

Article

Colocalised Genetic Associations Reveal Alternative Splicing Variants as Candidate Causal Links for Breast Cancer Risk in 10 Loci

André Besouro-Duarte ^{1,2}, Beatriz Carrasqueiro ², Sofia Sousa ², Joana M. Xavier ^{1,3} and Ana-Teresa Maia ^{1,2,3,*}

¹ CINTESIS@RISE, Universidade do Algarve, 8005-139 Faro, Portugal; a65440@ualg.pt (A.B.-D.); jgxavier@ualg.pt (J.M.X.)

² Faculty of Medicine and Biomedical Sciences, Gambelas Campus, Universidade do Algarve, 8005-139 Faro, Portugal

³ Centro de Ciências do Mar (CCMAR), Universidade do Algarve, 8005-139 Faro, Portugal

* Correspondence: atmaia@ualg.pt

Simple Summary: Hundreds of common genetic variants have been linked to breast cancer, but their exact mechanisms of action remain unclear. Understanding these mechanisms could lead to better prevention strategies and improved survival rates. Our study focused on how these variants influence splicing—a process by which a gene’s coding elements are rearranged to produce different proteins. By analysing data from healthy breast tissue, we identified 43 variants within twelve genes associated with both alternative splicing and breast cancer risk. We then used advanced computational tools and existing experimental data to explore the biological significance of these findings.

Abstract: Genome-wide association studies (GWASs) have revealed numerous loci associated with breast cancer risk, yet the precise causal variants, their impact on molecular mechanisms, and the affected genes often remain elusive. We hypothesised that specific variants exert their influence by affecting cis-regulatory alternative splice elements. An analysis of splicing quantitative trait loci (sQTL) in healthy breast tissue from female individuals identified multiple variants linked to alterations in splicing ratios. Through colocalisation analysis, we pinpointed 43 variants within twelve genes that serve as candidate causal links between sQTL and GWAS findings. In silico splice analysis highlighted a potential mechanism for three genes—*FDPS*, *SGCE*, and *MRPL11*—where variants in proximity to or on the splice site modulate usage, resulting in alternative splice transcripts. Further in vitro/vivo studies are imperative to fully understand how these identified changes contribute to breast oncogenesis. Moreover, investigating their potential as biomarkers for breast cancer risk could enhance screening strategies and early detection methods for breast cancer.

Keywords: breast cancer risk; genome-wide association studies; alternative splicing; splice quantitative trait loci; colocalisation analysis



Citation: Besouro-Duarte, A.; Carrasqueiro, B.; Sousa, S.; Xavier, J.M.; Maia, A.-T. Colocalised Genetic Associations Reveal Alternative Splicing Variants as Candidate Causal Links for Breast Cancer Risk in 10 Loci. *Cancers* **2024**, *16*, 3020. <https://doi.org/10.3390/cancers16173020>

Academic Editor: Carlos S. Moreno

Received: 1 July 2024

Revised: 23 August 2024

Accepted: 27 August 2024

Published: 29 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Genome-wide association studies (GWASs) have pinpointed numerous variants linked to alterations in breast cancer risk, predominantly situated in non-coding regions of the genome [1,2]. Some of these variants have been identified as regulators of gene expression mechanisms through SNPs in enhancer/promoter regions or miRNA that modulate transcript stability and translation [3,4]. Nevertheless, the specific mechanisms underlying most of these associations remain unclear.

Alternative splicing (AS) is a process that enables a single gene to generate multiple proteins by splicing in/out functional domains. Approximately 70% of human genes exhibit alternative splice forms, expanding the repertoire of functional transcripts and proteins by adding/removing functional blocks [5,6]. Variations in splicing regulatory

elements can modify motif recognition by RNA-binding proteins, leading to differences in isoform ratios or even the creation/destruction of motifs, generating new isoforms through cryptic splicing or mis-splicing [7,8]. Furthermore, splicing is specific to tissue/cell types and developmental stages, resulting in distinct phenotypes that necessitate independent analysis [9].

Previous research demonstrated the influence of AS on various diseases, including cancer [8], where splicing changes can act as oncogenic drivers [10–12]. In cases with *BRCA1* germline mutations, cis-regulating mutations modulate AS, leading to its inactivation—a feature often replicated in tumours [13,14]. Another instance is *ESR1*, which encodes for the oestrogen receptor; alterations in the ratios of alternatively spliced isoforms are associated with an elevated cancer risk [15]. Another two studies attempted to establish more systematic associations between GWAS variants and alternative splicing but did not evaluate AS in the normal tissue, the most relevant for risk studies, and carried out sQTL in a small number of variants [16,17].

This paper proposes to comprehensively assess the impact of alternative splicing cis-regulatory variants on breast cancer risk by leveraging previously collected genomic data and employing robust analysis methods. Specifically, we aim to quantify alternative splicing in healthy breast tissue, identify variants associated with this mechanism, colocalise these variants with previously published GWAS findings on breast cancer risk to identify common variants, and conduct *in silico* analyses to propose candidate causal variants and elucidate the impacted molecular mechanisms.

2. Materials and Methods

2.1. Data Sources

RNA-seq SRA and WGS-based VCF files were obtained from the GTEx v8.p2 study (dbGAP Study accession phs000424.GTEX.v8.p2). Anonymised sample attributes, phenotypes, and transcript TPMs were downloaded from the GTEx Portal. Data access was authorised on 7 March 2018, with retrieval in April 2018.

GWAS data for breast carcinoma traits were sourced in June 2020 using the *gwas-rapidd* R package, leveraging the NHGRI-EBI GWAS Catalog’s REST API [18,19]. An EFO-based trait search was performed using the keyword “breast carcinoma”, and additional sub-traits were manually curated and filtered out: survival, prognosis, toxicity, therapy, density, response, radiotherapy, treat, induce, mortality, and signature. Further filtering was performed for the “European population” in either discovery or validation sample descriptions, yielding 41 studies detailing 1080 unique risk-associated variants published until 17 February 2021 (Supplementary Table S1).

2.2. Ancestry Analysis

We conducted Principal Component Analysis using PLINK2 (v2.00a3LM) and 1000 Genomes (1000 G) data, focusing on European ancestry [20]. VCF file processing utilised BCFTools (v1.10.2 with htlib1.10.2-3) for variant extraction and merging GTEx and 1000 G data [21,22]. As previously reported, we used the top four principal components (PCs) to improve ancestry granularity [23], classifying 711 individuals from the GTEx dataset as Europeans, with 119 having breast tissue RNA-seq data available.

2.3. RNA-Seq Analysis

SRA files from 119 normal breast tissue samples from female donors classified as European were converted from to FASTQ using *sratoolkit* (v2.10.8) [24], followed by quality control with *FastQC* (v0.11.9), *multiQC* (v1.9) [25], and *cutadapt* (v2.8) [26]. *STAR* (v2.7.5a) [27] aligned reads to the hg38 assembly [28], achieving alignment rates between 85.8% and 93.4%, resulting in 26.2 M to 62.7 M reads mapped per sample (Supplementary Figure S1).

2.4. Alternative Splicing Quantification

Psichomics (v1.16.0) [29] was used to quantify alternative splicing, with a minimum threshold of 10 reads per exon junction. The package maintainer provided event annotations. We only considered events with quantification for all samples and variability higher than 0 for further analysis.

2.5. sQTL Analysis

TensorQTL [30,31] was used for sQTL mapping in a ± 1 Mb window from the 5' boundary of the alternative event (A5SS), with the top 15 PCs as covariates. We conducted both nominal sQTL association, which utilises Pearson correlation and reports p -values for all tested variant–event pairs, and empirical-beta approximation, which uses a beta-approximated distribution based on permutation testing to report the best-associated variant for each event. p -values from both approaches were corrected for multiple testing using the Benjamini–Hochberg method (FDR), setting significant association at $FDR \leq 0.05$.

sGenes from GTEx v.8 were retrieved and compared to those identified by our sQTL analysis. The comparison was not limited to genes analysed in both datasets due to the lack of information on non-significant genes in the GTEx dataset.

2.6. Linkage Disequilibrium Analysis

Linkage Disequilibrium (LD) analysis between sQTL and GWAS risk variants used LDlinkR (assembly GRCh37) with an r^2 threshold ≥ 0.4 , based on European super populations (including CEU, TSI, FIN, GBR, and IBS) [32,33].

2.7. Colocalisation Analysis

Selected loci were tested using the SuSiE implementation of COLOC [34]. Variants found both to be sQTLs in breast tissue and associated with risk for breast cancer in published GWASs were selected, and an LD matrix was retrieved using LDlinkR [33]. Variants in perfect SNPs ($r^2 = 1$) were pruned, keeping the variant with the highest combined value of $\log p$ -values as suggested by COLOC authors, except for *BANF1* variants for which pruning was performed for variants with $r^2 \geq 0.98$ to allow computation in a very complex LD structure locus. SNP-wise priors were set at 1×10^{-4} for p_1 and p_2 and 1×10^{-5} for p_{12} and the colocalisation hypothesis at 0.9894755 for H_0 , 5×10^{-3} for H_1 and H_2 , 2.45×10^{-5} for H_3 , and 5×10^{-5} for H_4 . True colocalisation was criteria was (1) $PP(H_4) \geq 0.9$ and (2) $PP(H_4) \geq 3 \times PP(H_3)$.

2.8. In Silico Splicing Analysis

Variants identified as colocalised together with their proxies ($r^2 \geq 0.8$ in the European superpopulation) were annotated regarding gene location using BiomaRt (build GRCh38.p13) [35]. Variants within genes were assessed for potential impacts in splicing via in silico analysis.

NetGene2 [36] and SpliceAI [9] were used to detect changes in primary splicing elements. NetGene2 analysed a 201-nucleotide sequence from Ensembl, focusing on the variant at position 101, to record allele score changes. SpliceAI's Online API evaluated score deltas within 50 nucleotides of the variant on hg38. The destruction of splice site predictions was only considered for variants located within 50 bps of the splice site, while predictions on creating new splice sites were considered for all variants. Encore [37], Postar3 [38], and RBP-map [39] were applied to identify changes in auxiliary splicing elements, specifically RNA-binding protein (RBP) motifs. Encore and Postar3 queries targeted genes of interest. RBP-map processed sequences of 221 nucleotides for all human RBPs to predict binding changes. HSF, composed of multiple tools, was used to predict the overall impact of changes in primary sequence on splicing [40] based on rsID input. Additionally, both SpliceAI and HSF provided a threshold to establish significant differences in the prediction of splicing impact between the two alleles of the variant.

2.9. eQTL Analysis

Transcript-wise eQTL analysis involved GTEx data and gene boundary information from Ensembl using BiomaRt [35], with the top 15 PCs on normalised transcript-wise counts used as covariates. Results were considered significant when $FDR \leq 0.05$.

3. Results

To investigate the effect of alternative splicing cis-regulatory variants on breast cancer risk, we initiated our analysis with a genome-wide sQTL analysis in healthy breast tissue data, followed by a colocalisation analysis with previously reported GWAS risk-associated variants, and finishing with an in silico functional analysis (Figure 1).

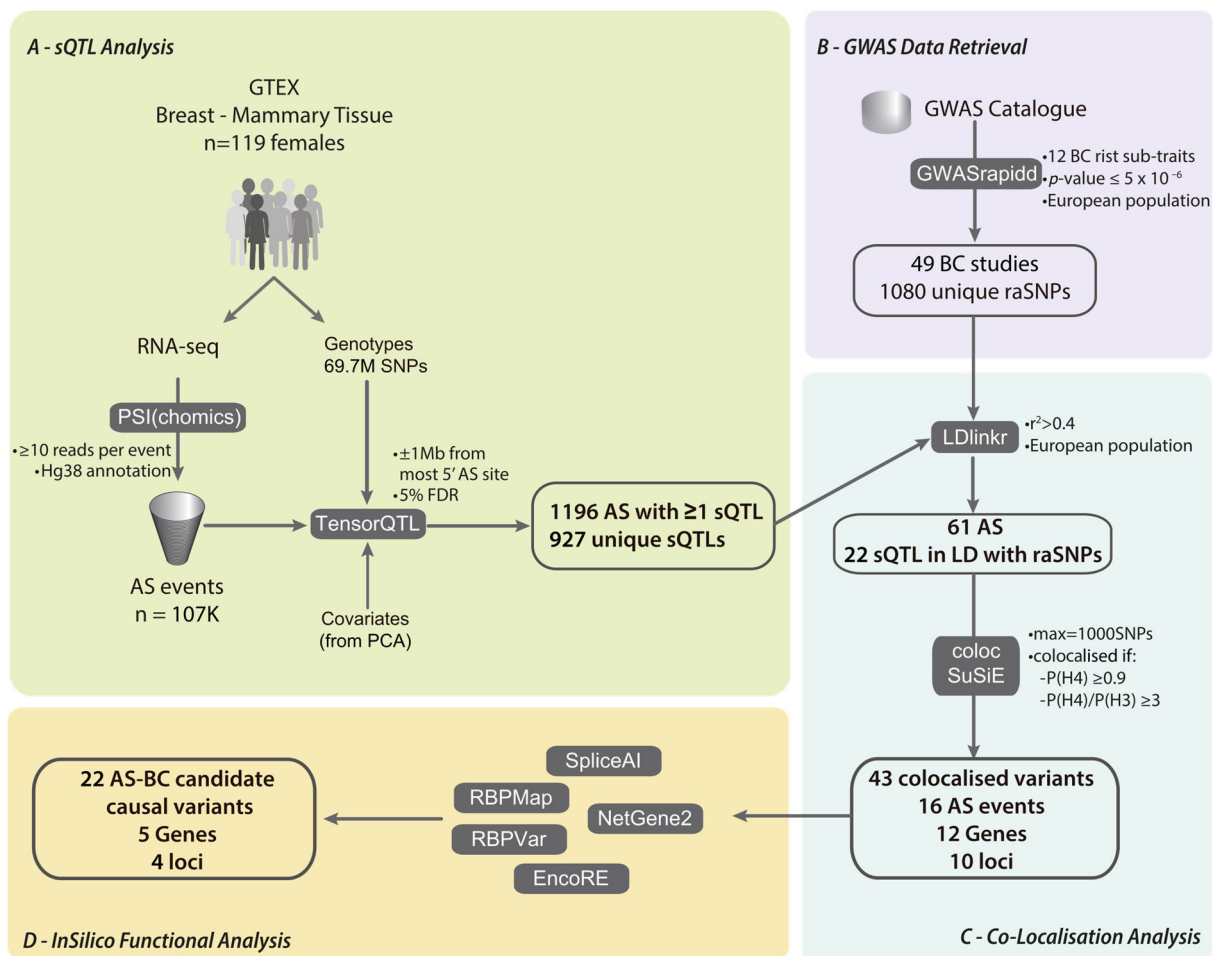


Figure 1. Analysis pipeline. (A) Quantification of alternative splicing events and sQTL mapping using data from GTEx project. (B) Retrieval of previously identified breast cancer risk-associated variants. (C) Colocalisation analysis to identify variants associated. (D) In silico functional analysis of putative causal variants.

3.1. Splicing QTL Hints at Underlying Gene Regulation

To reveal the effect of variants acting on alternative-splicing equilibrium in breast tissue, we performed sQTL mapping. We first quantified alternative splicing on 107,000 events across 119 samples, identifying 29,993 informative events. PSI at each event was then tested for associations with genotypes, linking over 152,000 genetic variants to changes in splicing (sQTLs) (5% FDR) in 3978 genes (sGenes). Some variants were associated with more than one event type, yielding over 241,000 unique variant–splice event pairs (Figure 2A, Supplementary Table S2). A total of 26% of the sGenes we identified overlapped with

fied 13 loci harbouring top sQTLs and breast cancer risk variants in LD (Supplementary Table S4), but one was dropped due to a lack of GWAS summary statistics.

Colocalisation analysis pinpointed 43 variants on twelve genes located in ten different loci (Supplementary Table S5), including the highly complex locus 1q22 where *FDPS*, *SCAMP3*, and *YY1AP1* share several colocalised SNPs (Figure 2D).

3.3. Linking Risk Variants to Mechanism at Three Loci

To pinpoint the candidate causal variant within each locus, we assessed the potential effect on splicing of the 133 variants (colocalised or in high LD with these) located within gene boundaries (Supplementary Table S6).

We identified 85 variants that showed evidence of splice site modulation. More specifically, four variants, located within *MRPL11*, *SGCE*, *FDPS*, and *YY1AP1*, were predicted to impact splicing elements (5' and 3' splice sites and the branching point) and were in close proximity (<50 bp) to the alternative splice element, further supporting their regulatory potential. Seventy-three variants were predicted to alter splice auxiliary elements (exonic and intronic splice enhancers and silencers), and 40 variants were found to modify RBP recognition (Supplementary Table S7).

Three variants (rs11264361, rs10247562, rs11110) showed multiple lines of evidence of altering splice elements and were analysed in more detail.

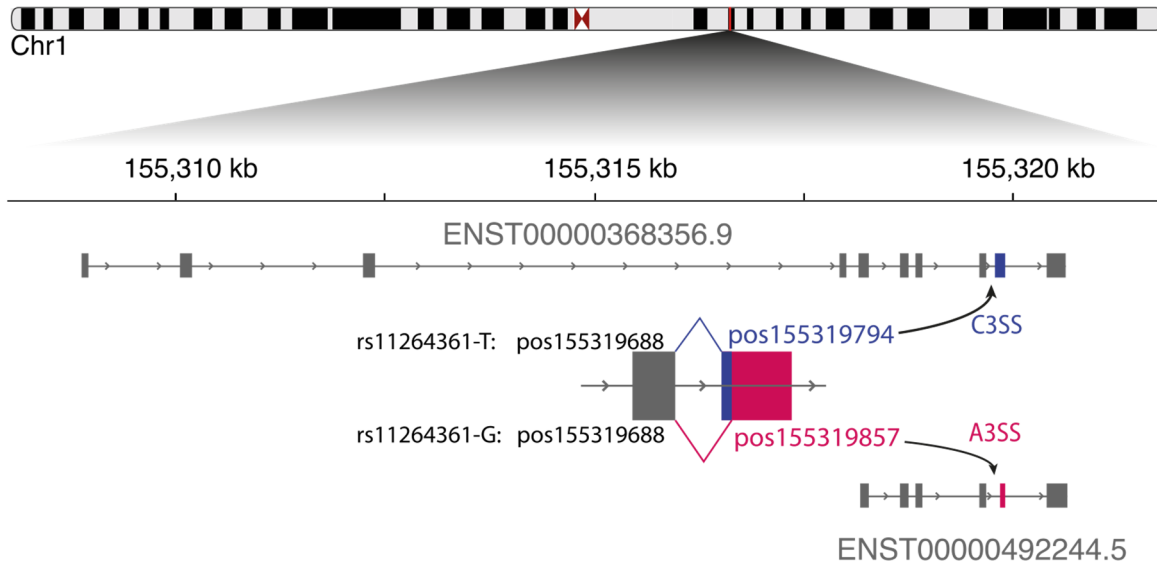
The minor allele (allele G) of rs11264361, located in intron 8 of *FDPS*, was predicted to break a branching point in exon 9, changing the 3' splice site from position 155,319,794 to position 155,319,857 (Figure 3A). This variant is in strong LD ($r^2 = 0.81$) with the sQTL rs12042020, whose minor allele G we found associated with a significant increase in the ratio of alternative 3' splice site usage for the same event (slope = 0.009, p -value = 1.07×10^{-7}) (Figure 3B, Supplementary Table S4) and was also predicted to disrupt crucial splicing mechanisms, impacting the spliceosome component U2AF and altering splicing ratios (Figure 3C). eQTL analysis corroborated these results, indicating that rs11264361 was also an eQTL for specific *FDPS* isoforms (Supplementary Figure S7). rs11264361 was in LD with rs12091730 ($r^2 = 0.67$), whose minor allele was associated with an increased risk for breast cancer [1]. None of these changes in alternative splicing have been reported in breast tumours [41].

Similarly, the major allele G of rs10247562, an exonic variant of *SGCE*, was predicted to decrease the exon usage where it is located (Supplementary Table S7). This variant was in LD ($r^2 = 0.98$) with the sQTL rs11508502, whose major allele was found to be associated with an increase in the exon skipping (slope = -0.014 , p -value = 5.61×10^{-18}) (Supplementary Table S4). Furthermore, rs10247562 showed RBP binding (PIP-Seq in He-La cells), and in silico analysis predicted a change in auxiliary splicing sequences with disruption of RBP binding (Supplementary Table S7). rs10247562 was in high LD with rs17268829 ($r^2 = 0.91$), whose major allele was associated with protection for BC [1,42]. Interestingly, breast tumours have shown higher inclusion of exon 12 than normal tissue (Supplementary Figure S8A) [41]. This finding aligns with our results, where lower inclusion of exon 12 was associated with protection against breast cancer.

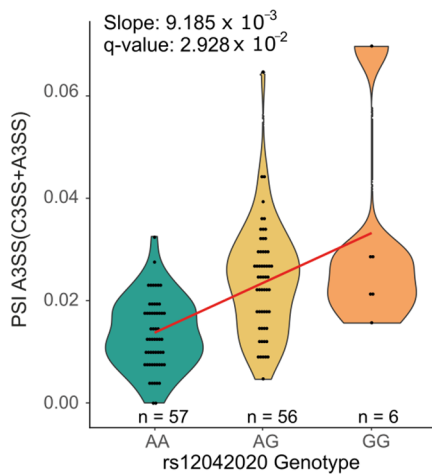
Lastly, we found that rs11110, an exonic variant of *MRPL11* located in an auxiliary splicing sequence (enhancer or silencer), is predicted to influence the use of a 5' splice site in *MRPL11* within the same exon, resulting in the inclusion of the variant site in the final isoform when the minor T allele is present (Supplementary Table S7). Concomitantly, this variant was among the top sQTLs (Supplementary Table S3), with the T allele associated with increased usage of the same 5' splice site in *MRPL11* (slope = -0.088 , p -value = 1.42×10^{-25}). Additionally, CLIP data show the binding of various RBPs at the variant site in multiple cell lines (Hek293, K562, HepG2), with the major allele C showing higher affinity for these RBPs (Supplementary Table S7). This variant was in LD with rs1134495 and rs7570 ($r^2 = 0.41$ and 0.47 , respectively), whose minor alleles have been reported to be associated with risk for breast cancer [1,42]. These findings are discor-

dant from previously reported preferential usage of the longer exon 1 in tumour tissue (Supplementary Figure S8B) [41].

A



B



C



Figure 3. Summary of the evidence collected linking rs11264361 to the splicing regulation of FDPS. (A) Gene model of FDPS showing the allele-specific effect of rs11264361 on the final isoform. (B) sQTL for rs12042020, the best FDPS-associated variant, at the event reported in panel A. (C) Consensus motif for U2 snRNA, variant, and flanking sequence at the rs11264361 variant site.

4. Discussion

Given the excess of intronic and intergenic breast cancer risk-associated variants identified in GWASs [4], previous studies have intended to establish a link between risk-associated variants and alternative splicing. However, they were limited by considering alternative splicing events in tumour tissue rather than normal [16,17], focusing on a single variant [15] or a small set [17], and not establishing causality by lack of colocalisation analysis [16,17].

This study provides the first compelling evidence that genetic variants contribute to breast cancer risk by impacting on alternative splicing in the normal breast. We identified ten loci where the top sQTL variant colocalises with previously reported breast cancer risk variants, suggesting that splicing modulation may be the mechanism driving risk at these

loci. Additionally, we identified rs11264361, rs10247562, and rs11110 as candidate causal variants that affect the splicing of the genes *FDPS*, *SGCE*, and *MRPL11*, respectively, in this context.

Our analysis began with an sQTL analysis identifying variants associated with splicing modulation for 3978 genes. Due to the prevalence of breast cancer GWASs in European ancestry populations [43], our analysis focused on individuals of this same ancestry. Furthermore, we exclusively analysed RNA-seq data from healthy female breast tissue as hormones, like oestrogen and progesterone, may alter gene expression regulation between genders [44] and because of the higher incidence of breast cancer in females [45].

We revealed a significant portion of unique sGenes in our study compared to published sGenes for GTEx [46], underscoring the importance of ancestry and gender in gene expression analysis. The comparison was not limited to genes analysed in both datasets, and our use of annotation-based tools for detecting alternative splicing (different from those used by GTEx) may have missed certain events, such as cryptic splice sites. This highlights the need for more comprehensive methods, like full-length mRNA sequencing, when performing such comparisons [28,47]. In addition, the presence of variants in trans, affecting RBPs expression, may also be an important factor to take into account in a further study.

Our findings show that the majority of sQTLs are located within 170 Kb of the associated splicing event and outside gene boundaries and beyond typically considered ranges in previous studies (5 Kb from the gene boundary or 100 Kb from the tested event) [46,48], suggesting the existence of distant regulatory elements affecting splicing, such as transcript-specific promoters/enhancers, and mechanisms related to chromatin accessibility and RNA polymerase II, as previously reported [28,48–51]. However, we cannot exclude the fact that the sQTLs we identified may be capturing the effect of high-LD regulatory variants within gene boundaries. The distribution of sQTLs by event type suggests differing underlying mechanisms. Mutually exclusive exon events appear to be more sensitive to proximal variants, while alternative first-exon events are more influenced by distant regulatory elements.

Our colocalisation analysis identified ten loci where variants cis-regulating splicing are candidate causal variants for breast cancer risk. This indicates that while splicing regulation may not be the most common mechanism driving risk, it is significant and should not be overlooked. Our results may even underestimate the impact of this mechanism because the colocalisation analysis was limited to loci with available GWAS summary statistics, excluding other loci where sQTLs were identified, such as *NR1H3*. In addition, extending this study to include other populations may yield broader insights into susceptibility to breast cancer-specific subtypes.

Some of the genes in the colocalised loci, including the ones functionally characterised, have been previously linked to tumour development. For instance, *FDPS* is involved in cholesterol biosynthesis and prenylation, processes vital for cell functions often exploited by cancer cells [52,53]. *SGCE*'s role in cell structure and signalling suggests its involvement in cancer progression [54,55], while *MRPL11*'s involvement in mitochondrial metabolism points to its potential role in cancer biology [56,57]. These changes are often linked to oncogenic processes, highlighting the potential for targeted therapies using transcript-specific RNAi or serving as neoepitopes, a target for immunotherapy.

The complexity of splicing and its context-dependent nature underscores the need for more targeted approaches in understanding and modelling this mechanism. This requires advanced techniques like eCLIP experiments on relevant tissue, such as breast, to improve our understanding of RBPs' roles in splicing and breast cancer risk. Hence, the addition of experimental validation in future studies will support the importance of the variants identified herein and their inclusion in risk assessment tools.

5. Conclusions

Our study sheds new light on how cis-regulatory variants involved in alternative splicing contribute to the increased risk of breast cancer. By identifying twelve genes within ten loci where top sQTL variants colocalise with previously reported breast cancer risk variants, we implicate splicing modulation as a key mechanism driving risk at these loci. Nevertheless, limitations in data availability prevented comprehensive colocalisation analysis for some genes associated with risk for breast cancer and for which sQTL were identified. In the future, analysis using updated GWAS, RNA-seq, and genotyping data, including data from different populations, and experimental validation will provide definite support for our findings.

Overall, our study emphasises the importance of exploring splicing regulation in the context of cancer risk, mainly through a tissue- and gender-specific lens.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/cancers16173020/s1>: Supplementary Data include seven tables and eight figures. Supplementary Table S1: List of GWAS retrieved from the GWAS Catalog. study_id—unique identifier assigned by GWAS Catalog, pubmed_id—PubMed identification number, publication_date—original publication date, publication_title—Title of paper, author_fullname—first author full name, author_orcid—first author orcid. Supplementary Table S2: List of significant sQTLs. phenotype_id—identification of alternative splicing events, variant_id—position-based identification of associated variant, tss_distance—distance between 5′ most splice site to associated variant, af—allele frequency, ma_samples—number of samples with the minor allele, ma_count—total count of minor alleles across all samples, pval_nominal—The nominal p-value of the association between the variant and the phenotype, slope—The beta (slope) of the linear regression, slope_se—The standard error of the beta, FDR_pval—FDR corrected p-value of association, log10FDR—negative logarithmic transformation of the FDR corrected p-value. Supplementary Table S3: Significant sQTLs using a beta-approximated distribution based on permutation testing. Supplementary Table S4: Variants of interest per loci and event. Supplementary Table S5: Colocalisation of sQTL and GWAS from summary statistics. Supplementary Table S6: Variants Colocalized or in high LD of interest. Supplementary Table S7: In silico Splice Analysis of variants of interest. Supplementary Figure S1: RNA-seq alignment rates of normal breast samples from the GTEx project. Supplementary Figure S2: sGenes overlapped with GTEx breast tissue analysis. Supplementary Figure S3: Relative position of each alternative splicing associated variant relative to the gene and the alternative splice event, corrected for strand. Supplementary Figure S4: Distribution of distances from alternative splice 5′ most edge to associated variant. Supplementary Figure S5: Distribution of the absolute slope value across negative logarithm of the p-value. Supplementary Figure S6: Example of an extreme effect, where alternative allele rs12898397-C content is associated with a decrease in PSI levels of 0.5125, depleting the reference splice pattern in favour of the alternative. Supplementary Figure S7: Transcript-wise QTL of FDPS at rs11264361. p-value and slope are provided for significant QTLs. Supplementary Figure S8: Changes in splicing between breast tumour and normal-matched tissue for the alternative splice events of interest.

Author Contributions: Conceptualisation, J.M.X. and A.-T.M.; methodology, A.B.-D.; formal analysis, A.B.-D., B.C. and S.S.; writing—original draft preparation, A.B.-D.; writing—review and editing, A.B.-D., J.M.X. and A.-T.M.; supervision, J.M.X. and A.-T.M.; funding acquisition, J.M.X. and A.-T.M. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by national Portuguese funding through FCT-Fundação para a Ciência e a Tecnologia and CRESC Algarve 2020: POCI-01-0145-FEDER-022184—“GenomePT”, PTDC/MED-GEN/30895/2017, the contract DL 57/2016/CP1361/CT0042 (J.M.X.), and within the scope of the project RISE—LA/P/0053/2020 and the R&D Units funding UIDB/4255/2020—CINTESIS and UIDB/04326/2020—CCMAR.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data generated in this study are provided as Supplemental Data, and the code can be made available upon request to the corresponding author. GWAS summary

statistics were obtained from the GWAS Catalog website via FTP download http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST004001-GCST005000/GCST004988/ (accessed on 28 January 2021) and http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST007001-GCST008000/GCST007236/ (accessed on 27 January 2021). SNP data were obtained from the Ensembl database (versions 92 and 75) and are available at www.ensembl.org (accessed on 30 June 2024). Gene and transcript expression and eQTL data for breast tissue from the GTEx Project (v7) were retrieved from the GTEx Portal at www.gtexportal.org (accessed on 30 June 2024). Other detailed results are available in Supplementary Data.

Acknowledgments: The authors sincerely appreciate the contributions made by all sample donors who allowed the datasets to be created and used for the good of medical care and science. The authors would also like to thank other Functional Genomics of Cancer group members at CINTESIS@RISE-UAIG for helpful discussions and Julieta Costa at UAIC for administrative support.

Conflicts of Interest: The authors declare no conflicts of interests.

References

1. Michailidou, K.; Lindström, S.; Dennis, J.; Beesley, J.; Hui, S.; Kar, S.; Lemaçon, A.; Soucy, P.; Glubb, D.; Rostamianfar, A.; et al. Association Analysis Identifies 65 New Breast Cancer Risk Loci. *Nature* **2017**, *551*, 92–94. [[CrossRef](#)]
2. Welter, D.; MacArthur, J.; Morales, J.; Burdett, T.; Hall, P.; Junkins, H.; Klemm, A.; Flicek, P.; Manolio, T.; Hindorff, L.; et al. The NHGRI GWAS Catalog, a Curated Resource of SNP-Trait Associations. *Nucleic Acids Res.* **2014**, *42*, D1001–D1006. [[CrossRef](#)]
3. Xavier, J.M.; Magno, R.; Russell, R.; de Almeida, B.P.; Jacinta-Fernandes, A.; Duarte, A.; Dunning, M.; Samarajiwa, S.; O’Reilly, M.; Rocha, C.L.; et al. Mapping of Cis-Regulatory Variants by Differential Allelic Expression Analysis Identifies Candidate Risk Variants and Target Genes of 27 Breast Cancer Risk Loci. *medRxiv* **2022**. [[CrossRef](#)]
4. Jacinta-Fernandes, A.; Xavier, J.M.; Magno, R.; Lage, J.G.; Maia, A.T. Allele-Specific miRNA-Binding Analysis Identifies Candidate Target Genes for Breast Cancer Risk. *NPJ Genom. Med.* **2020**, *5*, 4. [[CrossRef](#)]
5. Kalniņa, Z.; Zayakin, P.; Siliņa, K.; Line, A. Alterations of Pre-mRNA Splicing in Cancer. *Genes Chromosomes Cancer* **2005**, *42*, 342–357. [[CrossRef](#)] [[PubMed](#)]
6. Baralle, F.E.; Giudice, J. Alternative Splicing as a Regulator of Development and Tissue Identity. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 437–451. [[CrossRef](#)] [[PubMed](#)]
7. Kurtovic-Kozaric, A.; Przychodzen, B.; Singh, J.; Konarska, M.M.; Clemente, M.J.; Otrrock, Z.K.; Nakashima, M.; Hsi, E.D.; Yoshida, K.; Shiraishi, Y.; et al. PRPF8 Defects Cause Missplicing in Myeloid Malignancies. *Leukemia* **2015**, *29*, 126–136. [[CrossRef](#)]
8. Scotti, M.M.; Swanson, M.S. RNA Mis-Splicing in Disease. *Nat. Rev. Genet.* **2016**, *17*, 19–32. [[CrossRef](#)] [[PubMed](#)]
9. Jaganathan, K.; Panagiotopoulou, S.K.; McRae, J.F.; Darbandi, S.F.; Knowles, D.; Li, Y.I.; Kosmicki, J.A.; Arbelaez, J.; Cui, W.; Schwartz, G.B.; et al. Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **2019**, *176*, 535–548.e24. [[CrossRef](#)]
10. Zhang, J.; Manley, J.L. Misregulation of Pre-mRNA Alternative Splicing in Cancer. *Cancer Discov.* **2013**, *3*, 1228–1237. [[CrossRef](#)]
11. Climente-González, H.; Porta-Pardo, E.; Godzik, A.; Eyras, E. The Functional Impact of Alternative Splicing in Cancer. *Cell Rep.* **2017**, *20*, 2215–2226. [[CrossRef](#)] [[PubMed](#)]
12. Tian, J.; Chen, C.; Rao, M.; Zhang, M.; Lu, Z.; Cai, Y.; Ying, P.; Li, B.; Wang, H.; Wang, L.; et al. Aberrant RNA Splicing Is a Primary Link between Genetic Variation and Pancreatic Cancer Risk. *Cancer Res.* **2022**, *82*, 2084–2096. [[CrossRef](#)]
13. Li, D.; Harlan-Williams, L.M.; Kumaraswamy, E.; Jensen, R.A. BRCA1—No Matter How You Splice It. *Cancer Res.* **2019**, *79*, 2091–2098. [[CrossRef](#)]
14. de la Hoya, M.; Soukariéh, O.; López-Perolio, I.; Vega, A.; Walker, L.C.; van Ierland, Y.; Baralle, D.; Santamariña, M.; Lattimore, V.; Wijnen, J.; et al. Combined Genetic and Splicing Analysis of BRCA1 c.[594-2A>C; 641A>G] Highlights the Relevance of Naturally Occurring in-Frame Transcripts for Developing Disease Gene Variant Classification Algorithms. *Hum. Mol. Genet.* **2016**, *25*, 2256–2268. [[CrossRef](#)]
15. Dunning, A.M.; Healey, C.S.; Baynes, C.; Maia, A.-T.T.; Scollen, S.; Vega, A.; Rodríguez, R.; Barbosa-Morais, N.L.; Ponder, B.A.J.; Low, Y.-L.L.; et al. Association of ESR1 Gene Tagging SNPs with Breast Cancer Risk. *Hum. Mol. Genet.* **2009**, *18*, 1131–1139. [[CrossRef](#)] [[PubMed](#)]
16. Tian, J.; Wang, Z.; Mei, S.; Yang, N.; Yang, Y.; Ke, J.; Zhu, Y.; Gong, Y.; Zou, D.; Peng, X.; et al. CancerSplicingQTL: A Database for Genome-Wide Identification of Splicing QTLs in Human Cancer. *Nucleic Acids Res.* **2019**, *47*, D909–D916. [[CrossRef](#)] [[PubMed](#)]
17. Caswell, J.L.; Camarda, R.; Zhou, A.Y.; Huntsman, S.; Hu, D.; Brenner, S.E.; Zaitlen, N.; Goga, A.; Ziv, E. Multiple Breast Cancer Risk Variants Are Associated with Differential Transcript Isoform Expression in Tumors. *Hum. Mol. Genet.* **2015**, *24*, 7421–7431. [[CrossRef](#)] [[PubMed](#)]
18. Magno, R.; Maia, A.-T. Gwasrapidd: An R Package to Query, Download and Wrangle GWAS Catalog Data. *Bioinformatics* **2020**, *36*, 649–650. [[CrossRef](#)]
19. Sollis, E.; Mosaku, A.; Abid, A.; Buniello, A.; Cerezo, M.; Gil, L.; Groza, T.; Güneş, O.; Hall, P.; Hayhurst, J.; et al. The NHGRI-EBI GWAS Catalog: Knowledgebase and Deposition Resource. *Nucleic Acids Res.* **2023**, *51*, D977–D985. [[CrossRef](#)]
20. 1000 Genome Project Consortium. A Global Reference for Human Genetic Variation. *Nature* **2015**, *526*, 68–74. [[CrossRef](#)]

21. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve Years of SAMtools and BCftools. *Gigascience* **2021**, *10*, giab008. [[CrossRef](#)] [[PubMed](#)]
22. Bonfield, J.K.; Marshall, J.; Danecek, P.; Li, H.; Ohan, V.; Whitwham, A.; Keane, T. HTSlib: C Library for Reading/Writing High-Throughput Sequencing Data. *Gigascience* **2021**, *10*, giab007. [[CrossRef](#)] [[PubMed](#)]
23. Gay, N.R.; Gloudemans, M.; Antonio, M.L.; Abell, N.S.; Balliu, B.; Park, Y.; Martin, A.R.; Musharoff, S.; Rao, A.S.; Aguet, F.; et al. Impact of Admixture and Ancestry on EQTL Analysis and GWAS Colocalization in GTEx. *Genome Biol.* **2020**, *21*, 233. [[CrossRef](#)] [[PubMed](#)]
24. SRA-Tools—NCBI. Available online: <https://github.com/ncbi/sra-tools/> (accessed on 23 September 2020).
25. Ewels, P.; Magnusson, M.; Lundin, S.; Källér, M. MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report. *Bioinformatics* **2016**, *32*, 3047–3048. [[CrossRef](#)]
26. Martin, M. Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet J.* **2011**, *17*, 10–12. [[CrossRef](#)]
27. 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. [[CrossRef](#)]
28. Li, Y.I.; Knowles, D.A.; Humphrey, J.; Barbeira, A.N.; Dickinson, S.P.; Im, H.K.; Pritchard, J.K. Annotation-Free Quantification of RNA Splicing Using LeafCutter. *Nat. Genet.* **2017**, *50*, 151–158. [[CrossRef](#)]
29. Saraiva-Agostinho, N.; Barbosa-Morais, N.L. Psychomics: Graphical Application for Alternative Splicing Quantification and Analysis. *Nucleic Acids Res.* **2019**, *47*, e7. [[CrossRef](#)] [[PubMed](#)]
30. Ongen, H.; Buil, A.; Brown, A.A.; Dermitzakis, E.T.; Delaneau, O. Fast and Efficient QTL Mapper for Thousands of Molecular Phenotypes. *Bioinformatics* **2015**, *32*, 1479–1485. [[CrossRef](#)]
31. Taylor-Weiner, A.; Aguet, F.; Haradhvala, N.J.; Gosai, S.; Anand, S.; Kim, J.; Ardlie, K.; Van Allen, E.M.; Getz, G. Scaling Computational Genomics to Millions of Individuals with GPUs. *Genome Biol.* **2019**, *20*, 228. [[CrossRef](#)]
32. Machiela, M.J.; Chanock, S.J. LDlink: A Web-Based Application for Exploring Population-Specific Haplotype Structure and Linking Correlated Alleles of Possible Functional Variants. *Bioinformatics* **2015**, *31*, 3555–3557. [[CrossRef](#)]
33. Myers, T.A.; Chanock, S.J.; Machiela, M.J. LDlinkR: An R Package for Rapidly Calculating Linkage Disequilibrium Statistics in Diverse Populations. *Front. Genet.* **2020**, *11*, 157. [[CrossRef](#)] [[PubMed](#)]
34. Wallace, C. A More Accurate Method for Colocalisation Analysis Allowing for Multiple Causal Variants. *PLoS Genet.* **2021**, *17*, e1009440. [[CrossRef](#)] [[PubMed](#)]
35. Durinck, S.; Moreau, Y.; Kasprzyk, A.; Davis, S.; De Moor, B.; Brazma, A.; Huber, W. BioMart and Bioconductor: A Powerful Link between Biological Databases and Microarray Data Analysis. *Bioinformatics* **2005**, *21*, 3439–3440. [[CrossRef](#)]
36. Brunak, S.; Engelbrecht, J.; Knudsen, S. Prediction of Human mRNA Donor and Acceptor Sites from the DNA Sequence. *J. Mol. Biol.* **1991**, *220*, 49–65. [[CrossRef](#)] [[PubMed](#)]
37. Van Nostrand, E.L.; Freese, P.; Pratt, G.A.; Wang, X.; Wei, X.; Xiao, R.; Blue, S.M.; Chen, J.Y.; Cody, N.A.L.; Dominguez, D.; et al. A Large-Scale Binding and Functional Map of Human RNA-Binding Proteins. *Nature* **2020**, *583*, 711–719. [[CrossRef](#)]
38. Zhao, W.; Zhang, S.; Zhu, Y.; Xi, X.; Bao, P.; Ma, Z.; Kapral, T.H.; Chen, S.; Zagrovic, B.; Yang, Y.T.; et al. POSTAR3: An Updated Platform for Exploring Post-Transcriptional Regulation Coordinated by RNA-Binding Proteins. *Nucleic Acids Res.* **2022**, *50*, D287–D294. [[CrossRef](#)] [[PubMed](#)]
39. Paz, I.; Kosti, I.; Ares, M.; Cline, M.; Mandel-Gutfreund, Y. RBPmap: A Web Server for Mapping Binding Sites of RNA-Binding Proteins. *Nucleic Acids Res.* **2014**, *42*, 361–367. [[CrossRef](#)] [[PubMed](#)]
40. Desmet, F.-O.O.; Hamroun, D.; Lalande, M.; Collod-Bérout, G.; Claustres, M.; Bérout, C.; Collod-Bérout, G.; Claustres, M.; Bérout, C. Human Splicing Finder: An Online Bioinformatics Tool to Predict Splicing Signals. *Nucleic Acids Res.* **2009**, *37*, e67. [[CrossRef](#)]
41. Ryan, M.; Wong, W.C.; Brown, R.; Akbani, R.; Su, X.; Broom, B.; Melott, J.; Weinstein, J. TCGASpliceSeq a Compendium of Alternative mRNA Splicing in Cancer. *Nucleic Acids Res.* **2016**, *44*, D1018–D1022. [[CrossRef](#)]
42. Shu, X.; Long, J.; Cai, Q.; Kweon, S.-S.; Choi, J.-Y.; Kubo, M.; Park, S.K.; Bolla, M.K.; Dennis, J.; Wang, Q.; et al. Identification of Novel Breast Cancer Susceptibility Loci in Meta-Analyses Conducted among Asian and European Descendants. *Nat. Commun.* **2020**, *11*, 1217. [[CrossRef](#)] [[PubMed](#)]
43. Ju, D.; Hui, D.; Hammond, D.A.; Wonkam, A.; Tishkoff, S.A. Importance of Including Non-European Populations in Large Human Genetic Studies to Enhance Precision Medicine. *Annu. Rev. Biomed. Data Sci.* **2022**, *5*, 321–339. [[CrossRef](#)] [[PubMed](#)]
44. Oliva, M.; Muñoz-Aguirre, M.; Kim-Hellmuth, S.; Wucher, V.; Gewirtz, A.D.H.; Cotter, D.J.; Parsana, P.; Kasela, S.; Balliu, B.; Viñuela, A.; et al. The Impact of Sex on Gene Expression across Human Tissues. *Science* **2020**, *369*, eaba3066. [[CrossRef](#)]
45. Siegel, R.L.; Miller, K.D.; Fuchs, H.E.; Jemal, A. Cancer Statistics, 2022. *CA Cancer J. Clin.* **2022**, *72*, 7–33. [[CrossRef](#)]
46. Garrido-Martín, D.; Borsari, B.; Calvo, M.; Reverter, F.; Guigó, R. Identification and Analysis of Splicing Quantitative Trait Loci across Multiple Tissues in the Human Genome. *Nat. Commun.* **2021**, *12*, 727. [[CrossRef](#)] [[PubMed](#)]
47. Glinos, D.A.; Garborcauskas, G.; Hoffman, P.; Ehsan, N.; Jiang, L.; Gokden, A.; Dai, X.; Aguet, F.; Brown, K.L.; Garimella, K.; et al. Transcriptome Variation in Human Tissues Revealed by Long-Read Sequencing. *Nature* **2022**, *608*, 353–359. [[CrossRef](#)]
48. Walker, R.L.; Ramaswami, G.; Hartl, C.; Mancuso, N.; Gandal, M.J.; de la Torre-Ubieta, L.; Pasaniuc, B.; Stein, J.L.; Geschwind, D.H. Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. *Cell* **2019**, *179*, 750–771.e22. [[CrossRef](#)]

49. Schwartz, S.; Meshorer, E.; Ast, G. Chromatin Organization Marks Exon-Intron Structure. *Nat. Struct. Mol. Biol.* **2009**, *16*, 990–995. [[CrossRef](#)] [[PubMed](#)]
50. Spies, N.; Nielsen, C.B.; Padgett, R.A.; Burge, C.B. Biased Chromatin Signatures around Polyadenylation Sites and Exons. *Mol. Cell* **2009**, *36*, 245–254. [[CrossRef](#)]
51. Herzel, L.; Ottoz, D.S.M.; Alpert, T.; Neugebauer, K.M. Splicing and Transcription Touch Base: Co-Transcriptional Spliceosome Assembly and Function. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 637–650. [[CrossRef](#)]
52. Seshacharyulu, P.; Rachagani, S.; Muniyan, S.; Siddiqui, J.A.; Cruz, E.; Sharma, S.; Krishnan, R.; Killips, B.J.; Sheinin, Y.; Lele, S.M.; et al. FDPS Cooperates with PTEN Loss to Promote Prostate Cancer Progression through Modulation of Small GTPases/AKT Axis. *Oncogene* **2019**, *38*, 5265–5280. [[CrossRef](#)]
53. Reilly, J.E.; Neighbors, J.D.; Hohl, R.J. Targeting Protein Geranylgeranylation Slows Tumor Development in a Murine Model of Prostate Cancer Metastasis. *Cancer Biol. Ther.* **2017**, *18*, 872–882. [[CrossRef](#)]
54. Peall, K.J.; Smith, D.J.; Kurian, M.A.; Wardle, M.; Waite, A.J.; Hedderly, T.; Lin, J.P.; Smith, M.; Whone, A.; Pall, H.; et al. SGCE Mutations Cause Psychiatric Disorders: Clinical and Genetic Characterization. *Brain* **2013**, *136*, 294–303. [[CrossRef](#)] [[PubMed](#)]
55. Zhao, L.; Qiu, T.; Jiang, D.; Xu, H.; Zou, L.; Yang, Q.; Chen, C.; Jiao, B. SGCE Promotes Breast Cancer Stem Cells by Stabilizing EGFR. *Adv. Sci.* **2020**, *7*, 1903700. [[CrossRef](#)]
56. Bell, J.L.; Hagemann, S.; Holien, J.K.; Liu, T.; Nagy, Z.; Chulte, J.H.; Misiak, D.; Hüttelmaier, S. Identification of Rna-Binding Proteins as Targetable Putative Oncogenes in Neuroblastoma. *Int. J. Mol. Sci.* **2020**, *21*, 5098. [[CrossRef](#)] [[PubMed](#)]
57. Koc, E.C.; Haciosmanoglu, E.; Claudio, P.P.; Wolf, A.; Califano, L.; Friscia, M.; Cortese, A.; Koc, H. Impaired Mitochondrial Protein Synthesis in Head and Neck Squamous Cell Carcinoma. *Mitochondrion* **2015**, *24*, 113–121. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.