *Breast Cancer Res* **16** R24.
Wiener D, Gajardo-Meneses P, Ortega-Hernandez V, Herrera-Cares C, Diaz S, Fernandez W, Cornejo V, Gamboa J, Tapia T, Alvarez C, et al. 2015 BRCA1 and BARD1 colocalize mainly in the cytoplasm of breast cancer tumors, and their isoforms show differential expression. *Breast Cancer Res Treat* **153** 669-678.
Xiong H, Zhao W, Wang J, Seifer BJ, Ye C, Chen Y, Jia Y, Chen C, Shen J, Wang L, et al. 2017 Oncogenic mechanisms of Lin28 in breast cancer: new functions and therapeutic opportunities. *Oncotarget* **8** 25721-25735.
Yang J, Bennett BD, Luo S, Inoue K, Grimm SA, Schroth GP, Bushel PR, Kinyamu HK & Archer TK 2015 LIN28A Modulates Splicing and Gene Expression Programs in Breast Cancer Cells. *Mol Cell Biol* **35** 3225-3243.
Yoshida K & Ogawa S 2014 Splicing factor mutations and cancer. *Wiley Interdiscip Rev RNA* **5** 445-459.
Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, Sato Y, Sato-Otsubo A, Kon A, Nagasaki M, et al. 2011 Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478** 64-69.
Zhao W, Hoadley KA, Parker JS & Perou CM 2016 Identification of mRNA isoform switching in breast cancer. *BMC Genomics* **17** 181.
Zhou Q, Derti A, Ruddy D, Rakiec D, Kao I, Lira M, Gibaja V, Chan H, Yang Y, Min J, et al. 2015 A chemical genetics approach for the functional assessment of novel cancer genes. *Cancer Res* **75** 1949-1958.

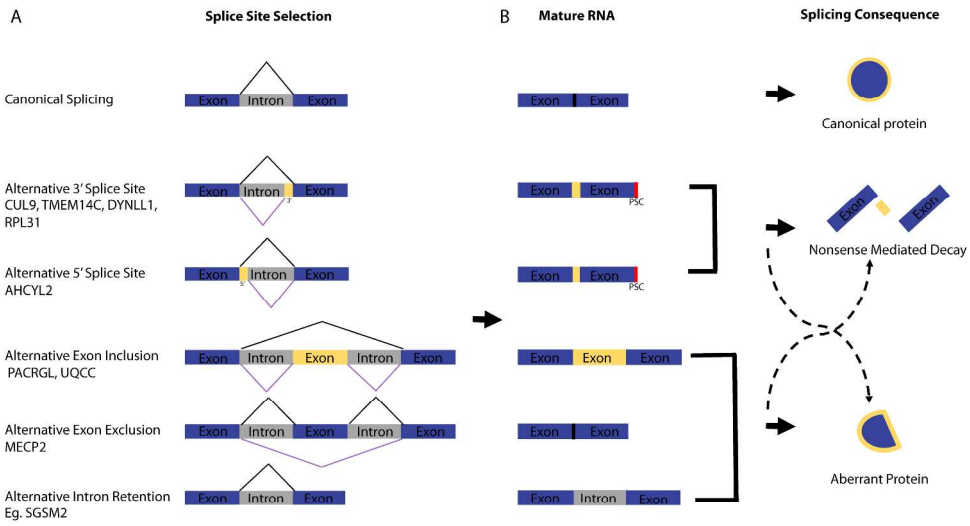
Figures

Figure 1: Mechanisms of alternative splicing in cancer

(A) Schematic of the possible ways in which alternative splicing can change the mRNA product. The product of canonical splicing is shown as well as the products of alternative splicing. Yellow represents non-canonical areas of the mRNA that are present in alternatively spliced transcripts. The black lines above the mRNA show where canonical splice sites are selected and the purple lines below the mRNA show where alternative splice sites are selected. Examples of genes for each event were obtained from (Darman et al. 2015). (B) the most likely product of the mRNA is indicated with solid dark arrows and the less likely but still possible products are indicated with dashed black arrows.

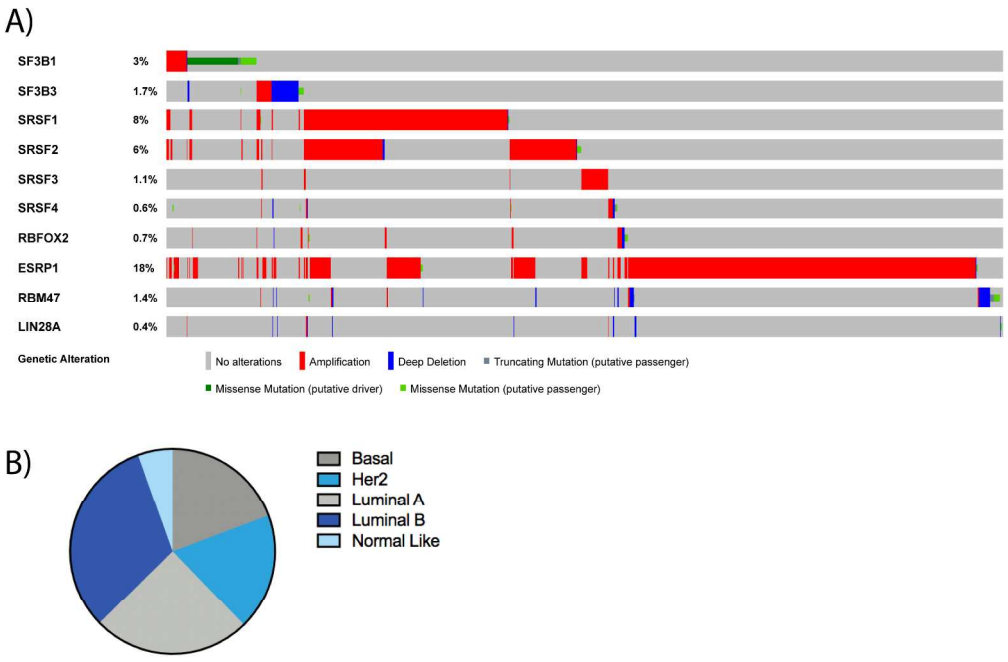
Figure 2: Summary of spliceosomal gene alterations in breast cancer

(A) cBioportal analysis of alterations in spliceosomal component genes from all available breast cancer data sets (Cerami et al. 2012; Gao et al. 2013). (B) Breakdown of patients with alterations by subtype from METABRIC and TCGA data with available PAM50 subtype calls. Basal= 19.3%, Her2= 18.5%, Luminal A= 24.9%, Luminal B= 31.9%, Normal-like= 5.5%.



Mechanisms of alternative splicing in cancer

297x210mm (300 x 300 DPI)



Summary of spliceosomal gene alterations in breast cancer

247x172mm (300 x 300 DPI)

Table 1

Summary of spliceosome component genes and RNA binding proteins found altered in breast cancer.

Splicing Factor/ RNA Binding Protein	Gene Name	Alteration	Occurrence in BrCa	Functional Impact
SF3B1	Splicing factor 3B subunit 1	Mutation and CNA	3%	Change of function, oncogenic
SF3B3	Splicing factor 3B subunit 3	CNA	1.7%	Oncogenic
SRSF1	Serine/arginine rich splicing factor 1	CNA	8%	Oncogenic
SRSF2	Serine/arginine rich splicing factor 2	CNA	6%	Oncogenic
SRSF3	Serine/arginine rich splicing factor 3	CNA	1.1%	Oncogenic
SRSF4	Serine/arginine rich splicing factor 4	CNA	0.6%	Oncogenic
RBFOX2	RNA binding protein fox-1 homolog 2	CNA	0.7%	EMT Regulator
ESRP1	Epithelial splicing regulatory protein 1	CNA	18%	EMT Regulator
RBM47	RNA binding motif protein 47	CNA	1.4%	Downregulation
LIN28A	Lin-28 Homolog A	CNA	0.4%	Loss of function

Sourced from all breast cancer studies available in cBioportal. CNA= Copy number alteration. n=4587 sequenced cases.

Table 2: Number and percentage of patients pertaining to each subtype with an alteration in the specified spliceosome component genes.

	SF3B1	**SF3B3**	**SRSF1**	**SRSF2**	**SRSF3**	**SRSF4**	**RBFOX2**	**ESRP1**	**RBM47**	**LIN28A**
Basal n= 391	12 (3.07)	10 (2.56)	11 (2.81)	22 (5.63)	15 (3.84)	6 (1.53)	5 (1.28)	84 (21.48)	13 (3.32)	1 (0.26)
Her2 n= 287	9 (3.14)	3 (1.05)	39 (13.59)	25 (8.71)	2 (0.70)	1 (0.35)	2 (0.70)	99 (34.49)	6 (2.09)	1 (0.35)
Luminal A n= 909	40 (4.40)	17 (1.87)	31 (3.41)	26 (2.86)	5 (0.55)	2 (0.22)	2 (0.22)	103 (11.33)	5 (0.55)	3 (0.33)
Luminal B n= 590	19 (3.22)	5 (0.85)	99 (16.78)	50 (8.47)	3 (0.51)	2 (0.34)	4 (0.68)	159 (26.95)	2 (0.34)	1 (0.17)
Normal Like n= 179	4 (2.23)	3 (1.68)	10 (5.59)	7 (3.91)	0 (0.00)	1 (0.56)	0 (0.00)	23 (12.85)	1 (0.56)	0 (0.00)

Data was derived from METABRIC and TCGA samples with available PAM50 subtype scores (n=2363). Percentages in brackets.