UNIVERSITY OF CRETE

MEDICAL SCHOOL

GRADUATE PROGRAM
"MOLECULAR BASIS OF HUMAN DISEASES"

Splicing modulation of the lncRNA PVT-1 with CRISPR Artificial Splicing Factors and its effect on MYC expression

Master Thesis

SOFIA PERDIKARI

HERAKLION, 2023

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

ΤΜΗΜΑ ΙΑΤΡΙΚΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
“ΜΟΡΙΑΚΗ ΒΑΣΗ ΝΟΣΗΜΑΤΩΝ ΤΟΥ ΑΝΘΡΩΠΟΥ”

Τροποποίηση του ματίσματος του lncRNA PVT-1 με χρήση CRISPR Artificial Splicing Factors και η επίδραση του στην έκφραση του MYC

Μεταπτυχιακή Διατριβή

ΣΟΦΙΑ ΠΕΡΔΙΚΑΡΗ

ΗΡΑΚΛΕΙΟ, 2023

This study has been conducted at the RNA and Gene Regulation Lab, at the Institute of Molecular Biology and Biotechnology of the Foundation for Research and Technology Hellas, under the supervision of Dr Evgenia Ntini.

ADVISORY COMMITTEE

1. Dr Evgenia Ntini, Researcher B
(Institute of Molecular Biology and Biotechnology of the Foundation for Research and Technology Hellas)

2. Professor Charalambos Spilianakis
(Department of Biology, University of Crete)

3. Professor Dimitris Kardassis
(School of Medicine, University of Crete)

Thank note

With the successful completion of my Master's thesis as part of the postgraduate program 'Molecular Basis of Human Diseases,' I would like to express my gratitude to all those who played a crucial role in its accomplishment.

First and foremost, I would like to express my heartfelt appreciation to Dr. Evgenia Ntini, my supervisor, for providing me with the opportunity to be a part of this research team. Her mentorship, guidance, and invaluable advice throughout this period were instrumental.

I would also like to convey my thanks to Prof. Charalampos Spilianakis and Prof. Dimitris Kardasis, both esteemed members of my committee, for their unwavering support and expert guidance during this journey.

I must express a special appreciation to Christos Katsioulas, the PhD candidate in our lab, for his invaluable assistance, unwavering support, and the wealth of knowledge and insights he shared, particularly in the context of laboratory life.

Of course, my gratitude extends to all the other members of our lab, whose friendly and joyful atmosphere contributed significantly to this experience.

Last but certainly not least, I would like to express my profound gratitude to my family for their unwavering support throughout these years.

Table of Contents

Table of Contents

Abstract

In recent years, a shift in research focus has been observed, with a transition from mRNAs to non-coding RNAs, highlighting their crucial roles in the regulation of gene expression. Among those RNA species, long non-coding RNAs (lncRNAs), characterized by lengths exceeding 200 nucleotides, have gained prominence. These lncRNAs are implicated in gene regulation through interactions with proteins and other RNAs. These interactions are often associated with specific sequence motifs and features, such as splicing. Despite their relatively low splicing efficiency, lncRNAs exhibit a high degree of alternative splicing, resulting in the generation of various isoforms that can profoundly impact cellular functions and contribute to disease development, including cancer. Notably, Plasmacytoma Variant Translocation 1 (PVT1), located on chromosome 8 in the human genome and positioned closely to the MYC gene, has emerged as a prominent lncRNA. Its expression has been linked to various cancer types, including breast cancer, a leading cause of cancer-related deaths in women. Recent studies have characterized PVT1 as a lncRNA that is retained within chromatin. Machine learning experiments have suggested an association between splicing efficiency and chromatin association. Therefore, the aim of this study is to enhance the splicing efficiency of PVT1, in the worst processed 3' splice sites, as indicated by the bioinformatic analysis of RNA-seq data for MCF-7 cell lines (representing luminal A breast cancer) and MCF-10A/MCF-12A (control) cell lines. To achieve splicing enhancement, the CRISPR Artificial Splicing Factors (CASFx) tool, a modified version of CRISPR-Cas9 containing as splicing activator the fused protein RBFOX1-dCasRx, is employed. A modified CASFx vector has been successfully constructed, allowing for the insertion of guide RNAs for any target in the transcriptome. Initial experiments have revealed promising results, suggesting that the constructed plasmid can influence exon inclusion and potentially impact MYC gene expression levels. However, further investigations are necessary to elucidate the tool's effect on PVT1 chromatin association, and further evaluate its downstream effects. This study holds the potential to uncover critical insights into the role of PVT1 in breast cancer and may open avenues for novel therapeutic approaches.

Περίληψη

Τα τελευταία χρόνια έχει παρατηρηθεί μια μετατόπιση του ερευνητικού ενδιαφέροντος, με τη μετάβαση από τα mRNAs στα μη κωδικά RNAs, αναδεικνύοντας τον κρίσιμο ρόλο τους στη ρύθμιση της γονιδιακής έκφρασης. Μεταξύ αυτών των ειδών RNA, τα μακρά μη κωδικά RNA (lncRNAs), που χαρακτηρίζονται από μήκη που υπερβαίνουν τα 200 νουκλεοτίδια, έχουν αποκτήσει εξέχουσα θέση. Αυτά τα lncRNAs εμπλέκονται στη γονιδιακή ρύθμιση μέσω αλληλεπιδράσεων με πρωτεΐνες και άλλα RNAs. Οι αλληλεπιδράσεις αυτές συχνά συνδέονται με συγκεκριμένα μοτίβα και χαρακτηριστικά της αλληλουχίας, όπως η ωρίμανση. Παρά τη σχετικά χαμηλή αποτελεσματικότητα του ματίσματός τους, τα lncRNAs παρουσιάζουν υψηλό βαθμό εναλλακτικού ματίσματος, με αποτέλεσμα τη δημιουργία διαφόρων ισομορφών που μπορούν να επηρεάσουν τις κυτταρικές λειτουργίες και να συμβάλουν στην ανάπτυξη ασθενειών, συμπεριλαμβανομένου του καρκίνου. Ειδικότερα, το Plasmacytoma Variant Translocation 1 (PVT1), που βρίσκεται στο χρωμόσωμα 8 του ανθρώπινου γονιδιώματος και τοποθετείται κοντά στο γονίδιο MYC, έχει αναδειχθεί ως ένα εξέχον lncRNA στον ερευνητικό τομέα. Η έκφρασή του έχει συνδεθεί με διάφορους τύπους καρκίνου, συμπεριλαμβανομένου του καρκίνου του μαστού, που αποτελεί την κύρια αιτία θανάτου από καρκίνο στις γυναίκες. Πρόσφατες μελέτες έχουν χαρακτηρίσει το PVT1 ως ένα lncRNA που διατηρείται προσδεμένο στη χρωματίνη. Πειράματα μηχανικής μάθησης έχουν υποδείξει μια συσχέτιση μεταξύ του ματίσματος και της πρόσδεσης στην χρωματίνη. Ως εκ τούτου, στόχος της παρούσας μελέτης είναι η ενίσχυση του ματίσματος του PVT1, στις χειρότερα επεξεργάσιμες 3' θέσεις ματίσματος, όπως υποδεικνύεται από τη βιοπληροφορική ανάλυση δεδομένων RNA-seq για κυτταρικές σειρές MCF-7 (που αντιπροσωπεύουν τον καρκίνο του μαστού Luminal A) και MCF-10A/MCF-12A, που αντιπροσωπεύουν τα φυσιολογικά κύτταρα. Για να επιτευχθεί η ενίσχυση του ματίσματος, χρησιμοποιείται το εργαλείο CRISPR Artificial Splicing Factors (CASFx), μια τροποποιημένη έκδοση του CRISPR-Cas9 που περιέχει τη συντηγμένη πρωτεΐνη RBFOX1-dCasRx, η οποία ενεργοποιεί το μάτισμα. Στη διάρκεια της πτυχιακής, κατασκευάστηκε με επιτυχία ένας τροποποιημένος φορέας CASFx, ο οποίος επιτρέπει την εισαγωγή RNA-οδηγών για οποιονδήποτε στόχο στο μεταγράφωμα. Τα αρχικά πειράματα αποκάλυψαν ελπιδοφόρα αποτελέσματα, υποδηλώνοντας ότι το πλασμίδιο που κατασκευάστηκε μπορεί να επηρεάσει την διατήρηση των εξονίων στο μετάγραφο και ενδεχομένως να επηρεάσει τα επίπεδα έκφρασης του γονιδίου MYC. Ωστόσο, απαιτούνται περαιτέρω έρευνες για να διευκρινιστεί η επίδραση του εργαλείου στην πρόσδεση του PVT1 στην χρωματίνη και να αξιολογηθούν περαιτέρω οι μεταγενέστερες επιδράσεις του. Η μελέτη αυτή έχει τη δυνατότητα να αποκαλύψει κρίσιμες γνώσεις σχετικά με το ρόλο του PVT1 στον καρκίνο του μαστού και μπορεί να ανοίξει δρόμους για νέες θεραπευτικές προσεγγίσεις.

Introduction

1. Non-coding RNAs

The human genome plays a crucial role in encoding proteins that are essential for various cellular functions and regulations. However, this protein-coding region constitutes only 1.5-2% of all genes, while the majority of genes produce non-coding RNAs. Traditionally, messenger RNAs (mRNAs) have been the primary focus of research as they serve as templates for protein synthesis, while non-coding RNAs (ncRNAs) were often regarded as by-products of massive transcription with less biological significance. However, since the discovery of ribosomal RNA (rRNA) and transfer RNA (tRNA) in the late 1950s, a diverse array of RNA species has been gradually unveiled, shedding light on a previously unrecognized non-coding world. In the past, the non-coding genome was often perceived as "junk RNA," seemingly remaining in the genome after years of evolution. Nonetheless, with advancements in research and technology over the decades, it has become evident that non-coding RNAs (ncRNAs) indeed play significant roles in cellular functions (1–4).

The groundbreaking human genome sequencing marked a major milestone in today's functional genomics landscape, fueling biomedical research by providing detailed nucleotide sequence information for protein-coding genes. As part of this effort, the US National Human Genome Research Institute (NHGRI) initiated the Encyclopedia of DNA Elements (ENCODE) project to decipher the elements within the human genome. The project revealed that over 80% of the genome consists of DNA elements that, after transcription, do not translate into proteins but instead serve regulatory functions, such as binding sites for transcription factors and regulators. These non-coding RNAs are vital for maintaining genomic integrity, cell differentiation, and development, and their dysregulation has been linked to various human diseases (3).

1.1 Categories of non-coding RNAs

Based on their regulatory role, ncRNAs are separated into two distinct categories: the housekeeping ncRNAs and the regulatory ncRNAs (Figure 1)Housekeeping ncRNAs include rRNAs, tRNAs, small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs) and telomerase RNAs. This category of RNAs are abundantly and ubiquitously expressed in cells, primarily regulating generic cellular functions (4). Regulatory ncRNAs are usually considered as key regulatory RNA molecules, function as regulators of gene expression at epigenetic, transcriptional, and post-transcriptional levels. This category is divided into two subcategories, small non-coding RNAs, with transcripts sizing up to 200nt, and long non-coding RNAs (lncRNAs), whose size is greater than 200nt(2,4).

The small non-coding RNAs include microRNAs (miRNAs), small interfering RNA (siRNA), piwi-interacting RNA (piRNA), transcription initiating RNA (tiRNA), transcription start-site-associated RNA (TSSaRNA), promoter-associated small RNA (PASRs) and promoter upstream RNAs (PROMPTs). MiRNAs are conserved single-stranded RNA molecules, usually around 18-22 nucleotide long, and they are the most abundant and localized in both the cytoplasm and the nucleus, participating in vital functions of cellular processes and homeostasis. This type of small ncRNAs, is usually known for gene silencing at a post-

transcriptional level, usually by interacting with mRNAs (miRNA-mRNA interactions) by binding to complementary sequences in the 3' UTRS. These miRNAs act as oncogenes by suppressing the function of tumor suppressor genes, or by enhancing the expression of mRNA associated with tumorigenesis. Moreover, miRNAs have been found to interact with more than one mRNAs or vice versa, underlining that they form multiple networks, whose dysregulation can be associated with many diseases (3–5).

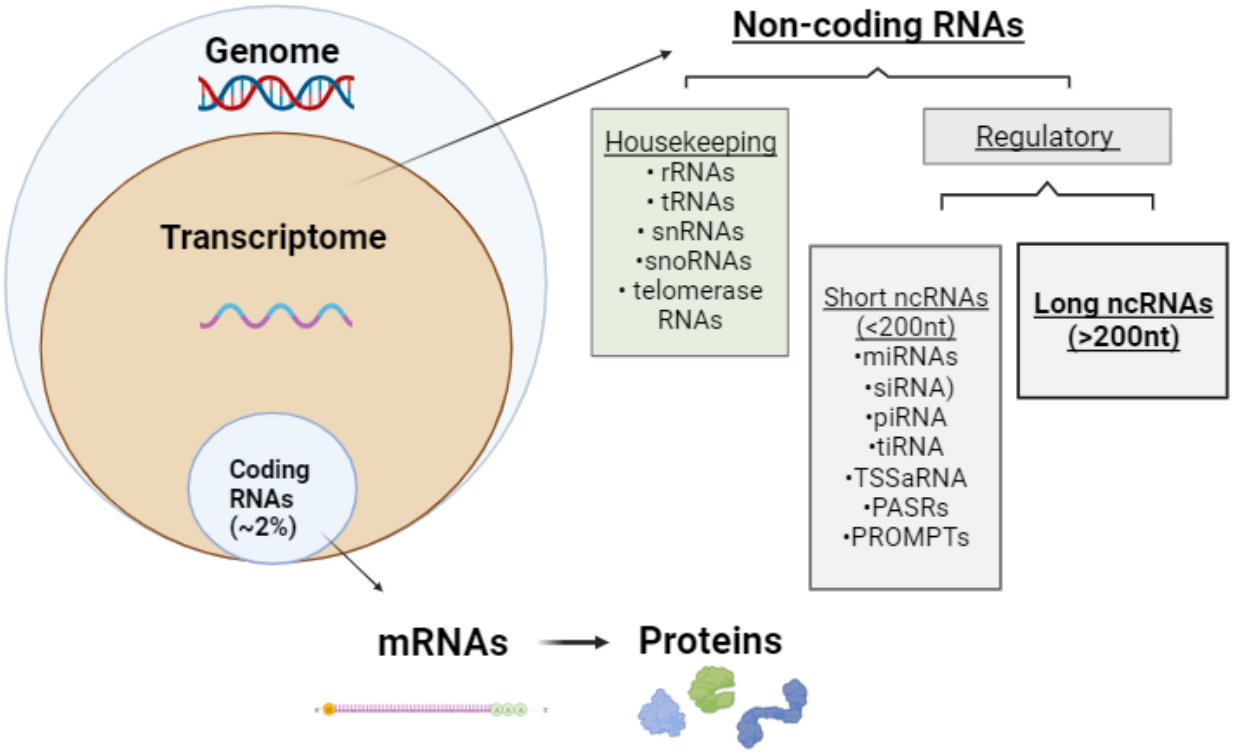

Figure 1. Schematic representation of the categories of RNA, including the subtypes in the non-coding RNAs.

2. The long non-coding RNAs

Long non-coding RNAs (lncRNAs), along with other non-coding RNAs (ncRNAs), lack the ability to encode proteins. It is estimated that the human genome encodes more than 13,000 lncRNAs, although this number may vary across different databases (6). They are primarily transcribed by RNA polymerase II (RNA polII), although transcripts can also be generated by RNA polymerase III (RNA polIII) (7,8). Those transcribed by RNA polII possess modifications, such as 5' end m^7G caps and 3' end poly(A) tails, considered to be processed similarly to mRNAs(6,9). Recent research (10) using mNET-seq methodology, has indicated that the phosphorylation of RNA polII in the carboxy-terminal domain (CTD) is associated with different transcription stages. For example, long non-coding RNAs that are weakly co-transcriptionally spliced and their transcription termination is independent of the polyadenylation signals, are transcribed by phosphorylation-dysregulated RNA polII. Those lncRNAs have been found to accumulate on chromatin, followed by rapid degradation by RNA exosome. Moreover, the chromatin-

tethered lncRNAs are correlated with increased levels of U1 small nuclear RNA binding sites, recruiting the U1 snRNP to transcriptionally engaged RNA polII(9).

Long non-coding RNAs' functionality is closely linked to their sequence composition, as well as their primary, secondary, and tertiary structures. Typically, the primary sequence lays the foundation for subsequent structural configurations. However, in certain cases, lncRNAs may contain specific sequences that contribute to their functional roles. Examples of this phenomenon are observed in lncRNAs like PVT1, MALAT1, and BORG, which harbor nuclear localization sequences within their primary sequences, facilitating their distinctive functions. Moreover, some lncRNAs possess sequences that are complementary to recognition motifs in U1 snRNPs, which are crucial in pre-mRNA splicing (11).

2.1 Categories of lncRNA based on their location characteristics and regulatory effects

2.1.1 Categories of lncRNAs based on their genomic position

LncRNAs can be categorized into five distinct subtypes based on their genomic positions and characteristics (Figure 2). Firstly, there are the sense-overlapping lncRNAs are transcribed from the sense strand and may encompass exons from protein-coding genes, although they may not harbor a functional open-reading frame. Furthermore, the antisense lncRNAs are transcribed from the antisense strand, overlapping with exonic or intronic regions, or even spanning the protein-coding sequence within an intron. Then, there are long intronic ncRNAs that are derived from the introns of protein-coding genes, arising from alternative splicing of pre-mRNA, and can originate from both the sense and antisense strands. In addition, the long intergenic ncRNAs (lincRNAs), are transcribed from intergenic regions on both strands, situated up to 10kb away from the nearest protein-coding gene. Lastly, bidirectional lncRNAs are transcribed from the promoter of a protein-coding gene, with their transcription initiated in a divergent manner (3,4,10,12). This type of lncRNAs, usually give rise to enhancer RNAs (eRNAs) and have been implicated in enhancer-promoter communication and gene regulation (?). The relative position between a lncRNA and its neighboring genes is a key determinant of their regulatory relationship (9).

2.1.2 Categories of lncRNAs based on their region of function

LncRNAs possess the capability to engage with proteins and other smaller RNAs, such as microRNAs (miRNAs), thereby positioning them as significant regulators of gene expression. Depending on their regulatory impact, they can be classified as either cis-lncRNAs, which oversee genes in close proximity, involving a single gene or multiple genes within a specific region. Alternatively, they can influence the expression of more than one gene, or they can take the form of trans-lncRNAs, which exert their regulatory influence over distant genes to a different chromosomal region (3,4). This category also encompasses lncRNAs that function as miRNA sponges, modulating miRNA expression, either by suppressing or promoting it, thus making them key players in of gene expression regulation (Figure 3) (5).

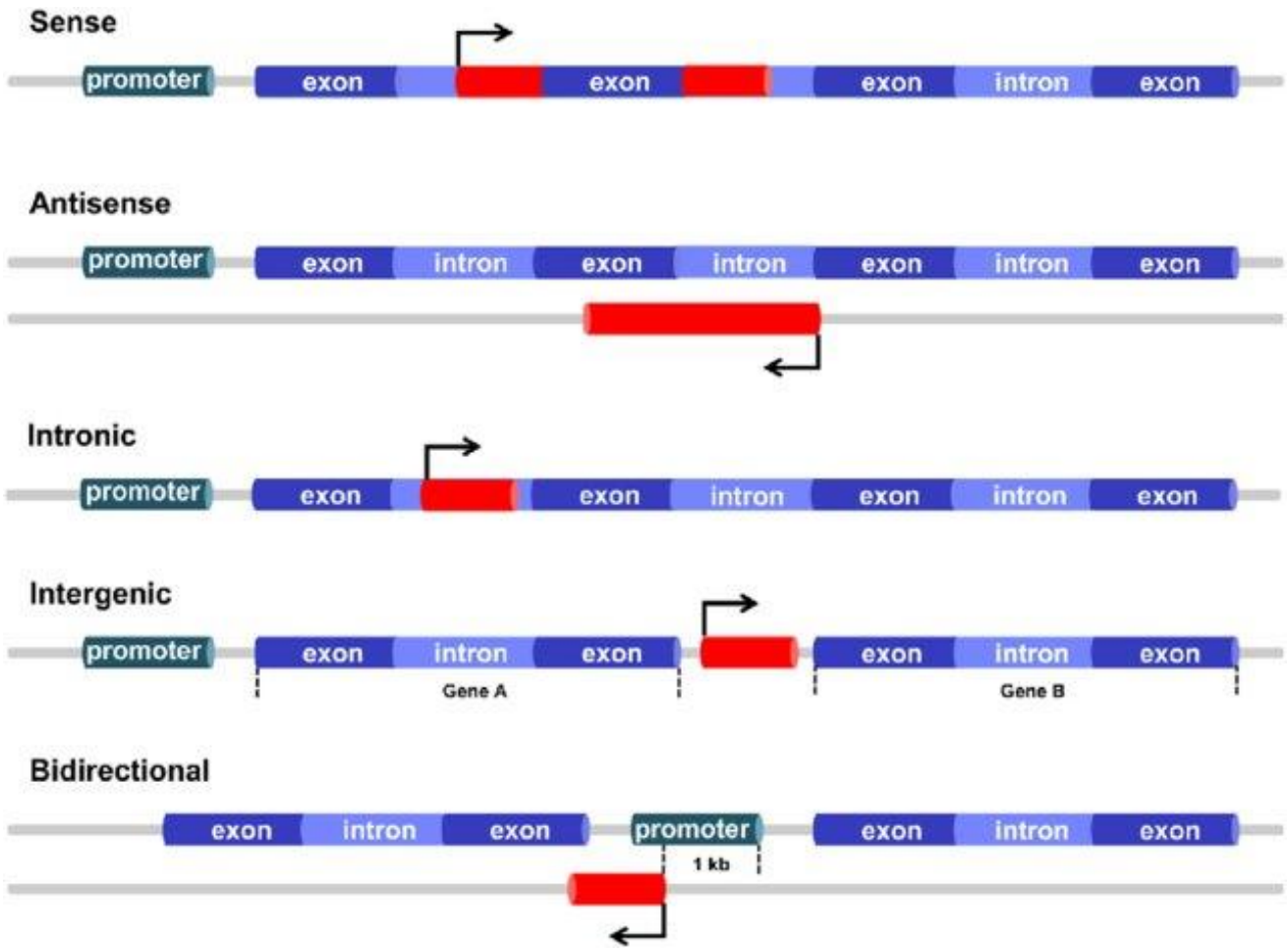

Figure 2.. Schematic representation of the different categories of lncRNA based on their genomic position: sense, antisense, intronic, intergenic, bidirectional; Figure taken from ref (12).

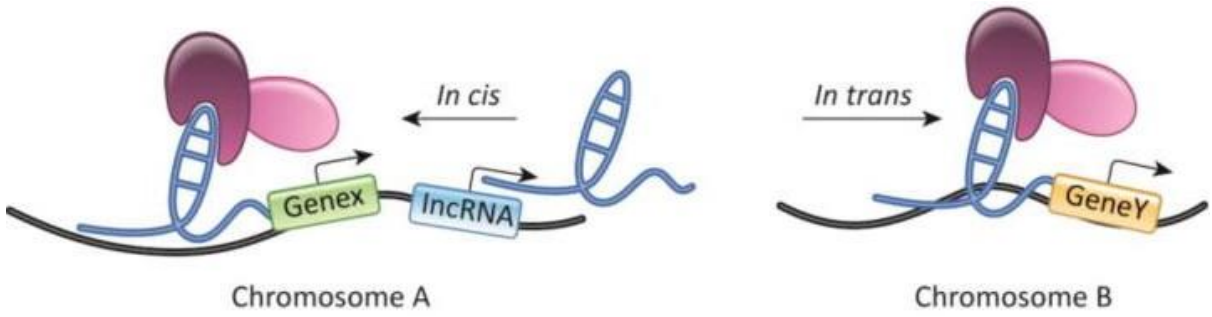

Figure 3. Schematic representation of the different functions of lncRNA: cis-acting, trans-acting; Figure taken from ref (13).

2.1.3 Categories of lncRNAs based on their function

An alternative classification, based on the functional roles of lncRNAs, divides them in categories such as signaling, decoy, guide and scaffold and enhancer (**Figure *4***). This classification predominantly hinges on the structural aspects of lncRNAs, taking into account their primary, secondary, and tertiary configurations. Signal lncRNAs oversee signaling cascades and particularly modulate transcription in response to stimuli, all without the involvement of a protein factor. These lncRNAs generally function as indicators of transcriptional activity. Conversely, decoy lncRNAs possess binding sites that hinder the interaction of other regulatory components, such as transcription factors, catalytic proteins, chromatin-modifying complexes, and even miRNAs. Scaffold lncRNAs operate as templates for assembling complexes, incorporating proteins, miRNAs, other lncRNAs, or even RNP complexes. This assembly process subsequently results in transcriptional activation or repression. Guide lncRNAs are responsible for directing RNP to specific gene targets, emphasizing their significance in ensuring the accurate localization of these complexes. Lastly, the enhancer are a class of non-coding RNAs transcribed from enhancer regions of the genome. It is thought that these lncRNAs are not released from ERs (Enhancer Regions) but tether proteins to ERs affecting gene expression. Also, the "chromatin interaction" of DNA is influenced by the enhancer regions as a result of enhancer RNAs (eRNAs). The disruption of any of aforementioned functionalities has been extensively investigated in recent years, revealing a strong association with various pathological conditions, including cancer (13–15).

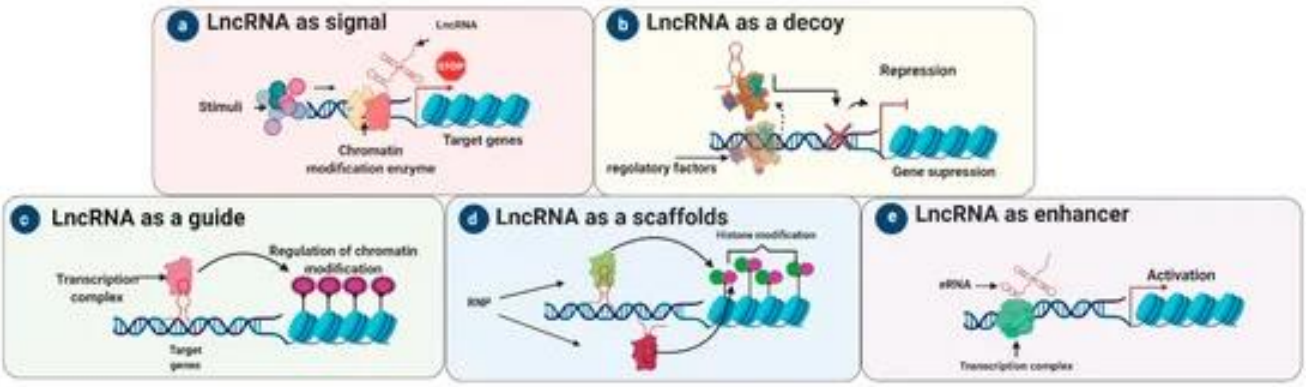

Figure 4. Schematic representation of the different functions of lncRNA: signal, decoy, guide, scaffold, enhancer; Figure taken from ref *(15)*.

2.2. Localization of lncRNAs and gene regulation by lncRNAs.

Initially, lncRNAs were believed to be unstable, but subsequent studies have shown that the majority of them are stabilized by modifications, particularly polyadenylation (16,17). Additionally, lncRNAs can form various structural elements, such as bulges, junctions, hairpin loops, stem loops, internal loops, helices, subdomains, and pseudoknots (**Figure *5***) (11,18).

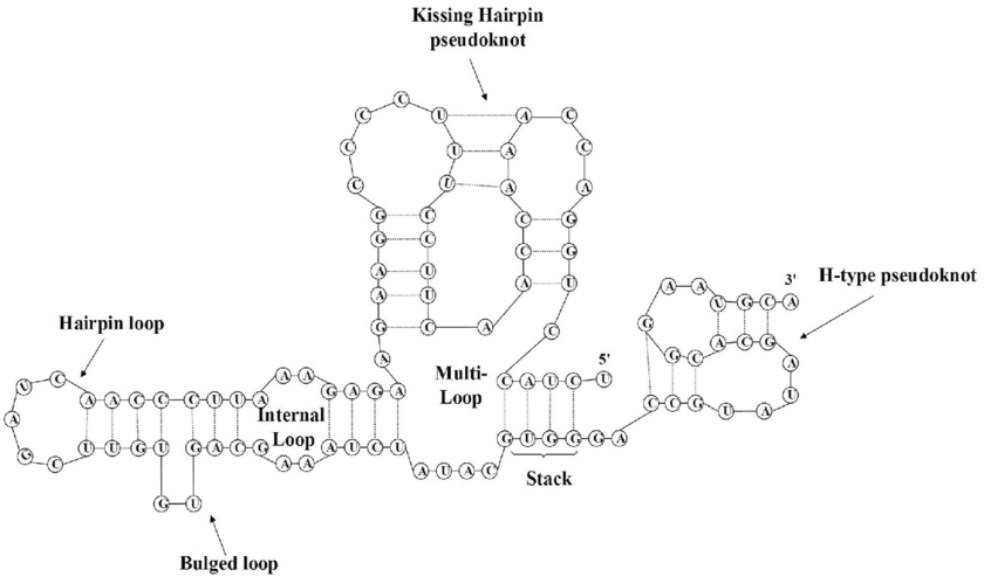

Figure 5: Schematic representation of the structural conformations of RNAs. Figure taken from ref (18)

These modifications not only mediate their interactions with DNA, proteins, and other RNAs (11) but also regulate their localization, facilitating their transport to the nucleus (17). Furthermore, their primary sequence can influence their retention in the nucleus, particularly due to the presence of repeat elements. For example, C-rich sequences generated by Alu repeats can bind to hnRNPK, a protein that promotes the nuclear retention of these lncRNAs. Moreover, lncRNAs often contain repetitive motifs that recruit nuclear factors, further enhancing their nuclear localization. Recent studies have identified a C-rich sequence from Alu repeats that enhances lncRNA nuclear retention through interaction with hnRNPK and the nuclear matrix protein (9). Consequently, the majority of lncRNAs are predominantly found in the nucleus (19,20), although there are higher levels of cytoplasmic lncRNAs than initially assumed (21).

RNA FISH experiments conducted by Cabilli et al. have revealed diverse subcellular localization patterns for lncRNAs, including large nuclear foci, large nuclear foci with scattered single molecules throughout the nucleus, predominantly nuclear localization without foci, a combination of cytoplasmic and nuclear localization, and predominantly cytoplasmic localization (23). Additionally, lncRNAs exhibit distinct organization within various cellular compartments and organelles. Within the nucleus, they are primarily present in nucleoli, responsible for ribosomal synthesis, as well as in chromatin speckles, where they participate in transcription and pre-mRNA maturation processes. Furthermore, lncRNAs are localized in paraspeckles, heterochromatin domains involved in mRNA nuclear retention (21).

However, it's worth noting that nuclear lncRNAs are generally less stable than those found in the cytoplasm, which is linked to their functional role in regulating gene expression (16). This regulation encompasses various mechanisms, including chromatin organization, interactions with the transcription machinery, and post-transcriptional modifications.

2.2.1. Chromatin and nuclear organization

The lncRNAs that participate in chromatin remodelling are characterized as chromatin-enriched RNAs (cheRNAs) (19). Examples include Xist, which represses most genes on the X chromosome and mediates dosage compensation in mammals, and H19, associated with heterochromatin formation (6,22). Some nuclear lncRNAs interact with SWI/SNF or PCR subunits, influencing chromatin architecture. Alternatively, specific lncRNAs have repeating RNA domains (RRDs) that interact with hnRNPU, mediating trans-chromatin interactions (19). Some chromatin-associated lncRNAs interact with proteins like CCCTC-binding factor (CTCF), either facilitating or inhibiting their binding to DNA regions. Some lncRNAs recruit chromatin-modifying complexes to promote gene activation (9). Moreover, negatively charged RNA can interact with positively charged histone tails, leading to chromatin decompaction, highlighting the role of these interactions in gene expression regulation (9).

2.2.2. Transcriptional regulation

One well-documented function of lncRNAs is transcriptional regulation. They interact with transcription factors or chromatin-modifying complexes, either locally (cis-regulation) or distally targeting genes on other chromosomes (trans-regulation) (6,19). Another area of interest is the direct interaction of lncRNAs with chromatin, forming hybrid structures called triple helices or R-loops. These structures influence chromatin accessibility, impacting downstream gene expression. Various methodologies, like TrIP-seq, have been developed to study these conformations in vivo (23). Recent research has intensified the study of chromatin-associated lncRNAs' effects on nearby genes (within 300kb), revealing their role as enhancers (24). Additionally, lncRNA-protein complexes (lncRNPS) are associated with processes such as mRNA splicing and signaling pathways. These interactions occur due to sequence motifs or structures that enable RNA binding protein (RBP) binding, such as UGCAU or GCAUG motifs, which recruit splicing factors, including RBFOX2 (25). There are many different motifs that recruit different RBPs (9).

2.2.3 Post-transcriptional regulation

LncRNAs also regulate gene expression through interactions with miRNAs, acting as competing endogenous RNAs (ceRNAs) (3,5). CeRNAs suppress the function of other non-coding RNAs, such as miRNAs, by reducing the number of miRNAs available to interact with their mRNA targets, affecting gene regulation (26). CeRNAs often interact with multiple miRNAs, carrying different miRNA recognition elements (MREs). This complex network of mRNAs, ceRNAs, miRNAs, and their interactions forms a ceRNA network (4). This lncRNA-miRNA-mRNA network plays a crucial role in cancer progression by activating key pathways and mechanisms associated with the disease (27).

Another mechanism involves lncRNAs influencing mRNA stability, exemplified by Staufen homologue 1 (STAU1), a double-stranded RNA binding protein. In this case, lncRNAs containing Alu

retroelements in humans promote STAU1-mediated mRNA decay of mRNAs that partially or fully complement these repeats by recruiting STAU 1 (28).

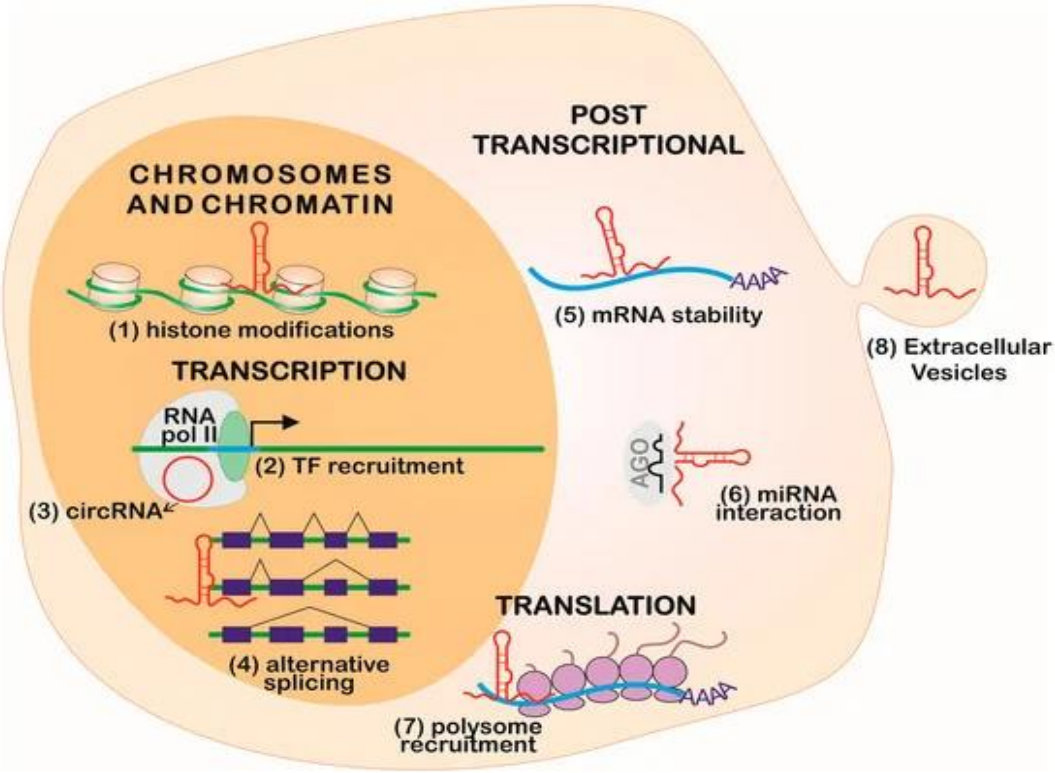

Figure 6. Illustration of the localization of lncRNAs and how it can affect the gene regulation in different levels, including chromatin modification, in transcriptional and post-transcriptional level. Figure taken from ref (26).

2.3 Involvement of Long Non-Coding RNAs in Diseases and their potential as therapeutic targets

Due to their extensive interactions, long non-coding RNAs (lncRNAs) play a crucial role in regulating genes and, consequently, in various cellular processes associated with diseases. For instance, consider the lncRNA H19, which promotes angiogenesis and is linked to several types of cancer, including bladder, colorectal, and hepatocellular carcinoma, due to its interaction with p53. PVT1 and NEAT1 are two additional lncRNAs associated with cancer development, particularly in breast cancer, where they contribute to chemotherapy resistance. Other lncRNAs like MALAT1, HOTAIR, and BCAR4 are also subjects of study for their roles in cancer development (29).

Furthermore, lncRNAs are implicated in the development of hereditary diseases. For example, NEAT1 has been associated with amyotrophic lateral sclerosis (ALS) (30). In the realm of neurological diseases, lncRNA BC-200 has been found to be dysregulated in Alzheimer's disease (31), and the presence of BACE1-AS lncRNA has been linked to this condition (32). Another disease that has been associated with dysregulated expression of lncRNA is diabetes, such as TUG1 that affects the function of pancreatic β-cells, or MEG3 that is associated with insulin resistance. Diabetes is another disease with altered lncRNA expression, where TUG1 affects pancreatic β-cell function, and MEG3 is associated with insulin

resistance. Cardiovascular diseases have also been connected to changes in lncRNA expression, such as DACH1, which is upregulated in heart failure patients, and MALAT1 (29).

As previously mentioned, a substantial number of lncRNAs are deeply involved in fundamental aspects of hereditary diseases and cancer, impacting processes like uncontrolled cell growth and resistance to cell death. They can act directly as oncogenes or tumor suppressors or indirectly by interacting with established cancer-related genes like *MYC* or p53 (33). Moreover, long noncoding RNAs exhibit specific expression patterns in particular cells, tissues, or tumors, making them highly promising for potential therapeutic applications. Over the years, various strategies have been developed (34).

One of the earliest RNA-based therapeutic approaches involves RNA interference (RNAi), which employs double-stranded RNA to silence specific target genes (35). This method has evolved with the creation of synthetic siRNAs, with numerous pharmaceutical companies utilizing this technique for advancing treatment development. Chemical enhancements like 2'-O-methyl (2'-O-Me) sugar modifications and phosphothioate linkages have been incorporated to enhance the stability of therapeutic molecules (36). For example, siRNA has been successfully used in preclinical studies for genes like CASC9 in esophageal squamous cell carcinoma and DANCR in a triple-negative breast cancer mouse model (29).

Another well-known strategy is the use of antisense oligonucleotides (ASOs). ASOs are short synthetic nucleotide sequences (15-20nt) that can be designed to bind specifically to complementary RNA sequences, leading to downstream RNA degradation. Chemical modifications like phosphothioate linkages or 2'-O-Me have addressed challenges related to low cell permeability and poor stability, significantly improving their pharmacological properties (34,37). ASOs provide a versatile toolset for precise interventions, enabling the selective degradation of oncogenic lncRNAs, modulation of their splicing patterns, or disruption of their critical interactions. By directly targeting the molecular dysregulations associated with cancer, ASOs hold promise for developing highly specific and effective therapeutic strategies. While challenges such as efficient delivery and safety persist, the strategic application of ASOs presents a promising avenue for advancing precision medicine in cancer therapy (38). Notably, the U.S. Food and Drug Administration (FDA) has approved ASO drugs for treating conditions like spinal muscular atrophy, Duchenne muscular dystrophy, and familial amyloid polyneuropathy, with several other ASO-based drugs in clinical trials (29).

CRISPR-based therapies have also emerged, particularly focusing on gene knockouts in diseases like cancer, metabolic disorders, and neurodegenerative diseases (39). Clinical trials employing CRISPR–Cas for ex vivo genome editing are actively underway. For example, attempts are being made to disrupt the CCR5 gene, the predominant co-receptor for entry of the HIV virus into human T cells, in CD34+ hematopoietic stem cells (HSCs) of HIV-infected individuals with hematological malignancies (40). Moreover, there have been pioneering CRISPR human trials for treating patients with metastatic non-small cell lung cancer who did not respond to conventional therapies. This involves CRISPR-edited T cells with a knocked-out PD-1 gene (41).

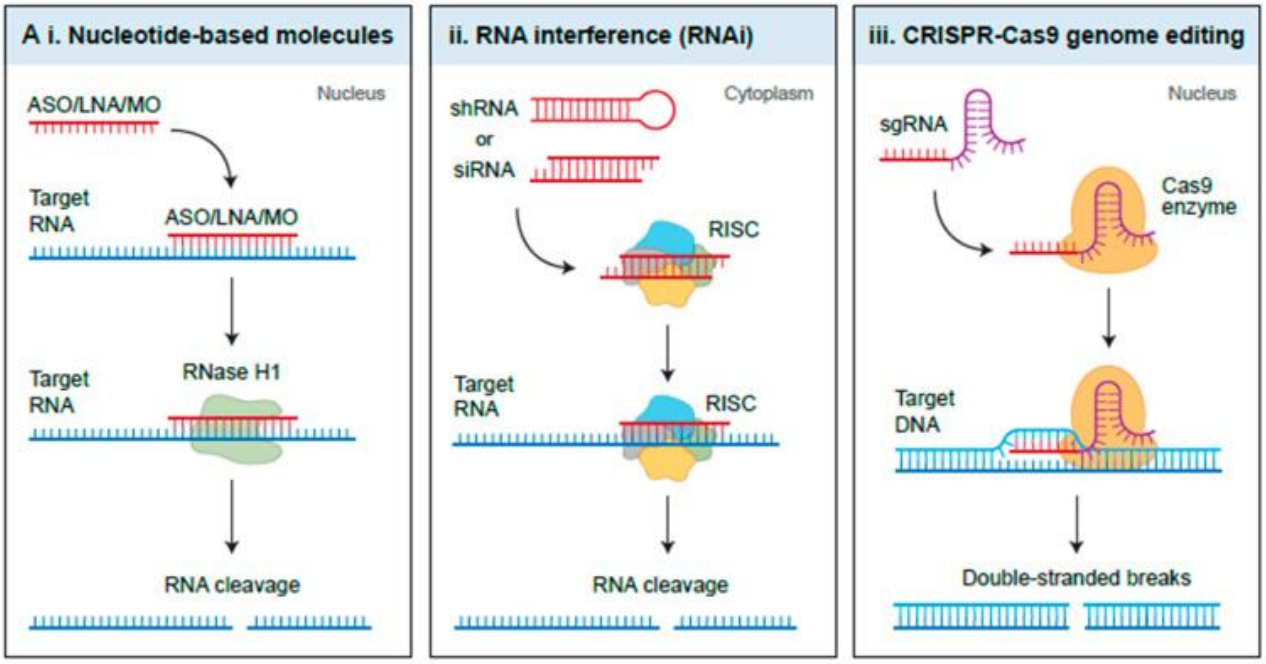

Figure 7. Schematic representation of the RNA-based therapeutic approaches: A. Anti-sense oligonucleotides (ASOs), B. RNA interference (RNAi), C. CRISPR-Cas9 genome editing. Figure taken from ref (42).

In summary, despite non-coding RNAs not being directly associated with protein production, their presence significantly impacts cellular functions and is linked to various diseases. Consequently, targeting these ncRNAs holds potential as a therapeutic approach for these diseases. So far, the FDA has approved 11 RNA-based therapeutics for diseases related to the liver, muscle, and nervous system, including the use of antisense oligonucleotides (ASOs) and small interfering RNAs (siRNAs). However, ongoing clinical development includes various other RNA-based therapeutics like short hairpin RNAs (shRNAs), ASO anti-microRNAs (antimiRs), miRNA mimics, miRNA sponges, therapeutic circular RNAs (circRNAs), and CRISPR–Cas9-based gene editing (40).

3. RNA Splicing

The concept of non-coding sequences, called introns, that interrupt the protein-coding sequence of genes, is nowadays an established and proven consensus in the field of Molecular Biology. To transform messenger RNA precursors (pre-mRNA) into mature mRNA, these introns must be excised. This pivotal process, named pre-mRNA splicing, is heavily reliant on the activity of a complex known as spliceosome (42). In this procedure, approximately 10% of pre-mRNA sequence is retained as mature mRNA, while the remaining 90% is eliminated as introns. It's widely acknowledged that the majority of splicing events occur co-transcriptionally, with intricate interplay between the splicing and transcriptional machinery (43).

3.1 The mechanics of RNA splicing

The orchestration of this process revolves around the sequence of introns, with particular emphasis on three conserved segments: the 5' splice site (5'SS) marked by a highly conserved GU sequence, the branch point (BP) adenosine, and the polypyrimidine tract (Y(n)), coupled with the 3' splice site (3'SS) housing a conserved AG dinucleotide. The splicing reaction unfolds through a sequence of steps, beginning with branching and finishing with exon ligation. In branching, the 2' hydroxyl group of the BP adenosine forms a loop with the phosphodiester group at the 5'SS, leading to the cleavage of the 5' exon. Subsequently, in exon ligation, the 3' hydroxyl group of the 5' exon combines with the phosphodiester group at the 3'SS, resulting in the connection between the 5' and 3' exons, with the concurrent release of the intron (42–46).

3.1.1 The splicing machinery

The splicing process is conducted by the spliceosome, a complex of five small nuclear RNAs (U1, U2, U4, U5, and U6). U1 engages in base-pairing interactions with the 5'SS, while U2 interacts with the branch point. The remaining three small nuclear ribonucleoprotein particles (snRNPs) converge to construct an integrated complex, from which U4 subsequently disengages. This results in U6 replacing U1, facilitating interactions between U2 and U6. This interaction, in turn, brings the 5'SS and branch point into close proximity. Ultimately, U5 brings the two exons together, enabling splicing and exon ligation (**Figure *8***) (43,45,47). This splicing machine catalyzes 99.5% of splicing reactions. The remaining splice sites are recognized by the minor (or U12-dependent) spliceosome, comprising the U5 snRNP as well as unique functional analogs of the major spliceosome's U11, U12, U4atac, and U6atac snRNPs (45).

3.2 Alternative splicing

Splice sites are also distinguished based on their consistent recognition as splice sites ('constitutive') or their occasional recognition ('alternative'). The phenomenon of alternative splicing (AS) is responsible for the creation of several isoforms from a single gene. The transcriptional diversity affects the transcriptome and proteome, introducing a pivotal layer to gene expression. The production of distinct isoforms profoundly modulates cellular functions, influencing interactions of these transcripts with proteins, DNA and other RNAs, as well as their subcellular localization. In addition, in the case of protein coding RNAs, alternative transcripts can result in a variety of differential functions, such as distinct enzymatic activity of different protein isoforms. In some instances, the final product of this process might be a truncated protein susceptible to degradation. Notably, the tight association between alternative splicing and the emergence of numerous diseases has been identified, and studies focused on the pathological role of this process are emerging in the field in recent years (43,47).

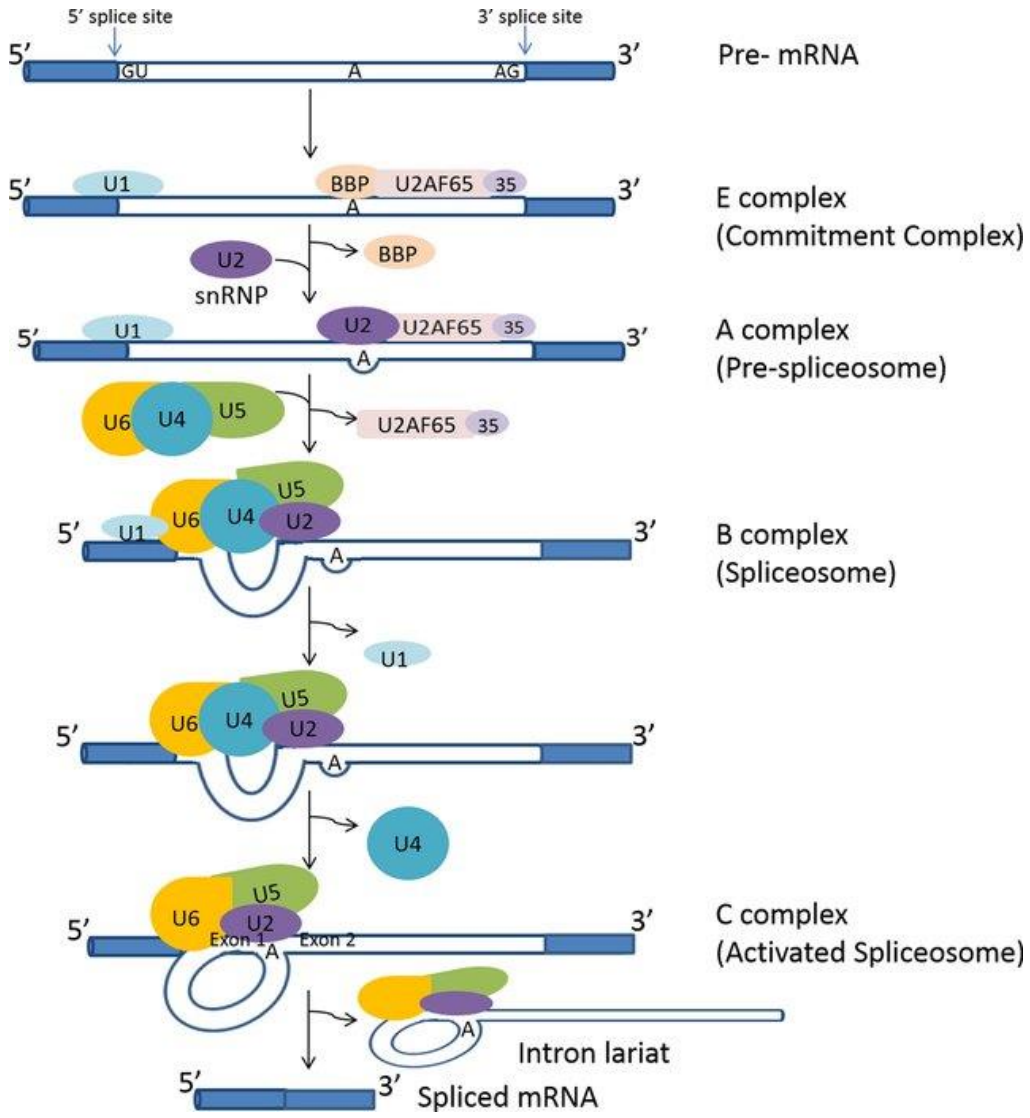

Figure 8. Illustration demonstrating the process of splicing and the role played by components of the splicing machinery. Figure taken by ref (46).

Fundamental alternative splicing patterns include exon skipping in which entire exons can be included or excluded from the mature mRNA, alternative 5' and 3' splice sites, in which different splice sites within an exon can be chosen, mutually exclusive exons that two or more exons are mutually exclusive, meaning only one is included in the mature mRNA, intron retention, in which case introns are not removed, resulting in their presence in the mature mRNA, and alternative splicing coupled with alternative first or last exons (**Figure *9***) (48). Moreover, complex alternative splicing patterns merge multiple of these elemental patterns (44). Alternative splicing events can impact an entire cassette exon or only a portion of it by utilizing alternative splicing sites within a variant exon. This phenomenon is observed in genes that possess multiple transcription start sites or employ multiple polyadenylation sites (49).

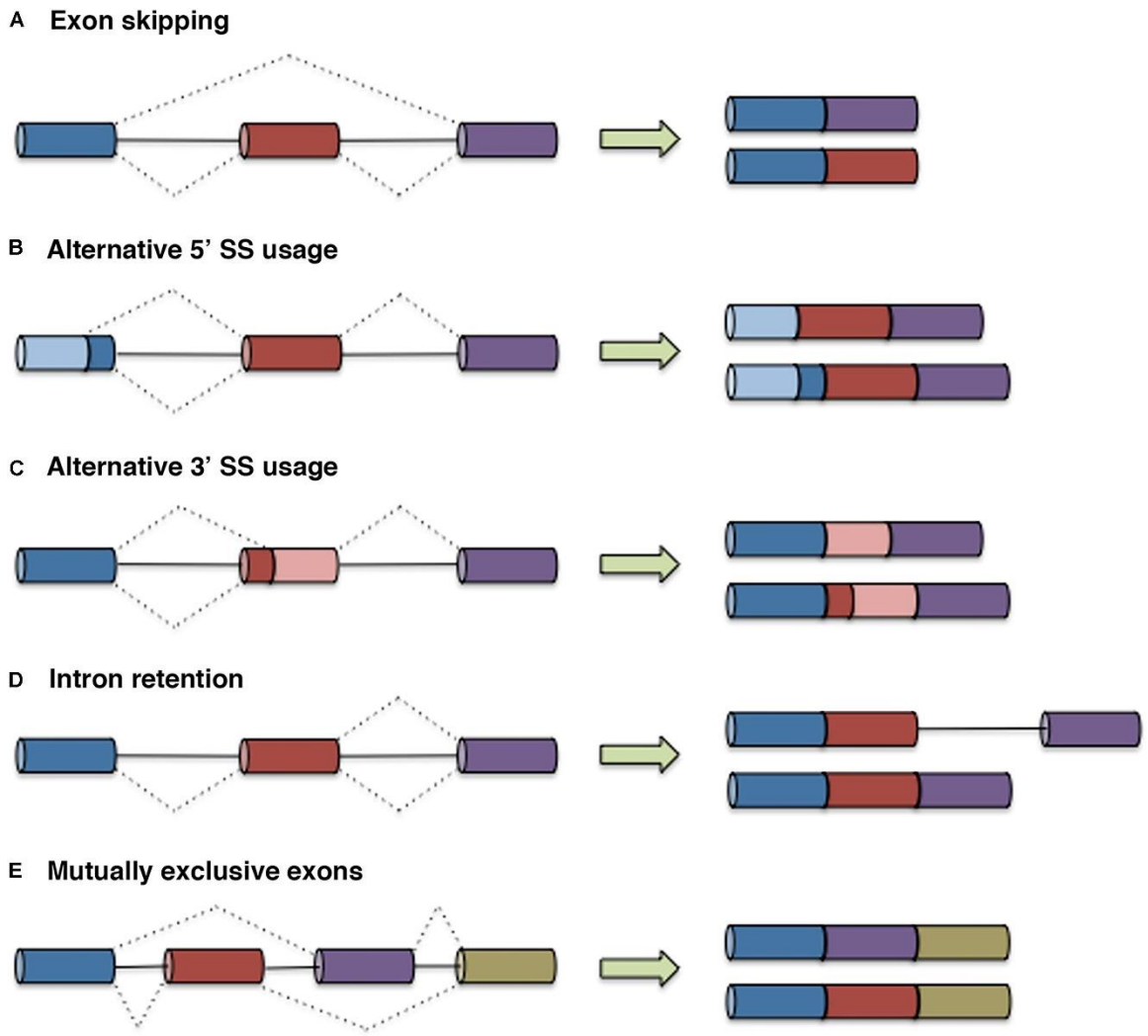

Figure 9. Schematic illustration depicting various forms of alternative splicing and their resulting products. Figure taken by ref (48).

Alternative splicing (AS) primarily hinges on the coordinated regulation of cis-elements, such as exon splice enhancers (ESEs), exon splice silencers (ESSs), intron splice enhancers (ISEs), and intron splice silencers (ISSs), as well as trans-acting factors like serine- and arginine-rich splicing factors (SR proteins) and heterogeneous nuclear ribonucleoproteins (hnRNPs) (50,51). Both SR factors and hnRNP proteins share a common structural framework characterized by distinct domains. These domains serve specific functions, including their ability to bind to RNA, exemplified by the RNA Recognition motif (RRM), and their capacity to engage in interactions with other proteins (50). Thereby, these sequences influencing the assembly of the spliceosome and thereby either activating or repressing splicing (45). SR proteins are commonly believed to facilitate alternative splicing by attaching to ESE sequences situated within variable exons and engaging with spliceosome components. This interaction promotes the inclusion of variable exons. In contrast, hnRNPs typically bind to ESS or ISS elements, leading to the inhibition of variable exon inclusion. They do so by disrupting the connection between the spliceosome and weak

splice sites. Nevertheless, mounting evidence has indicated that SR proteins and hnRNPs can also act in a complementary manner, with SR proteins repressing and hnRNPs activating variable exons under certain circumstances(52).

Furthermore, post-transcriptional modifications, such as N6-Methyladenosine (m6A), exert an influence on alternative splicing. Mutations in splicing-related sequences can lead to abnormal alternative splicing, thereby contributing to the development of diseases, including cancer. Research has also indicated that long non-coding RNAs can emerge as products of alternative splicing and wield biological functions, either promoting or suppressing cancer (49,51).

3.2.1. Alternative splicing in cancer

It has become increasingly evident that the dysregulated expression of splicing variants or the inadequate production of the correct isoforms constitutes a fundamental aspect of cancer development, and this AS can be potentially used as a potential therapeutic biomarker (53). Typically, the splicing patterns of numerous genes undergo changes as cells advance along the oncogenic pathway. These alterations are associated with the development of cancer traits, such as rapid proliferation, the ability to generate new blood vessels, invasion into surrounding tissues, and survival under adverse conditions (50). An example of this is the *TP53* gene, which presents 12 alternative isoforms. Their expression patterns change between various cancer types, and this affects how they respond to therapy. Similarly, to *TP53*, genes such as *PTEN* and *BCL6* have synonymous mutations, which often disrupt exonic motifs involved in splicing regulation. This leads to alterations in how oncogenes are spliced, specifically in certain types of cancer (54). The implications of splicing in cancer development have also been observed in the interaction between splicing factors and RNA polymerase II, which depends on the presence of promoters and enhancers (53).

3.3 CRISPR technologies in gene expression and splicing regulation

The CRISPR-Cas9 system stands as a widely employed genome editing tool, using custom-designed guide RNA sequences to pinpoint specific loci within the genome (55). This system was derived from a naturally occurring genome editing mechanism employed by bacteria as part of their immune defense strategy. In a nutshell, the process initiates with the transcription of the CRISPR array into a single RNA molecule. Subsequently, this molecule is cleaved at the repeat sequences, yielding shorter CRISPR RNAs (crRNAs), each carrying a single spacer sequence. These crRNAs then join forces with a small trans-activating CRISPR RNA (tracrRNA). Together, they create a complex that can be recognized and bound by Cas9, leading to the formation of a ribonucleoprotein (RNP) complex. This RNP complex functions by associating with a phage genome and embarking on a quest to locate sequences that match the spacer encoded within the crRNA. Upon identifying a matching sequence, Cas9 acts as a nuclease, cleaving the DNA and generating a double-strand break (DSB). This DSB effectively thwarts the phage's ability to complete its life cycle (39,55).

Over the years, modifications in this system have led to its use in gene activation (CRISPRa) or inhibition (CRISPRi) (56), but several other functions can be achieved as well, by substituting the Cas9 with other types of Cas proteins with expanded capabilities, as in the case of RCas9, which is an RNA-targeting protein. Among the RNA-targeting Cas9, the most recently identified protein is CasRx, which belongs to the type IV-D CRISPR-Cas ribonucleases and originated from Ruminococcus flavefaciens XPD3002. Studies had shown that by deactivating the catalytic domain of the protein (dCasRx), the system could be used to affect splicing, by either blocking it or enhancing it (57).

The employment of these broad abilities of Cas proteins, has led to the development of new methodologies, such as the CRISPR Artificial Splicing Factors (CASFx) system, which can be utilized in order to promote the splicing in specific splice sites of transcripts of interest. The CASFx system overcomes many barriers that exist with other methods that are related to prominent splicing alterations, such as the use of ASOs, due to its higher specificity and its modulation potential, since the targeted transcript can be changed by modifying the guide RNAs. Moreover, comparing dCasRx with other Rcas9 proteins, CASFx outweighs the presence of the PAMer sequence that is not required compared to other RNA-targeting Cas proteins.

4. The lncRNA plasmacytoma variant translocation 1 (PVT1)

4.1 The Genomic Landscape and Transcripts of *PVT1*

The initial identification of the long non-coding RNA plasmacytoma variant translocation 1 (PVT1) dates back to 1984 when it was recognized as a regulator of the *MYC* gene in murine plasmacytoma variant translocations (58,59). Subsequent studies by Huppi K et al revealed PVT1 involvement in chromosomal translocations and amplifications in murine lymphocytic-B neoplasms (60). The human *PVT1* gene is situated on chromosome 8 at location 8q24.21, spanning a substantial region of over 30 kilobases (kb) from nucleotide position 128806779 to 129113499, residing 57 kb downstream of the *MYC* gene. The *PVT1* locus produces diverse transcripts, but so far, no protein product associated with *PVT1* has been identified (58,59).

4.2 Implications of *PVT1* Dysregulation in Cancer

The expression levels of *PVT1* are low in normal conditions (59). However, abnormal expression of *PVT1* has been observed in various types of human malignancies, encompassing lung cancer (61), gastric cancer (GC) (62), colorectal cancer (CRC) (63), hepatocellular carcinoma (HCC) (64), lymphoma (65)melanoma (66), pancreatic cancer (PC) (67), breast cancer (BC) (68), cervical cancer (CC) (69), malignant pleural mesothelioma (70), multiple myeloma (MM) (71), renal cell carcinoma (RCC) (72), bladder cancer (73), prostate cancer (PCa) (74) glioma (75), osteosarcoma (OS) (76) and thyroid cancer (TC) (77). Moreover, the genomic region of 8q24 harboring *PVT1* is one of the most frequently amplified regions in breast cancer (78). Additionally, studies have also found that *PVT1* expression was associated with clinicopathological characteristics and reduced survival times in patients (27,79). What is also worth

mentioning is that in different studies, associated with different types of cancer, it was noticed a correlation of PVT1 with cell proliferation in cancer, in addition to increased cell migration and invasion, after inhibition and overexpression experiments (27,68,80).

4.2.1 Alternative isoforms of PVT1

Last decade, many studies have been focused on the dysregulation of splicing and its pathological effects, especially on the development of tumorigenesis. Transcriptomic analysis in cancer samples, compared to normal tissues, has indicated a significant difference in alternatively spliced transcripts (81,82). Specifically, in a study performed in a large cohort of human tumors, in breast cancer (BRCA), different cancer subtypes can be distinguished based on exon skip splicing features (82).

PVT1 exemplifies a long non-coding RNA (lncRNA) that undergoes complicated alternative splicing events, resulting in the emergence of diverse transcript variants. These transcripts feature polyadenylated tails at their 3' ends and are capped at their 5' ends (60). The complexity of *PVT1* transcriptome is noteworthy, as it comprises of a minimum of 12 exons, generating a multitude of alternatively spliced non-protein coding transcripts. This variety includes 26 circular and 26 linear RNA isoforms, accompanied by 6 miRNAs: miR-1208, miR-1207-5p, miR-1207-3p, miR-1205, and miR-1204 (59,83).

These diverse isoforms of PVT1 play varied roles, and their dysregulation has been linked to the development of various types of cancer. For instance, within the context of aggressive prostate cancer (PCa), one particular variant, PVT1 exon 9, consistently displays higher expression levels in aggressive PCa cell lines compared to non-tumorigenic mother cell lines. This pattern indicates a strong association between the overexpression of PVT1 exon 9 and the aggressiveness of this specific PCa model (84). In addition to these splicing variants, a distinct transcript form of PVT1, referred to as PVT1ΔE4, lacking exon 4, has been identified in renal cell carcinoma (ccRcc). Intriguingly, this variant demonstrates a higher endogenous expression level than the full-length transcript. Notably, PVT1ΔE4 retains the ability to promote cell invasion and proliferation, similar to the full-length transcript (72). In another study, which was focused on a specific isoform of PVT1, named PVT1b, that suppresses *MYC* expression, due to alteration in chromatin conformation, in a p53-dependent manner (85).These findings emphasize the complex landscape of alternative splicing in PVT1, wherein distinct transcript variants contribute to various aspects of cancer biology(86).

4.2.2 PVT1's Interplay with miRNAs in Cancer

PVT1 engagement with miRNAs, particularly while acting as a miRNA sponge, holds significance in various cancer contexts due to the observed correlation in their expression levels. Notably, suppressing *PVT1* expression leads to increased expression of these miRNAs, and conversely, an increase in PVT1 corresponds to reduced miRNA levels (27). In most of the abovementioned cases of cancer development, *PVT1* expression is regulated by a miRNA axis. A notable instance of this interaction is evident in the study by Paci et al., which was based on a previous study of the same team, wherein

computational methodologies unveiled a distinct "sponge interaction network" between PVT1 and the miR-200 family, in breast cancer. Consequently, PVT1 exerts its influence by sequestering miR-200, thereby preventing its interaction with mRNA targets. This dynamic interplay outlines downstream gene expression, underlining a potential role in the progression of disease (87,88). Likewise, a separate investigation within the context of breast cancer has revealed PVT1 modulatory effect on TRPS1 expression. TRPS1, a transcription factor pivotal in various cellular processes, particularly cancer-related pathways, is affected by *PVT1* expression. This is achieved through interaction of PVT1 with miR-543, a known regulator of gene expression that exerts its effect by binding to mRNA molecules, resulting in their degradation or translational suppression. Consequently, this interaction leads in the augmentation of TRPS1 expression, significantly impacting breast cancer development (89).

4.2.3 *PVT1* Chromosomal Aberrations and Fusion Genes in Cancer Development

The 8q24 segment, housing the *PVT1* gene, contains two common fragile sites, FRA8C and FRC8D, making *PVT1* susceptible to chromosomal abnormalities triggered by genetic mutations or factors like hypoxia, viral infections, and cytotoxic drugs (33). These conditions can cause various structural changes in chromosomes, including breakage, amplification, translocation, and more. This genetic instability near *PVT1* can lead to alterations in its copy numbers (59). Through amplification processes, *PVT1* generates new fusion genes, with over 98 distinct *PVT1* fusion transcripts identified (86). These fusions arise from rearrangements within the chromosomal region housing *PVT1* and play a profound role in cancer development and progression. One extensively studied fusion pattern is the *PVT1-MYC* fusion, observed in various cancer types, including colorectal carcinoma, acute myelogenous leukemia, and human gastric cancer. This exon displacement event instigates DNA rearrangements, creating a complex regulatory network involving oncogenes and tumor suppressors that promote tumorigenesis (59,86)

4.3 Interaction of MYC with *PVT1*

The interplay between the *MYC* oncogene and *PVT1* reveals a complex relationship that significantly influences cancer development. Its proximity to *PVT1* within the human genome, a mere 53 kb upstream (Figure 10), underscores their functions and collaborative involvement in cancer progression. Both genes have been independently implicated in malignancies, indicating their collective impact on cancer development (59,90,91).

The *MYC* oncogene, also known as c-*MYC*, belongs to a gene superfamily frequently activated in human cancers. It serves as a master regulator of various biological processes primarily by functioning as a transcription factor, controlling the expression of thousands of genes, either directly or indirectly. MYC has two paralogs, MYCL and MYCN (92). The first evidence of MYC amplification was found in the human leukemia cell line HL60 back in 1982 (93). Initially, it was thought that MYC played a ubiquitous role in human cancers, including both hematological and solid tumors, while L-MYC was associated with small-cell lung cancer and N-MYC with neuroblastoma. However, subsequent studies revealed that most human cancers exhibit genetic activation of at least one MYC family member(92).

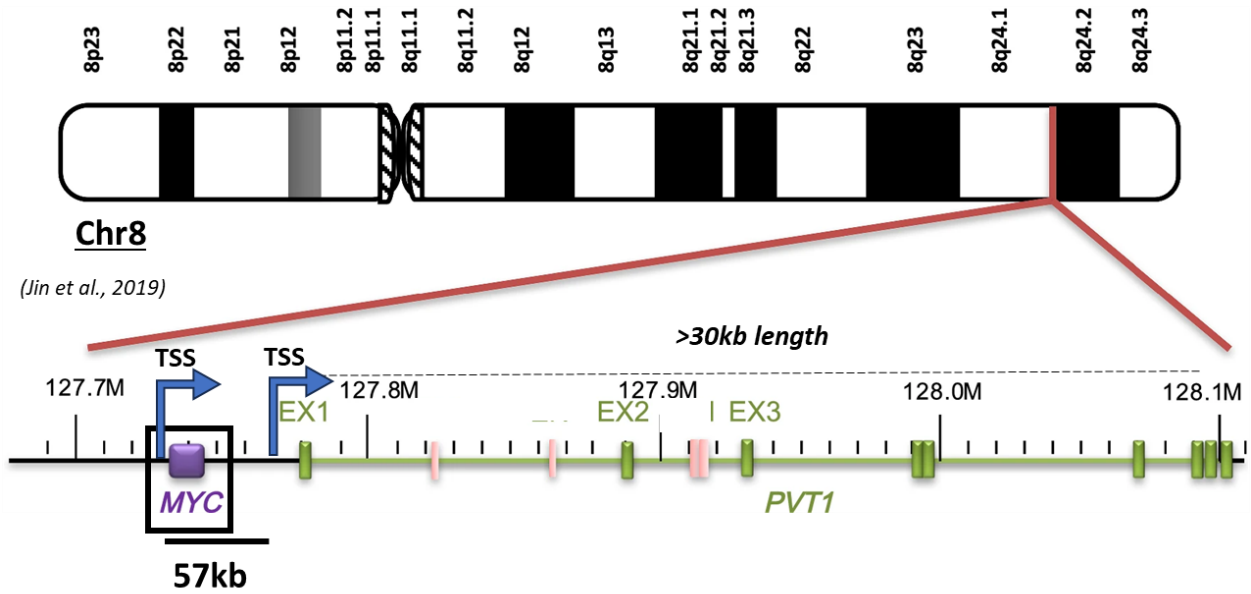

Figure 10. Schematic representation of location of PVT1 and MYC. Figure taken and modified from ref (59).

MYC activation contributes to various cancer hallmarks, including enhanced proliferation, self-renewal, cell survival, genomic instability, altered metabolism, invasiveness, angiogenesis, and immune evasion. *MYC* deregulation can be classified into alterations that render *MYC* RNA or protein unresponsive to negative regulatory signals, resulting in sustained *MYC* expression, or those that elevate the basal levels of *MYC* expression (90,92).

Elevated *PVT1* expression has been found to exert a substantial influence on c-MYC levels, primarily by modulating c-MYC stability. This interaction between *PVT1* and c-MYC represents a dynamic regulatory mechanism, shedding light on their coordinated involvement in malignancy. This relationship is further emphasized through *PVT1* recognition as a MYC target. This interaction hinges on the E-box region of *PVT1* proximal promoter, suggesting a crosstalk between the two. Notably, evidence suggests their potential co-localization within the nucleus, possibly operating within the same molecular complex (58,91).

Numerous cancer-related studies have underscored the correlation between c-MYC and *PVT1*, indicating a parallel upregulation of both genes. Additionally, silencing *PVT1*, by performing siRNA-mediated knockdown experiments in SK-BR-3 and MDA-MB-231, cell lines led to a significant reduction in MYC protein levels while leaving transcript levels unaffected. This intricate interplay is claimed to occur through PVT1 potential role in preventing MYC proteolytic degradation, by decreasing MYC phosphorylation at the Threonine 58 residue (MYCT58) (90,91). In the context of breast cancer, research suggests a synergistic action between MYC and PVT1. They appear to collaboratively promote the expression of RSPO1, a protein engaged in signaling pathways that regulate cell growth and is linked to breast cancer progression. This effect may arise from their synergistic recruitment of transcriptional machinery to the RSPO1 gene promoter or through MYC's impact on *PVT1* expression (94). Another layer of this complex interaction might involve transcription factors. *PVT1* and *MYC* are proximal within the

8q24 locus. FOXM1, for instance, has binding sites in *PVT1* promoter, implying its potential to influence both genes. Increased *PVT1* transcription by FOXM1 correlates with elevated c-MYC levels, suggesting a possible role of FOXM1 in modulating c-MYC expression (91). A noteworthy facet of the MYC-PVT1 relationship centers around the function of *PVT1* promoter as a DNA boundary element. Recent insights highlight the crucial role of DNA boundary elements in segregating distinct gene regulatory domains, a role that can be perturbed in diseases including cancer. Manipulating *PVT1* promoter through CRISPRi experiments led to increased MYC expression, potentially due to the competitive interaction between *PVT1* and *MYC* promoters for enhancer binding. The dynamic interplay between *PVT1* and *MYC* enhancers underscores their interdependence, where *PVT1* promoter restricts *MYC* transcription while PVT1 RNA product sustains MYC protein levels (95).

Objective of the research

Breast cancer is among the most common malignancies and a leading cause of cancer-related deaths in women (96). Various factors, including gender, age, hormonal status, diet, and genetics, contribute to its development (97). Breast cancer is categorized into subtypes based on the immunohistochemical expression of hormone receptors: estrogen receptor positive (ER+), progesterone receptor positive (PR+) and human epidermal growth factor receptor positive (HER2+). Each of these subtypes have different characteristics and responses to treatment. Luminal A tumors are characterized by the presence of ER and/or PR and the absence of HER2 (ER/PR+, HER2-). Luminal B tumors are of higher grade and worse prognosis compared to Luminal A. They are usually HER2 positive (ER/PR+, HER2+). Triple-negative breast cancer is ER-negative, PR-negative, and HER2-negative, with the worst prognosis (97,98)

Non-coding RNAs (ncRNAs), including PVT1, play crucial roles in cancer progression, including breast cancer. *PVT1* is significantly overexpressed in breast cancer and can serve as a prognostic indicator. PVT1 was also found to be strongly linked with tumor size and pathological grade. It acts as a "miRNA sponge" due to the presence of specific miRNA response elements (MREs) withing its sequence by interacting with specific miRNAs, thereby regulating downstream pathways. *PVT1*'s proximity to the *MYC* oncogene further influences its stability and expression (3,54,66,81,84,89)

As it has already been stated, many lncRNAs may stay in proximity to chromatin and in that way affect the expression of genes in specific regions. Thereby, lncRNAs are often enriched in the nucleus and in some cases tethered to chromatin suggesting an involvement in epigenetic regulation. Preliminary experiments aiming to identify parameters that affect chromatin dissociation of lncRNAs, support that splicing efficiency per transcript plays an important role in this process and specifically in the regulation of their target gene expression, as in the case of A-ROD and its activating gene DKK1. The two loci of the target gene and the lncRNA gene are brought into proximity by loop conformation in MCF-7 cells, still the lncRNA is not regulatory functional while chromatin-associated (99). One of the main parameters that affects this association for lncRNAs, based on predictions of machine learning experiments, is the highest splicing efficiency per transcript (Figure 11A).

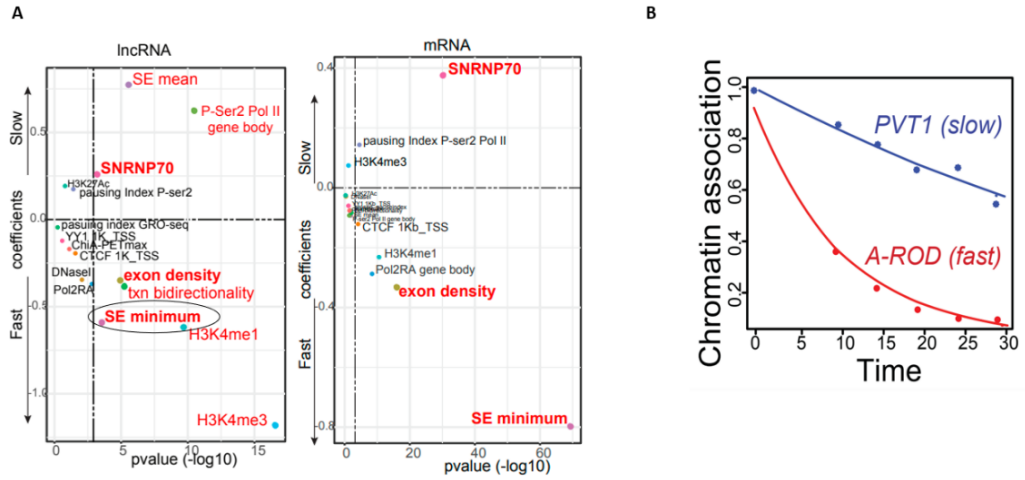

Figure 11. A. Logistic regression with 10 x cross-validation. Coefficients > 0.3 or < −0.3 and with a p-value < 0.001 are marked red. Among the parameters that predict fast chromatin dissociation of nascent RNA transcripts is the efficiency of the worst spliced intron per transcript (Splicing Efficiency (SE) minimum). Representation of chromatin association of A-ROD characterized as fast-released and PVT1 characterized as slow-releases. Figures taken from A. ref (98), B. ref (100).

The aim of this study is to assess the significance of alternative splicing of the lncRNA PVT1, which seems to be a co-transcriptionally poorly spliced and chromatin retained lncRNA (Figure 11B) (100). Our interest is directed particularly in breast cancer, where PVT1 is well-established for its oncogenic role. PVT1 involvement in cancer, along with its dysregulation in various cancer types, underscores its status as a functionally vital lncRNA. Thereby, the hypothesis is that by affecting the chromatin-association of PVT1, by enhancing the splicing, it could affect the expression of *MYC*, that is proximity with *PVT1*, or other interactions, such as with miRNAs.

To achieve this objective, I will employ a newly developed splicing enhancement tool known as CRISPR-Artificial Splicing Factors (CASFx). The experimental focus will be on MCF-7, a breast cancer cell line representing the Luminal A subtype. Initially, I will conduct a comprehensive bioinformatic analysis using published RNA-seq data from MCF-7 and control cell line MCF-10A to identify divergent splicing patterns within PVT1. Subsequently, I will proceed with the construction of CASFx, incorporating guide RNAs targeting specific exons, and evaluate the tool's effectiveness. Lastly, I will investigate the effects of splicing modulation, on *MYC* expression at gene level, based on the enhanced association between *PVT1* and *MYC*. To establish a baseline for comparison, I will include MCF-12A as a control cell line, which is very similar to MCF-10A, in my analysis. This research aims to shed light on the functional significance of alternative splicing in PVT1 and its potential implications in breast cancer. Given PVT1 oncogenic effects on various molecular pathways, it represents a promising therapeutic target. Modulating PVT1 splicing, influenced by its multiple isoforms, could impact cancer development and therapy resistance, including resistance to drugs (86).

Materials and Methods

Materials

1. Chemicals

Reagent	Company	Product Number
1Kb Plus DNA Ladder	NEB	N3200S
Acetic Acid	Honeywell	33209
Acrylamide/Bis Solution	Serva	10687.01
Agar-agar	Sigma	05040
Agarose	Biorad	161-3102
Bromophenol Blue	Sigma	B0126
Chloroform	Merck	1.02445.1000
Dimethyl sulfoxide	Sigma	D4540
Dithiothreitol	NEB	B1034A
dNTPs	Applied Biosystems	4367381
Dulbecco's Modified Eagle's Medium	Gibco	41966
Ethanol Abs	Merck	1.00983.2511
Ethylenediaminetetraacetic Acid	AppliChem	A5097
FBS	Gibco	10437-028
Glycerol	AppliChem	141339.1211
Glycine	AppliChem	A1067
Glycogen-carrier	Thermo Fisher Scientific	R0561
Hydrochloric Acid	Supelco	1.00317
Isopropanol	Merck	818766
Methanol	AppliChem	131091.1212
NP-40	Thermo Fisher Scientific	85125
PBS pellets	Gibco	18912-014
Phenol, acidic	Sigma	P4682
Potassium Chloride	Merck	1.04936
Prestain protein markers	ProteinTech	PL00001
RNasin Ribonuclease Inhibitor	Invitrogen	AM2696
SDS	Sigma	L5750
Sodium Chloride	Supelco	1.06404
TEMED	Sigma	1.10733.0100
Tris ultrapure	AppliChem	A1086
Triton-X-100	Sigma	9036-19-5
Trypsin EDTA 10X	Gibco	15400-054
Tryptone	Sigma	T.9410
Tween-20	Merck	8.22184
Xylene Cyanol	Sigma	X4126
Yeast Extract	Sigma	70161

2. Enzymes - Proteins

Reagent	Company	Product Number
Kapa SYBR Fast Rox Low (50X)	Kapa Biosystems	KD4603
Protein inhibitor cocktail	Sigma	58820-2TAB
Phusion-HF DNA Polymerase	NEB	M0530S
T4 Polynucleotide Kinase	NEB	M0201S
Quick CIP	NEB	M0525S
T4 DNA Ligase	NEB	M0202S
Taq DNA Polymerase	NEB	M0273S
ProtoScript II Reverse Transcriptase	NEB	M0368S
NsiI-HF	NEB	R3127S
BmtI-HF	NEB	R3658S
HindIII-HF	NEB	R3104S

3. Antibodies

Antibody	Company	Product Number
Anti-A2BP1 Polyclonal antibody	ProteinTech	22647-1-AP
Anti- GAPDH Monoclonal antibody	ProteinTech	60004-1-Ig
Peroxidase AffiniPure Goat Anti-Mouse IgG (H+L)	Jackson ImmunoResearch	115-035-146
HRP-conjugated Affinipure Goat Anti-Rabbit IgG(H+L)	ProteinTech	SA00001-2

4. Kits

Kit Name	Company	Product Number
Monarch® Plasmid Miniprep Kit	NEB	T1010S
Monarch® DNA Gel Extraction Kit	NEB	T1020S
Polyplus Transfection jetPEI	Polyplus	101000053/1ML

5. Oligos and primers

Oligo Name	Sequence (5'->3')
Seq1 gBlock 4	TACTCGTACGAATCAGCACGGGTTTGAAACCTTTG
Seq2 gBlock 4	CATGTACAAAGGTTTCAAACCCGTGCTGATTCGTACGAGTA
Seq3 gBlock 4	TACATGGAGCCCTTAAACAAGTAAACCCCTACCAACTGGTCGGGGT
Seq4 gBlock 4	TTTCAAACCCCGACCAGTTGGTAGGGGTTTACTTGTTTAAGGGCTC
Seq5 gBlock 4	TTGAAACCCGTGTCTAGGCATGTTCCCCACAAGTAAACCCCTACCAA
Seq6 gBlock 4	GACCAGTTGGTAGGGGTTTACTTGTGGGGAACATGCCTAGACACGG
Seq7 gBlock 4	CTGGTCGGGGTTTGAAACTTAGGTCTGGTCTTAGGCCTGT
Seq8 gBlock 4	ACAGGCCTAAGACCAGACCTAAGTTTCAAACCCC

Seq1 gBlock 5	TACTCGTACGAATCAGCACGGGTTTGAAACCTACA
Seq2 gBlock 5	GCCACTTGTAGGTTTCAAACCCGTGCTGATTCGTACGAGTA
Seq3 gBlock 5	AGTGGCGCATTCACTAGCAAGTAAACCCCTACCAACTGGTCGGGGTTT
Seq4 gBlock 5	TGTTTCAAACCCCGACCAGTTGGTAGGGGTTTACTTGCTAGTGAATGC
Seq5 gBlock 5	GAAACAGGTGTAACAGGATTCTAGAAACAAGTAAACCCCTACCAACT
Seq6 gBlock 5	CCGACCAGTTGGTAGGGGTTTACTTGTTTCTAGAATCCTGTTACACC
Seq7 gBlock 5	GGTCGGGGTTTGAAACCTAGACTCTGAGAACCAGGTCG
Seq8 gBlock 5	CGACCTGGTTCTCAGAGTCTAGGTTTCAAACC
Seq1 gBlock10	TACTCGTACGAATCAGCACGGGTTTGAAACTATGCAAT
Seq2 gBlock10	CACGTCATTGCATAGTTTCAAACCCGTGCTGATTCGTACGAGTA
Seq3 gBlock10	GACGTGGAGAACAGCAAGTAAACCCCTACCAACTGGTCGGGGTT
Seq4 gBlock10	GTTTCAAACCCCGACCAGTTGGTAGGGGTTTACTTGCTGTTCTC
Seq5 gBlock10	TGAAACCTGAGGGCTTAAGACACTTATACAAGTAAACCCCTAC
Seq6 gBlock10	CAGTTGGTAGGGGTTTACTTGTATAAGTGTCTTAAGCCCTCAG
Seq7 gBlock10	CAACTGGTCGGGGTTTGAAACGCTTTGACCTAGAGGTCTCTTG
Seq8 gBlock10	CAAGAGACCTCTAGGTCAAAGCGTTTCAAACCCCGAC
Seq1 gBlock 11	TACTCGTACGAATCAGCACGGGTTTGAAACCATAA
Seq2 gBlock 11	TGGACCTTATGGTTTCAAACCCGTGCTGATTCGTACGAGTA
Seq3 gBlock 11	GGTCCAGGTGGAGTCATCAAGTAAACCCCTACCAACTGGTCGGGGT
Seq4 gBlock 11	TTTCAAACCCCGACCAGTTGGTAGGGGTTTACTTGATGACTCCACC
Seq5 gBlock 11	TTGAAACGCAGCAGGTTCCTTTAGGTCTGCAAGTAAACCCCTACCA
Seq6 gBlock 11	ACCAGTTGGTAGGGGTTTACTTGCAGACCTAAAGGAACCTGCTGCG
Seq7 gBlock 11	ACTGGTCGGGGTTTGAAACCGTGGGCGATGAAGTTCGTACT
Seq8 gBlock 11	AGTACGAACTTCATCGCCCACGGTTTCAAACCCCG

Primer name	**Sequence (5'->3')**
Fw U6	TTGATGCATgagggcctatttcccatgattcctt
Rv U6	GCTAGCATAAAGCTTTGCTGCTGCTGCTGCTGCTG
Rv PolyT	GCATGCATAAAAAAGCTAGCATAAAGCTTTGCTGC
Rv Cloning_ver	TGTGTTCAGGAACGGTAGTTTGCCCAGTGTATT
Fw gBlock construct	AGTGAAGCTTTACTCGTACGAATCAGCACG
Rv construct gBlock 4	GATAGCTAGCACAGGCCTAAGACCAGACCTA
Rv construct gBlock 5	GATAGCTAGCCGACCTGGTTCTCAGAGTCTA
Rv construct gBlock10	GATGGCTAGCCAAGAGACCTCTAGGTCAAA
Rv construct gBlock 11	GATAGCTAGCAGTACGAACTTCATCGCCCAC
Fw SEval exon 3	GAACCATGCACTGGAATGACAC
Rv SEval exon 4	CAAATCTCAGTGTCCTGGCAGTAAA
Rv SEval exon 5	CGTCCAGAGTGCTGAAAGGATATGTA
Rv SEval span exon4	GTGTAACAGGATTCTACTCCAAATCTCAGTGT

Fw SEval span exon 10	AATAACGGGCTCCCAGATTCACAA
Rv SEval span exon 10	CAGACATGCAAAAAATATAAGTCTATAATAAGCTG
Fw SEval exon 9	GAATAACGGGCTCCCAGATTCA
Rv exon 10	GGTCAGTAGTGATTCCCATAGAAGG
Rv SEval exon 11	GGGCGATGAAGTTCGTACTCATCTT
Fw MYC	TCGGATTCTCTGCTCTCCTCG
Fw MYC Large	GGAGACATGGTGAACCAGAGTTTC
Rv MYC	AGGTGATCCAGACTCTGACCT
Random Primers	NNNNNN
Oligo d(T)23 VN	TTTTTTTTTTTTTTTTTTTTTTTVN

6. Cell lines

The main cell line of our experiment, that will also be used for transfection is MCF-7 that are adherent, epithelial cells, expressing the WNT7B oncogene, representing the Luminal A subtype of breast cancer. MCF-12A is the other cell line that is being used, as a control for our experiments. This cell line is a non-tumorigenic epithelial cell line from tissue taken at reduction mammoplasty from a nulliparous patient with fibrocystic breast disease. It is similar to MCF-10A that the RNA-seq data was conducted.

7. Culturing medium

Full cell culture medium: DMEM high glucose stable glutamine, 5-10% fetal bovine serum (FBS)

Freezing Buffer: DMEM high glucose stable glutamine, 10% DMSO

8. Buffers – Solutions

Annealing Buffer: 5M NaCl, 0.5M EDTA, 1M Tris-HCl pH= 7.9

TAE (50x): Buffer for agarose gel electrophoresis of DNA: Tris base 24.2% w/v, acetic acid 5.71% v/v, EDTA 50 mM, pH 8.6.

Loading buffer (6x): Buffer for electrophoresis of DNA: 0.09% v/v Bromophenol Blue, 0.09% v/v Xylene Cyanol, 60% v/v Glycerol, EDTA 60mM.

Lysis buffer: 20mM Tris-HCl, pH 7.4, 120mM KCl, 1mM DTT, 1mM EDTA, pH 8, 2% v/v NP-40, 1x protease. inhibitors

TRIS- HCl pH 8.8 1.5M: Buffer for polyacrylamide gel preparation.

TRIS- HCl pH 6.8 0.5M: Buffer for polyacrylamide gel preparation.

SDS 10% w/v: Soap buffer for denaturation of proteins in polyacrylamide gel, that also charges them equally with negative charge.

APS 10% και TEMED: Buffers essential for the catalysis of acrylamide polymerization reaction

Transfer Buffer: 25mM Tris, 150mM Glycine, 20% v/v Methanol pH 8.3.

TBS/Tween: 50mM Tris-Cl, pH 7.5, 150mM NaCl, 0.05% v/v Tween 20.

Blocking Buffer: Buffer for blocking the non-specific binding on the membrane: TBS/Tween, 5% w/v fatless powdered milk.

LB medium: 1% Peptone, 1% NaCl, 0,5%Yeast Extract

LB-agar: 1% Peptone, 1% NaCl, 0,5% Yeast Extract, 1,5% agar-agar

Methods

1. CRISPR Artificial Splicing Factors (CASFx)

CRISPR Artificial Splicing Factors (CASFx) is a tool developed as a means to promote splicing in specific splice sites of given transcripts of interest. In this system, dCasRx which is engineered form of Cas13 but without causing RNA cleavage, is fused with RBFOX1 (RNA Binding Fox-1 Homolog 1), essentially replacing the RNA recognition motif (RRM) of the RBFOX1 molecule. RBFOX1 is a splicing activator that belongs to the FOX family of RNA binding proteins and this tool utilizes a pool of guide RNAs or a polycistronic pre-guide RNA, that drives the fused dCasRX-splicing activator to the targeted loci. This system was developed in order to alter the splicing of transcripts, by either targeting a specific exon that leads to exon exclusion, or by binding on a downstream intron that promotes exon exclusion. Targeting of multiple exons, introns, or combination of both is also possible, granting a strong advantage to this system, as it gives the ability to simultaneously affect different loci (57).

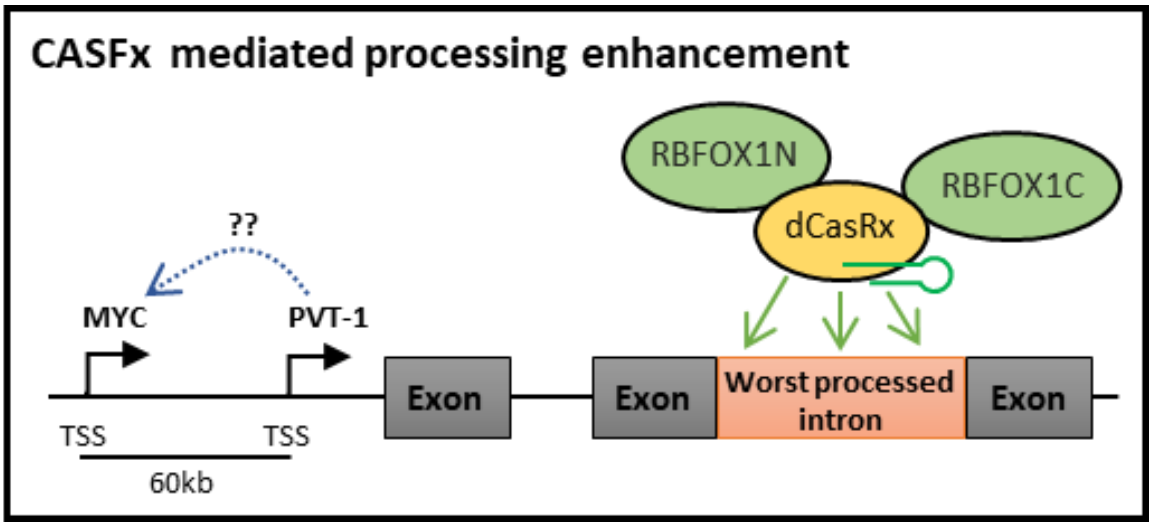

Figure 12. Schematic illustration of (A) MCF-7 cell culture with CASFx plasmid and gRNAs, (B) the principal of this approach. By improving the splicing of the least efficiently processed intron, the anticipated outcome is the release of lncRNA PVT-1 from chromatin, which may potentially influence the regulation of MYC.

2. Construction of CASFx-gBlock plasmids

2.1 Polycistronic gRNA construction

To design the guide RNAs (gRNAs), the CRISPR-Cas9 guide RNA design checker tool from IDT (Integrated DNA Technologies) is utilized. The oligos that would form the polycistronic gRNA (gBlock) are designed as single strand DNA sequences and ordered from Macrogen. The complementary single stranded oligos are annealed in a final concentration of 100mM, using Annealing Buffer, and incubated for 5min in 95°C. The double stranded DNA products, which are designed to have 6nt overhangs, are phosphorylated using the T4 Polynucleotide Kinase (NEB #M0201S). In order to construct the

polycistronic gRNA (gBlock), the phosphorylated duplexes are ligated, using T4 DNA ligase (NEB #M0202S) by incubating them in 14°C for 16 hours. The double stranded product is amplified in PCR using primers with extra overhang in order to incorporate sequences that are recognized by restriction enzymes that will be later needed for the insertion of gBlocks in the CASFx plasmid vector. For the PCR amplification, Phusion-HF DNA polymerase is used (NEB #M0530S) based on the indicated protocol. The final product is analyzed by electrophoresis in a 1% agarose gel, in order to proceed with gel extraction of the correct fragment using Monarch DNA Gel Extraction Kit (NEB #T1020S).

2.2 Plasmid Cloning

The U6 promoter is isolated from the pAC1812_pCR8_gCasRx-CUG plasmid and amplified using primers that introduced NsiI restriction sites and a 3' end poly-T sequence. Phusion-HF DNA polymerase (NEB# M0530S) is used for amplification. The PCR product was gel-purified using the Monarch DNA Gel Extraction Kit (NEB# T1020S). The U6-polyT fragment and pAC1802_pmax-RBFOX1N-dCasRx-C plasmid is digested with NsiI-HF (NEB# R3127S), followed by the phosphorylation of the 5' ends of the U6-polyT fragment and the dephosphorylation of the linear vector using the Quick CIP (NEB# M0525S). Ligation of the plasmid and fragment in a 3:1 ratio is performed using T4 DNA ligase. The resultant plasmid was transformed into DH5A bacterial cells, and plasmid sequences are confirmed via Sanger sequencing. The cloning of the polycistronic gRNAs is done in the same way, excluding the use of different restriction enzymes.

2.3 Restriction Digestion reaction

Fragments and plasmids required for cloning are digested with the appropriate High Fidelity (HF) restriction enzymes. Digestions followed manufacturer protocols, with 1µg of DNA incubated with 1µL of enzyme at 37°C for 1 hour.

2.4 Bacterial transformation

Chemically competent DH5A bacteria are transformed with the constructed plasmids. The ligated product (6µl) is combined with unthawed Dh5a bacteria and incubated on ice for 30 minutes. Heat shock is induced by immersing the tube in a 42°C water bath for 45 seconds, followed by 2 minutes on ice. Pre-heated LB medium was added, and cells are incubated at 37°C for 1 hour. The transformed cells are plated on LB-agar containing kanamycin (50µg/ml) and incubated at 37°C for 16 hours.

2.5 Colony PCR for the selection of the constructed plasmids.

Random colonies from the transformed plate are diluted in LB and stored at 4°C. Colony PCR is performed using Taq DNA polymerase and primers targeting the U6 promoter start and a plasmid region. PCR products are analyzed by electrophoresis in 1% agarose gel, with a product size of 2500bp indicating successful transformation. Selected colonies are cultured in liquid LB with kanamycin (50µg/ml) at 37°C for 16 hours.

2.6 Plasmid DNA isolation

The protocol that is being followed is according to the manufacturer's instructions (Monarch® Plasmid Miniprep Kit, NEB, #T1010S). Around 1-5 mL from E. coli culture that has been grown in a tube are collected by a series of centrifugations at 16,000 x g for 30 seconds. Then the bacteria are resuspended in single cells and lysed by gently inverting the tube 5-6 times. The lysate is incubated for 1 minute at room temperature, and then the Plasmid Neutralization Buffer is added, followed by gentle inverting of the tube and 2-minute incubation at room temperature. After that, the sample is centrifuged for 2-5 minutes at 16,000 x g and the supernatant is transferred to a spin column to be centrifuged for 1 minute at 16,000 x g. The flow-through is removed and the bound plasmid DNA is washed by two different buffers. Then the sample is centrifuged one more time to get rid of residual buffer volume stuck in the column. Following the wash and drying steps, the column is transferred to a clean 1,5ml microfuge tube and more than 30ul of Elution Buffer or sterile water is added to it. After 1 minute of incubation, the sample is centrifuged at 16,000 x g and the plasmid DNA is eluted to the microfuge tube, which is stored at -20°C.

3. Routine of cell culture

3.1 Cell thawing

A cryovial stored in liquid nitrogen is thawed in 37°C until all the nutrient material containing the cells has melted. The cells are then transferred to a 25 cm2 (T-25) flask with a final cell culture medium volume of 6 ml. The culture is maintained in an incubator at 37°C with 5% CO2. The next day, the medium is removed, and fresh cell culture medium is added. This is done to remove all the dead cells, which affect negatively the growth of live cells.

3.2 Re-culturing of the cells

When cells cover from 80% to 100% of the flask or dish surface, before they begin to form double layers, they need to be detached, diluted, and reseeded to cover approximately 10% to 20% of the surface area. For the detachment of MCF-7 cells, treatment with trypsin is necessary. To allow trypsin to act, cells should be initially washed with PBS to remove any remaining nutrient traces. Trypsin-EDTA solution (1x) is added to cover the entire surface of the flask and left for approximately 30 seconds. Afterward, fresh nutrient medium is added to deactivate the trypsin and the cells are detached by gentle tapping. Subsequently, 1/6 of the cells are transferred to a new flask or dish and brought to the appropriate final volume regarding the plate or flask that is being used, with complete nutrient medium (DMEM+5%FBS). The cells are then incubated at 37°C with 5% CO2 for further culturing.

3.3 Cell freezing and storage

Cells of an appropriate confluency in a flask or a plate are treated with Trypsin-EDTA as previously described, centrifuged into a pellet, and then resuspended at a cell density of approximately $5x10^6$

cells/ml in freezing medium which contains 10% (v/v) dimethyl sulfoxide (DMSO), that acts as a cryoprotectant by not letting ice to form and damage the cells during the freezing process. The cell suspension is aliquoted into 1ml cryotubes and transferred to an isopropanol containing freezing box which is stored at -80°C for 16 hours. In this way the temperature drop is slowed to 1°C per minute. Then, the cryovials containing the cells are stored under liquid N_2 in dedicated tissue storage facilities.

3.4 Cell transfection

MCF-7 cell lines, in 70% confluency are cultured in 100mm plate using 6ml DMEM, with additional 5%FBS. After 16hrs of incubation, proceeding with transfection with Polyplus Transfection jetPEI (France) according to the manufacturer's instructions. The cells are incubated for 4hrs at 37°C and then exposed to heat shock, in which the cells are transferred at 42°C for 10 min, and then back to 37°C for 48hrs incubation. 24hrs after the transfection, the medium of the cells is changed to fresh DMEM-5%FBS. In order to evaluate the transfection efficiency, a GFP containing plasmid (px461, Addgene ID #48140) is used, as a positive control in another 100mm plate. The cells are collected after 48hrs of transfection with the constructed GFP plasmid, as proposed by already published studies (57,101). The GFP condition is observed in a UV microscope the next day to validate the transfection efficiency.

4. Cell fractionation and RNA extraction

MCF-7 cells at 90-100% confluency are scraped from 100mm plates, using PBS. The cells are centrifuged at 3000rpm for 5 min, and the whole cell pellet is used. 150µl of freshly prepared cell lysis buffer is used in order to disrupt the cell membranes, incubated for 5 min, strictly on ice to avoid the activity of RNases, and the sample is lied on top of 375µl Sucrose buffer. Centrifugation followed in 16.000g for 10min, in order to separate the nucleoplasmic phase (pellet) from the cytoplasmic (supernatant). The nuclei pellet is used with 800µl PBS-EDTA and then diluted in 250µL PBS. The same volume of acidic phenol is added, followed by intense vortex and incubation of ice for 5min. The samples are centrifuged at 12.500rpm for 15min, in order for the organic and aquatic phases to get separated. The aquatic phase is then collected in a new tube, and the same volume of chloroform-phenol is added. The same procedure followed and collected the aquatic phase that included the nuclear RNA. For the precipitation of the nuclear RNA, 10% of the volume of the sample NaOAc is added to the sample, as well as 3x this volume of etOH 100% and 1µL of glycogen-carrier. The next step is the snap-freezing of the samples in liquid nitrogen, and incubation at -20°C O/N. The nuclear RNA pellet is washed with etOH 70% after the centrifugation and diluted in nuclease-free water. The concentration of RNA is then measured with the use of a Nanodrop.

4.1 Reverse Transcriptase (RT) PCR

From the nuclear RNA extracted from the cells, 1µg is used as a template for the RT-PCR reaction. cDNA is synthesized in a two-step process using ProtoScript II Reverse Transcriptase (NEB #M0368S),

using Oligo d(T)23 VN and Random mix primers and in the presence of RNasin Ribonuclease Inhibitor (Promega), in a final volume of 20µL, following the protocol from NEB.

4.2 Polymerase Chain Reaction

In order to validate that the CASFx tool is effective, and the splicing of the targeted fragments occurred, as well as to check the different pattern of splicing efficiency between cell lines, Polymerase Chain Reaction (PCR) is used, using primers binding on the regions of interest. For this reaction, Taq DNA polymerase (NEB# M0273S) is used following the manufacturer's instructions to a final volume of 25µL. As a template, cDNA from reverse transcription reaction is used, with an amount of 80ng of the initial RNA. The samples are loaded in 1% agarose gel for electrophoresis.

4.3 Real time PCR (qPCR)

For the real time PCR (quantitative PCR, qPCR), the KAPA SYBR Fast qPCR kit is used. cDNA derived from 80ng of total RNA is used as template of the reaction. For the calculation of the fold change and analysis of the results (Ct values), the 2-ΔΔCT method is utilized. In this way, the relative changes in gene expression in the sample under study compared to the reference sample (calibrator) are calculated. The data represent the fold change in the expression of a gene that has been normalized to a reference gene (normalizer) and is compared to an untreated control sample (calibrator). The reference gene that is used is GAPDH.

5. RNA-sequencing data analysis

For the computational analysis of these data, GRCh38/hg38 is used as a human genome reference and Gencode v41 is used as gene annotation. The alignment of the reads, for all the conditions, is conducted using STAR tool (version 2.7.10a). The files are trimmed by keeping only the ones with STAR score == 255, which corresponds to the uniquely mapped reads, and proceeded with the strand correction and normalization of the files to create reads per million of uniquely mapped reads. For the splicing efficiency analysis and in order to extract the statistically significant regions, R script is used as a tool. In order to acquire the split and non-split read, the parameter -cigar is used in bedtools.

6. Western Blot

MCF-7 transfected cells with plasmid containing the gBlock are used for protein extraction. The cells are centrifuged to a pellet and lysed with radioimmunoprecipitation assay (RIPA) buffer. In order to facilitate the breaking of the membranes, the samples are submerged into liquid nitrogen and then back to 37°C incubator for 5 times, following a step of sonication. The samples are centrifuged in order to get rid of the debris. The calculation of sample concentrations is done by using the Bradford reagent and measured at 595nm, after first constructing a standard curve with known quantities of protein, using a 0.1% BSA solution (or 1µg/µl). The proportions used are 2µl of sample in 198µl of sterile ddH2O, followed by the addition of 800µl of the Bradford reagent.

Equal amount of protein samples is diluted in loading dye that contains β-mercaptoethanol and boiled for 5 minutes at 95ºC. SDS-polyacrylamide gel electrophoresis (PAGE) is used in order for the proteins to be denatured and separated by size. Specifically, the upper part of the gel (stacking gel) contains 4% acrylamide and has a pH of 6.8. The composition of the lower part of the gel (separating gel) contains 10% acrylamide, with a pH of 8.8. Electrophoresis is performed at 80V for 2.5-3 hours in 1x Running Buffer.

After that, proteins are transferred by electrophoresis from the SDS/polyacrylamide gel (SDS-PAGE) to a nitrocellulose membrane. In sequence, 2 pieces of Whatman 3mm paper, the separating gel, the membrane, and 2 pieces of Whatman paper (all soaked in Transfer Buffer) are placed, ensuring that no air bubbles are trapped. Subsequently, the "sandwich" is placed in the electrophoresis apparatus with the membrane facing the anode. Electrophoresis is conducted under 150A at 4°C for 2hrs.

Afterwards, the membrane is incubated in a blocking buffer (5% non-fat milk) at room temperature for 2 hours to block non-specific binding sites on the membrane. Following the blocking step, the membrane is washed with TBS-Tween solution three times for 10 minutes each. Then overnight incubation follows with a primary antibody that targets the protein of interest, diluted in 2% non-fat milk. The next day, the membrane is incubated with the appropriate secondary antibody, which is conjugated with Horse Radish Peroxidase (HRP), diluted in 2% non-fat milk.

Results

1. Bioinformatic analysis from RNA-seq data

To validate differences in expression and splicing patterns of *PVT1* between breast cancer cells and normal breast cells, I initiated the analysis using published available data from the MCF-7 cell line, representing the ER+ breast cancer phenotype, and the MCF-10A cell line, representing the corresponding normal phenotype. The dataset used was retrieved from the Gene Expression Omnibus (GEO) under accession number GSE71862, and each cell line had three replicates for analysis.

In the initial phase of the analysis, I focused on *PVT1* expression and observed elevated levels in the MCF-7 cell line when compared to MCF-10A, across all three replicates (Figure 13)To standardize these findings, I quantified read counts for each replicate and then normalized them against the total number of uniquely mapped reads (Figure 13). This normalization step was crucial in preventing differences due to variations in library sizes. The graph displayed in Figure 13 represents the average of these normalized results, in conjunction with the visualization from UCSC (Figure 14).

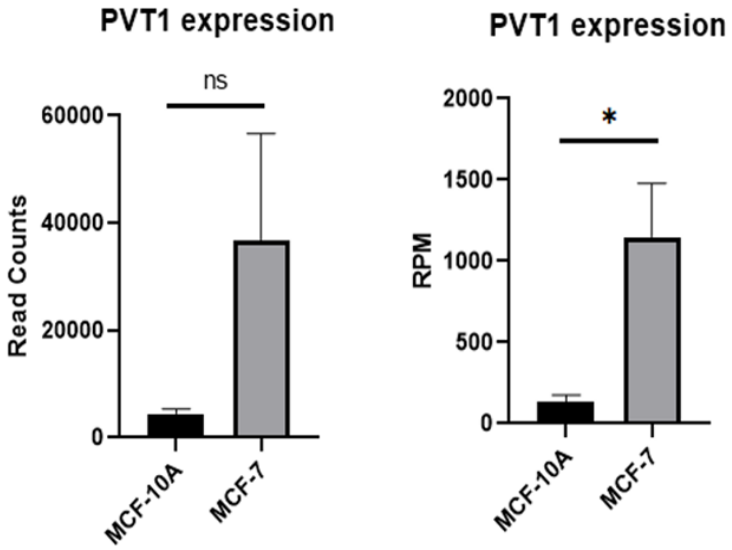

Figure 13. Representation of the expression levels of PVT1, between MCF-7 and MCF-10A cell lines, acquired from published RNA-seq data. On the left is the read counts of PVT1 expression from 3 replicated for each cell line and on the right is the normalization of read counts to uniquely map reads per million (RPM).

As depicted in the following figure (Figure 14), there is an increase in the expression of intronic regions in MCF-7 compared to MCF-10A, highlighting a reduction in splicing efficiency. Consequently, in the subsequent phase of the analysis, the attention was directed to evaluating the splicing efficiency of PVT1 exons in order to detect statistically significant differences between the two cell lines. The analysis was based on the ratio of split reads to the total of split and non-split reads, derived from the alignment to the genome reference, as illustrated in the schematic (Figure 15). The results acquired can range from 0, that means not spliced, to 1, that indicates completely spliced (102). To ensure the reliability of these findings, I implemented a threshold with a p-value of less than 0.05. Additionally, I refined the results to encompass AG nucleotides at the upstream part of introns (donor splice site) and GT nucleotides at the downstream part (acceptor splice site). These sites are recognized as canonical splice sites and are

targeted by the predominant class of spliceosomes, notably the U2-type spliceosome, which is widely employed in eukaryotic splicing processes (103).

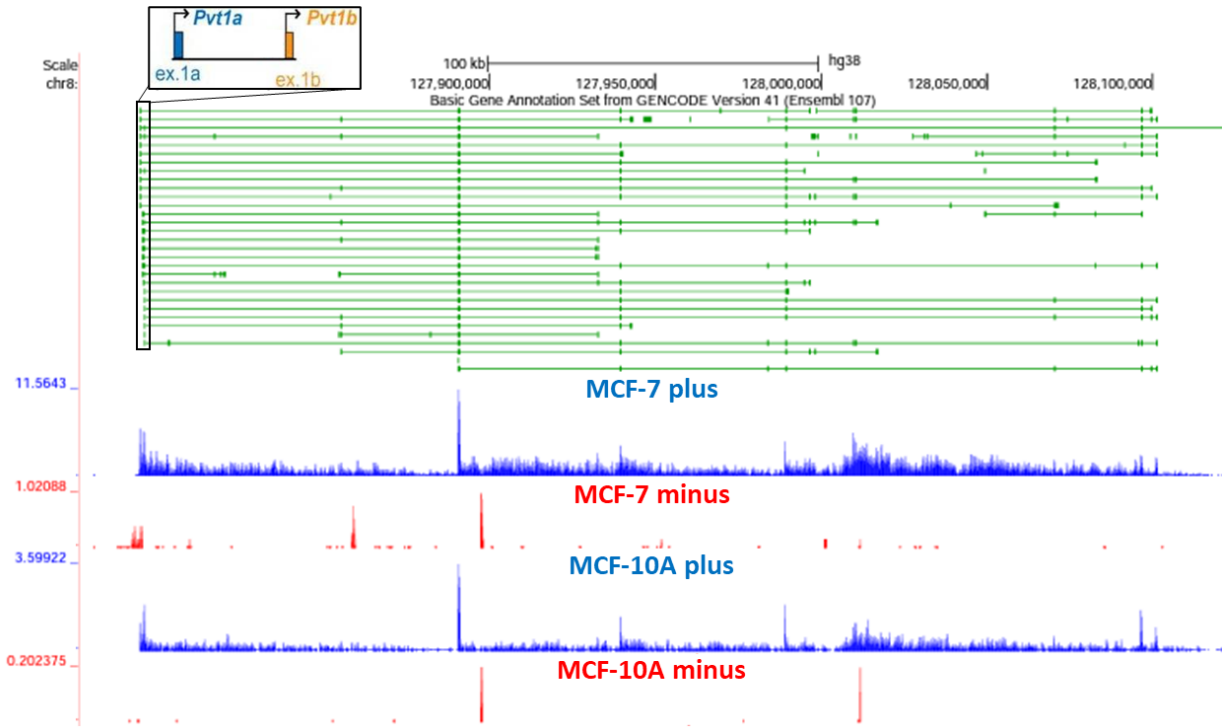

Figure 14. Screenshot of *PVT1* expression levels on UCSC, along with the transcript coordinates of PVT1. The distinct first exons of transcript variants PVT1a and PVT1b are also highlighted.

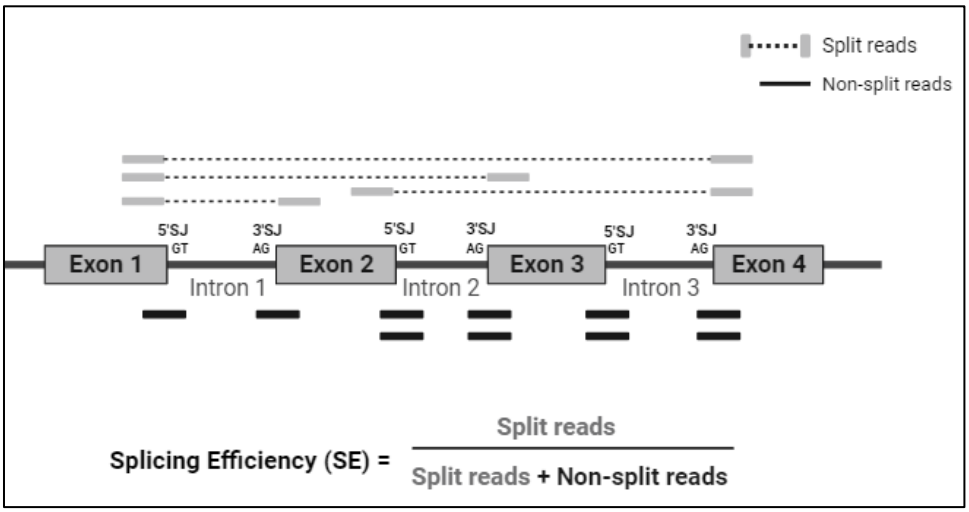

Figure 15. Schematic illustration of the splicing efficiency evaluation.

The analysis revealed that there were no statistically significant splicing events at the 5' splice sites of PVT1 transcripts between the two cell lines. However, we identified 16 isoforms associated with the 3' splice sites. Among these isoforms, those with the highest statistical significance and greater abundance in the various PVT1 transcripts were selected to proceed.

Based on this analysis, I designed guide RNAs to incorporate in the utilization of the CASFx tool (57), intending to specifically target sites displaying distinct patterns between the two cell lines. The initial focus was on augmenting splicing within a particular exon exhibiting a statistically significant difference in splicing efficiency. This exon corresponds to the genomic coordinates chr8:127939505-127939676. In order to facilitate the process, I named the exons based on the ENST00000657449.1 isoform, which has 11 exons in total (Figure 16A). So, the exon of interest, with a proven significance, is exon 4. Additionally, exon 10, which is located near the 3' end, is also being selected as a target. This exon could potentially be linked to PVT1 chromatin association, as 3' end cleavage and polyadenylation in nascent transcripts is associated with chromatin-retention (99,104). The expression patterns between the two cell lines, along with the positions of the guide RNAs, are provided below (Figure 16B,C).

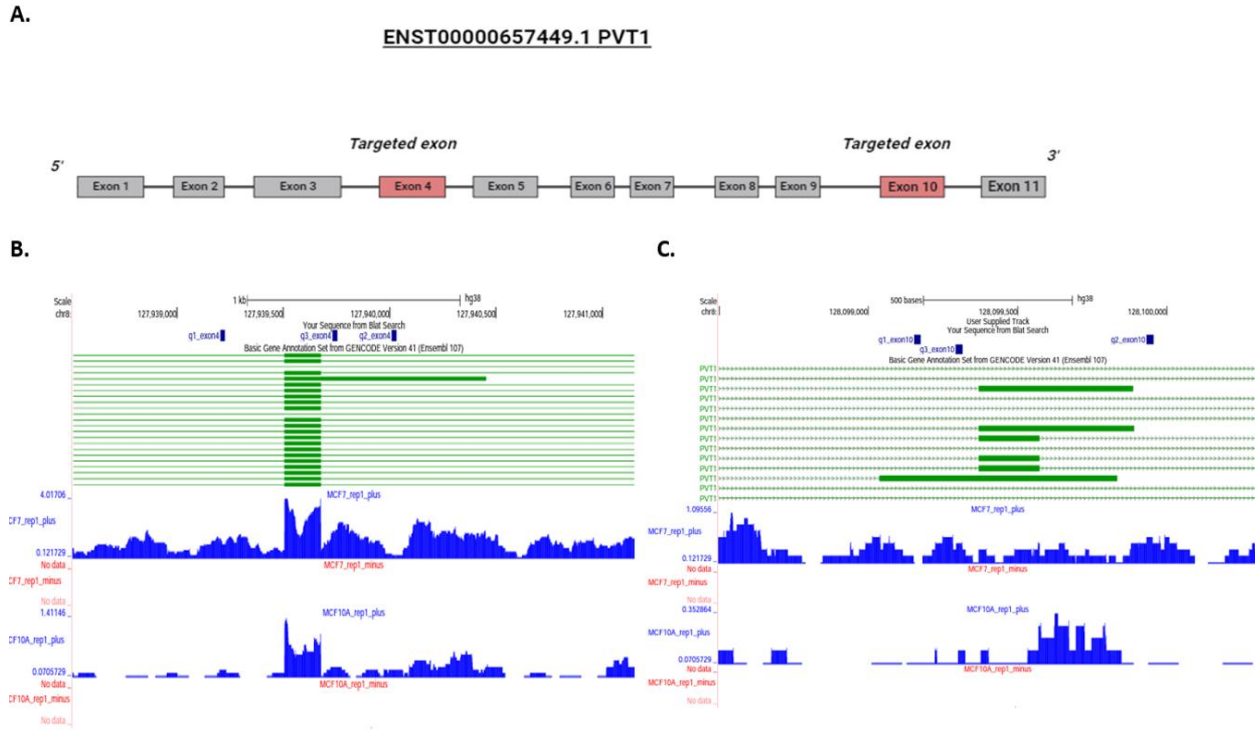

Figure 16. (A) Schematic illustration of the exon sequence of ENST00000657449.1, highlighting the specific exons targeted for modification (in red). (B) UCSC genome browser visualization depicting the expression levels of exon 4 in both MCF-7 and MCF-10A cell lines, with guide-RNA target sites indicated. (C) UCSC genome browser visualization showing the expression levels of exon 10 in MCF-7 and MCF-10A cell lines, with guide-RNA target sites highlighted.

2. Construction of the modified CASFx plasmids

2.1 *In silico* design and assembly of the guide RNA block inserts

Based on Du et al., (57) three guide RNAs (gRNAs) were selected for each target exon to improve splicing efficiency and precision. These gRNAs were structured as a polycistronic pre-gRNA with a direct repeat (DR) sequence positioned between each gRNA, creating a guide-RNA block (gBlock). The DRs

facilitate the guidance of Cas13d to its target and specifically the Cas13d direct repeats exhibit a high degree of conservation in terms of length and predicted secondary structure, consisting of a 36 nt length, an 8-10 nt stem with an A/U-rich loop, and a 5'-AAAAC motif at the 3' end of the direct repeat. Previous studies have shown that this conserved 5'-AAAAC motif is specifically recognized by a type II Cas1/2 spacer acquisition complex (82).

Apart from the gBlock designed for the specific exon, an additional gBlock was generated to target sites near the neighboring exons, namely exon 5 and exon 11. This additional step aims in potential further enhancement of splicing in these intronic sites, which appear to be less efficiently spliced in MCF-7 compared to MCF-12, ultimately promoting a stronger retention of these exons and more effective splicing. Consequently, the entire cassette, including both DRs and gRNAs, was designed and ordered as single-stranded oligonucleotides.Through this approach, the single-stranded DNA (ssDNA) have been annealed to double-stranded DNA (dsDNA), incorporating 6nt overhangs on each strand of the dsDNA. These overhangs are complementary to the next and previous dsDNAs, facilitating the assembly of the complete cassette through a ligation process. An intermediate step involving phosphorylation of each dsDNA at the 5' site is required for the ligation of these parts to occur efficiently.

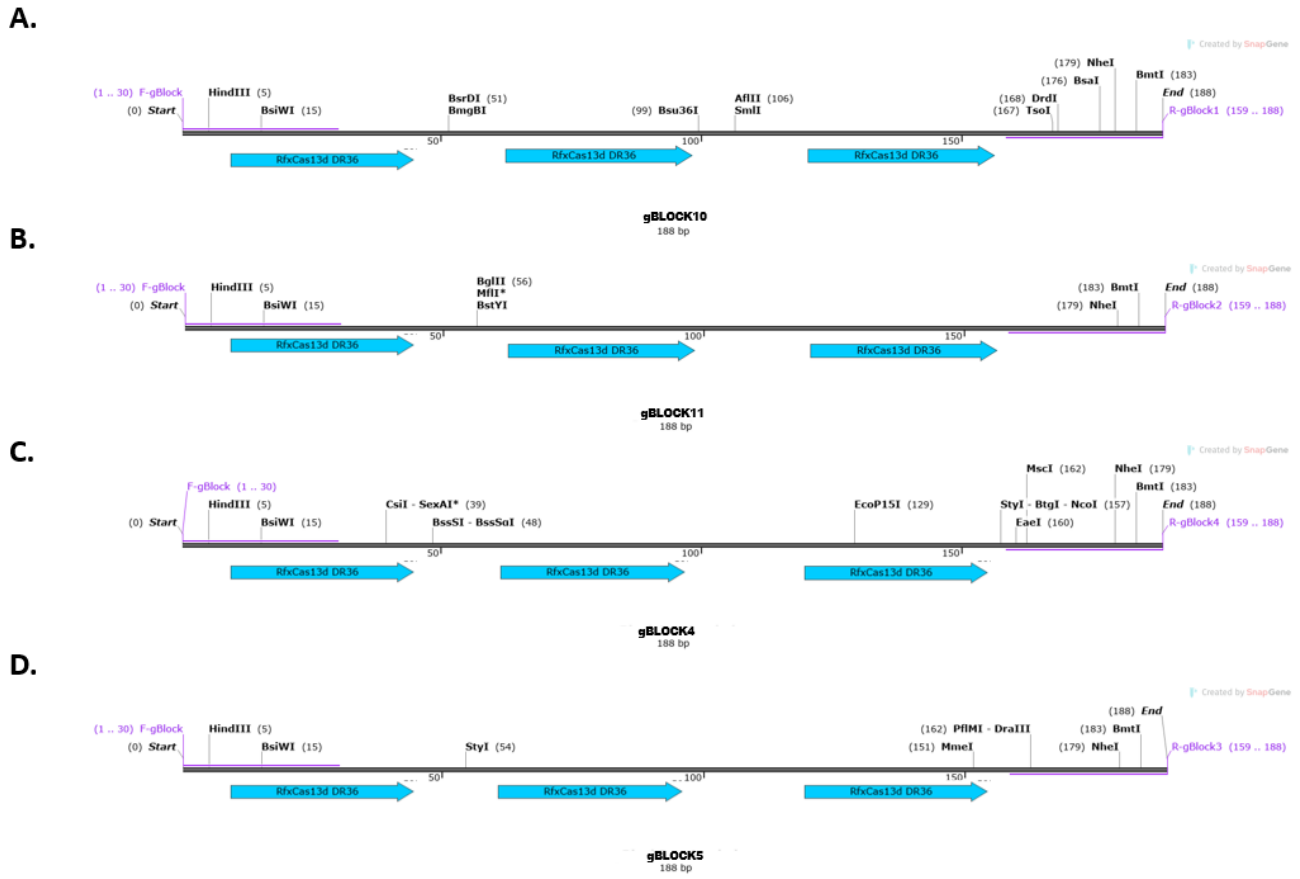

Figure 17. Schematic illustration of the *in-silico* design of the gBlocks. (A) Illustration of gBlock10 that targets exon10. (B) Illustration of gBlock11 that targets exon11. (C) Illustration of gBlock 4 that targets exon4. (D) Illustration of gBlock 5 that targets exon5.

Each cassette was amplified using primers containing overhangs for BmptI and HindIII restriction sequences, at each site. These overhangs facilitate the insertion of the cassette into pAC1802_pmax-RBFOX1N-dCasRx-C, in which I cloned the U6 promoter. Schematics of the final gBlock products are presented above (Figure 17).

2.2 Construction of the modified CASFx vector

Du et al. suggested that including both the fused RBFOX1-dCasRx, and the polycistronic guide RNA (gRNA) in the same plasmid enhances splicing efficiency (57). Therefore, pAC1802_pmax-RBFOX1N-dCasRx-C (Addgene #118635) was utilized as the basis of the final vector. The plasmid was modified by inducing a U6 promoter and the polycistronic gRNA. The U6 promoter sequence was obtained from pAC1812_pCR8_gCasRx-CUG (Addgene #118645) through PCR amplification, utilizing specific primers for this region, in order to isolate this specific fragment.

During the PCR process, an overhanging restriction site for NsiI was added to the forward primer, while the reverse primer contained overhanging restriction sites for BmtI and HindIII. The NsiI restriction enzyme is necessary for inserting the U6 promoter into the plasmid already containing RBFOX1-dCasRx, and the BmtI and HindIII restriction sites will be utilized for inserting the polycistronic gRNA into the final, modified CASFx vector. The expected product of 341bp was extracted after agarose gel electrophoresis. Picture has not been acquired in order to minimize the UV exposure on the U6 promoter sequence fragment, thus avoiding random strand breaks that could result in mutations.

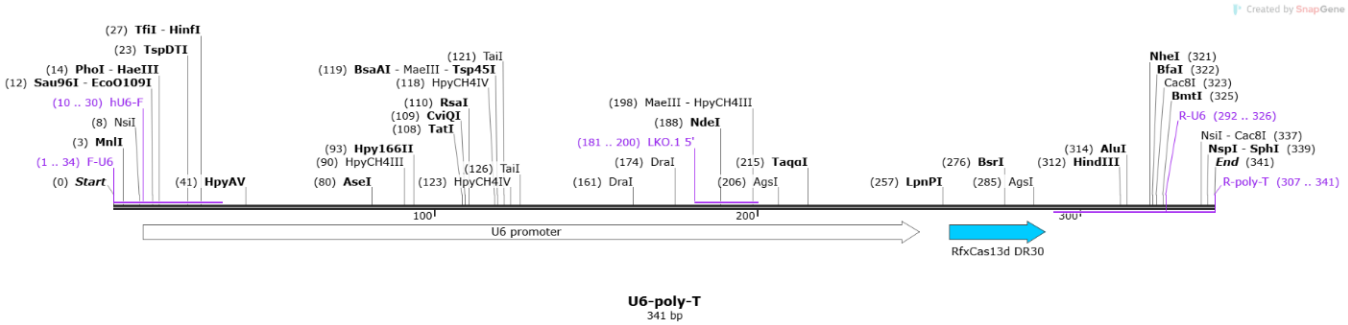

Figure 18. Schematic illustration of the *in silico* designed U6 promoter sequence, along with the added restriction sites and a poly-T sequence, that is associated with transcription termination.

Subsequently, a second PCR was performed using the U6 promoter as a template. This step aimed to incorporate a poly-T sequence, which serves as the transcription termination signal for PolIII (105). An additional overhanging restriction site for NsiI was also added during this PCR. The final product resulting from these procedures is depicted in the figure above (Figure 18).

The NsiI restriction enzyme will be employed to cleave both the pAC1802_pmax-RBFOX1N-dCasRx-C plasmid and the resulting U6-polyT fragment, facilitating the subsequent insertion by ligation of the fragment into the plasmid. The schematic of the final cloned plasmid is depicted in the image provided below (Figure 19).

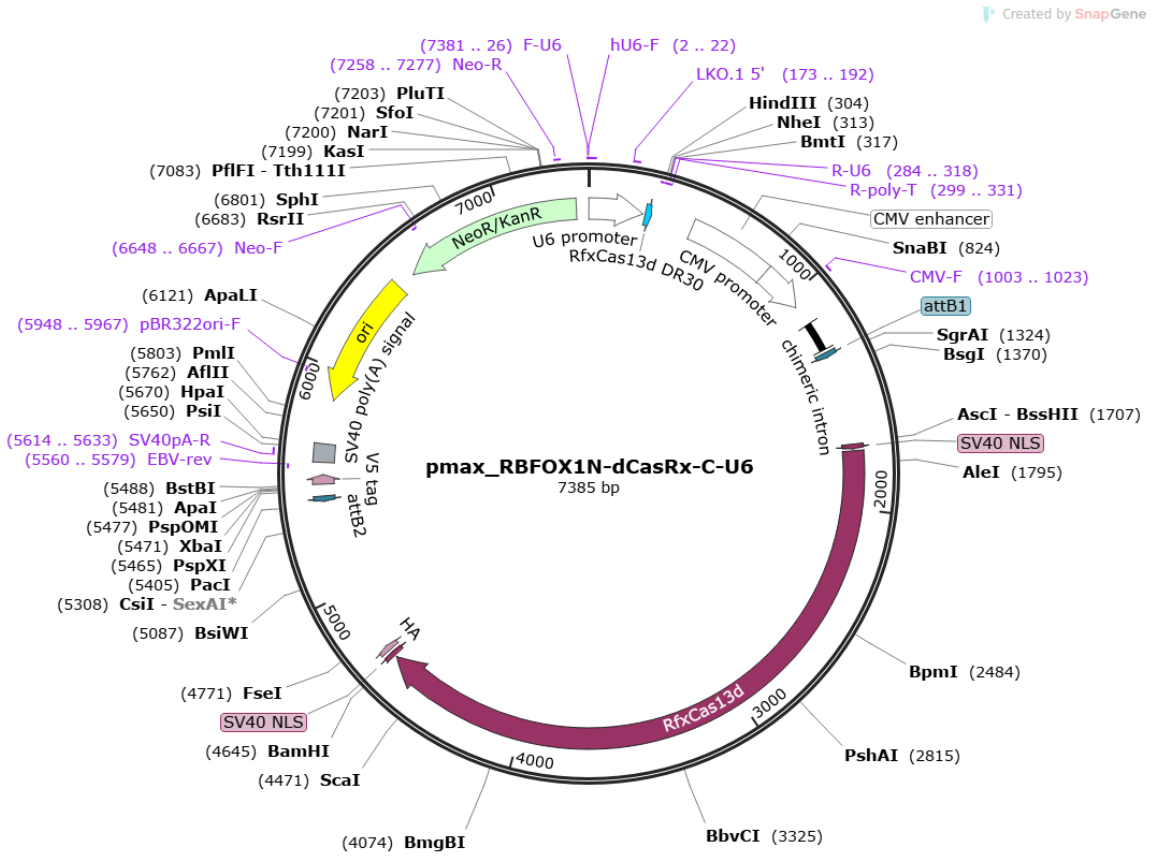

Figure 19. Schematic illustration of the pAC1802_pmax-RBFOX1N-dCasRx-C plasmid after the insertion of U6 promoter. This plasmid includes restriction sites for the insertion of any gBlock.

The ligation product was used to transform Dh5a bacteria, and the resulting clones were analyzed to confirm the presence of the desired fragment. To validate the incorporation of the U6 sequence in the plasmids, colony PCR was conducted on eight of these clones. The anticipated product from the colony PCR is shown in the schematic presented below (**Figure *20***). The colony PCR results revealed that only clone 6 (C6) displayed the expected 1532bp band, indicating successful incorporation of the U6 sequence, as well as the correct orientation. Subsequently, clone 6 was cultured for further validation (Figure 21).

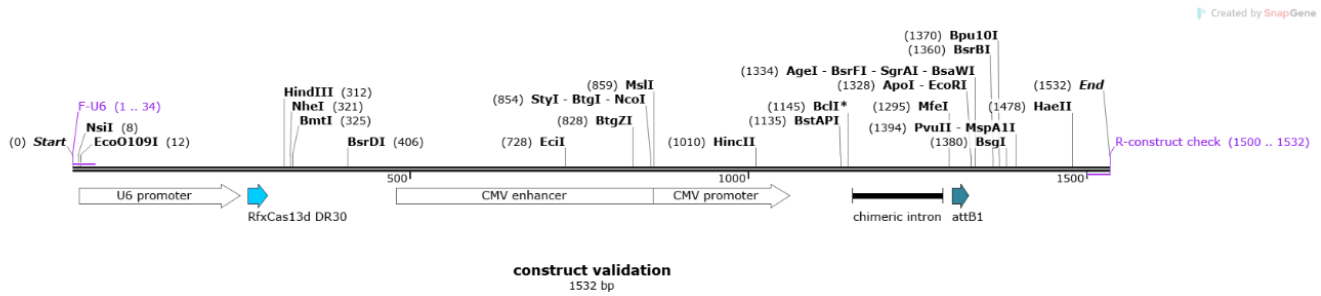

Figure 20. Representation of the validation construct that will be expected after PCR amplification, in order to validate the correct orientation and insertion of U6 promoter.

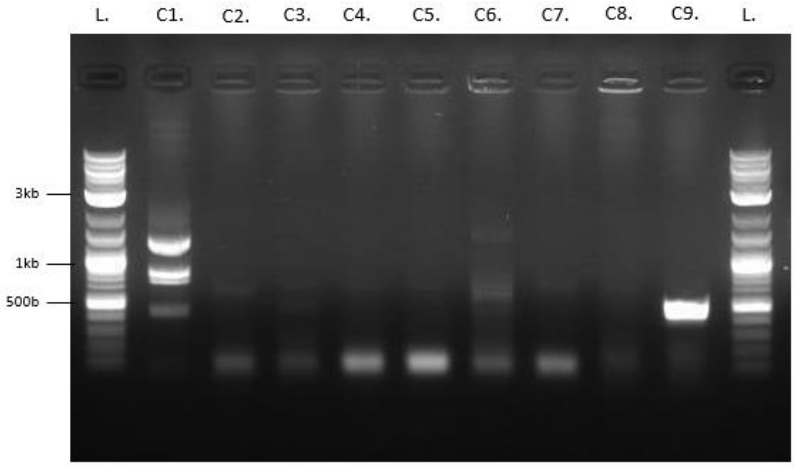

Figure 21. Agarose gel electrophoresis (1,2%) after colony PCR for the validation of incorporation of the U6-polyT sequence in the pAC1802_pmax-RBFOX1N-dCasRx-C plasmid. Only clone 6 (C6) seems to have a band of the expected size.

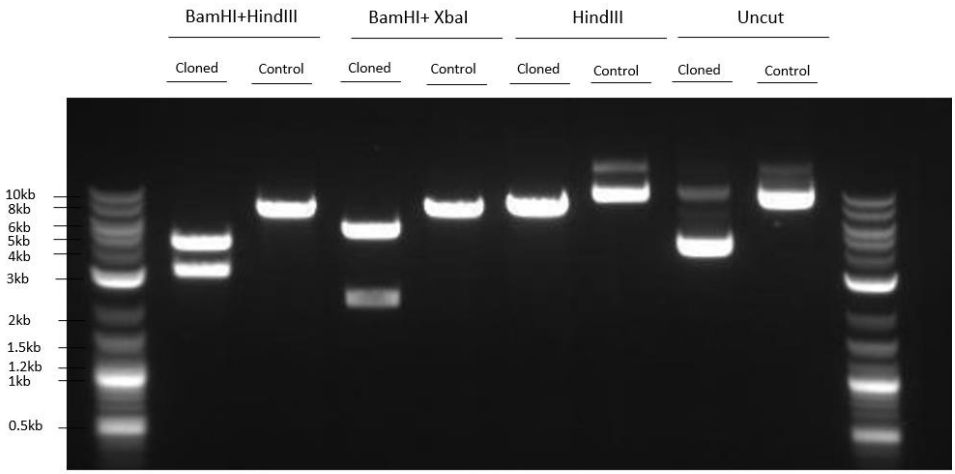

Figure 22. Agarose gel electrophoresis (1%) for the validation of clone 6, with restriction enzymes after insertion of the U6 fragment. The control plasmid represents the pmax_RBFOX1N-dCasRx-C, that does not include restriction sites for HindIII.

The validation of the constructed plasmid was further assessed through a double restriction digestion reaction (Figure 22). Two different reactions were conducted, one with HindIII and BamHI enzymes and the other with HindIII and XbaI enzymes. Restriction sites for XbaI and BamHI were present in the initial plasmid, as indicated by the map of the plasmid. The expected fragment sizes for these reactions were 4341bp + 3044bp and 5164bp + 2218bp, respectively. To serve as a control, the initial pAC1802_pmax-RBFOX1N-dCasRx-C plasmid, lacking the HindIII restriction site, was also tested with these enzymatic restrictions. This control plasmid should produce a linear band of the expected size of 7055bp. The control plasmid displayed linear bands (lanes 2 and 4) or remained uncut when a single enzyme reaction was performed (lane 6) due to the absence of the HindIII restriction site in its sequence. Furthermore, it can be observed that the plasmid from clone 6 appeared slightly higher in the gel, indicating a larger size in kilobases compared to the control plasmid (comparison between lanes 4 and 5). This can be attributed to the incorporation of an additional 341bp fragment into the plasmid.

2.3 Cloning of the guide RNA blocks into the CASFx vector

As soon as clone 6 was validated to be correct, it served as a template for the insertion of each distinct gBlock. As an important reminder, during the PCR amplification of the plasmid sequence, restriction sites for BmtI and HindIII were incorporated, like those present in the gBlocks. Both the modified CASFx vector and the gBlocks were subjected to a double restriction digestion using these two enzymes. Subsequently, the plasmid underwent dephosphorylation and purification by gel extraction, while the gBlocks were phosphorylated before ligating the linear fragments together. Again, picture for the dephosphorylated vector has not been acquired in order to minimize its UV exposure. The resulting plasmid was then transformed into Dh5a bacteria on an agar plate with kanamycin, generating multiple clones. These clones were subsequently tested to confirm the presence of the intended plasmid. Similarly, colony PCR was performed on selected plasmids, targeting the same region as before, with an expected size of 1385bp.

The results from the electrophoresis in 1% agarose gel, indicated that, in case of gBlock4, only clones 3 and 8 had been incorporated the constructed fragment regarding the expected size (Figure 23). The correct sequence of the plasmid containing gBlock4 was further validated with Sanger Sequencing (Figure 24). The same process was followed for the rest of the gBlocks constructed (Figure 34).

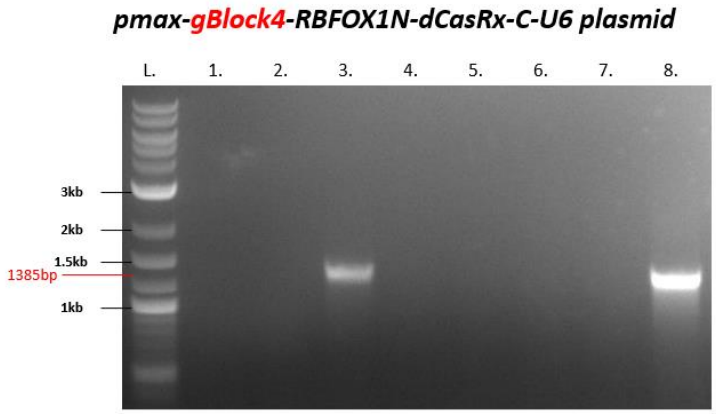

Figure 23. Indicative image of gel electrophoresis after colony PCR in 1% agarose gel. Each lane represents one of the tested clones of Dh5a with the plasmid that includes both U6 sequence and gBlock.

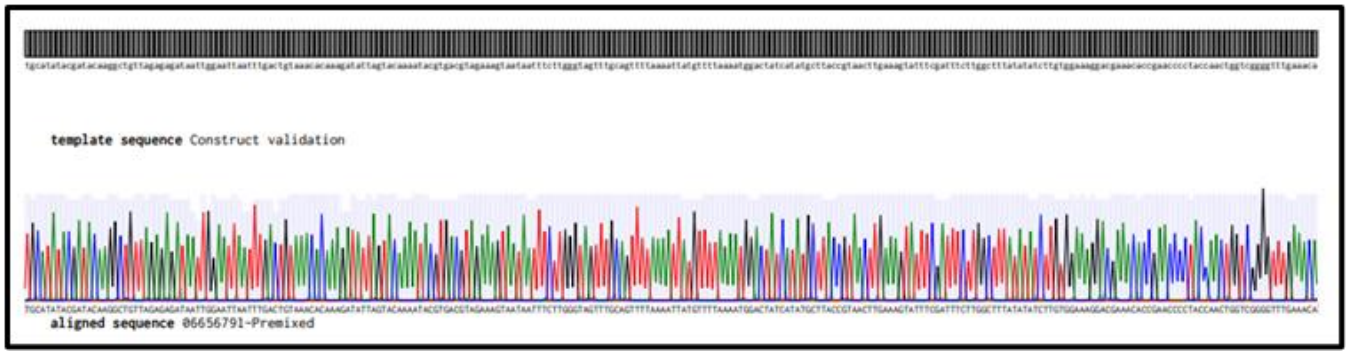

Figure 24. Screenshot from Benchling showing the Sanger Sequencing results for the modified CASFx vector cloned with the gBlock associated with the targeting of exon 4.

3. Validation of the transfection of CASFx plasmid by testing RBFOX1 expression

The modified CASFx plasmid that was engineered contains the RBFOX1 catalytic subunit, fused with the dCasRx subunit, along with the gRNAs and the necessary promoters for their expression. To confirm the expression of RBFOX1-dCasRx in the MCF-7 cell line after the transfection of the cells with the respective plasmid, as well as to determine the time point of its maximum expression, we utilized Western Blot analysis for the RBFOX1 protein at various time intervals: 0hrs, 24hrs, 48hrs, 56hrs, and 72hrs. In parallel, untransfected MCF-7 cells were used as a control. To standardize this expression, GAPDH, a housekeeping protein, was used as a control. Since RBFOX1 protein is modified and fused with dCasRx protein, the observed molecular weight is expected to be a lot higher than its normal molecular weight. Specifically, RBFOX1 is typically observed at 45-70kDa, but now based on the nucleotide sequence the expected molecular weight would be approximately 130-150kDa (Figure 25).

The results, as depicted in the figure, reveal no noticeable RBFOX1 expression in either the control or at the first time point (t=0hrs). At t=24hrs, the lowest expression was observed compared to other time points. Interestingly, there was no significant difference in RBFOX1 expression between t=48hrs, t=56hrs, and t=72hrs. These findings confirm the successful expression of the engineered plasmid post-transfection and indicate that the peak expression of the target protein occurs between 48hrs and 72hrs (Figure 25).

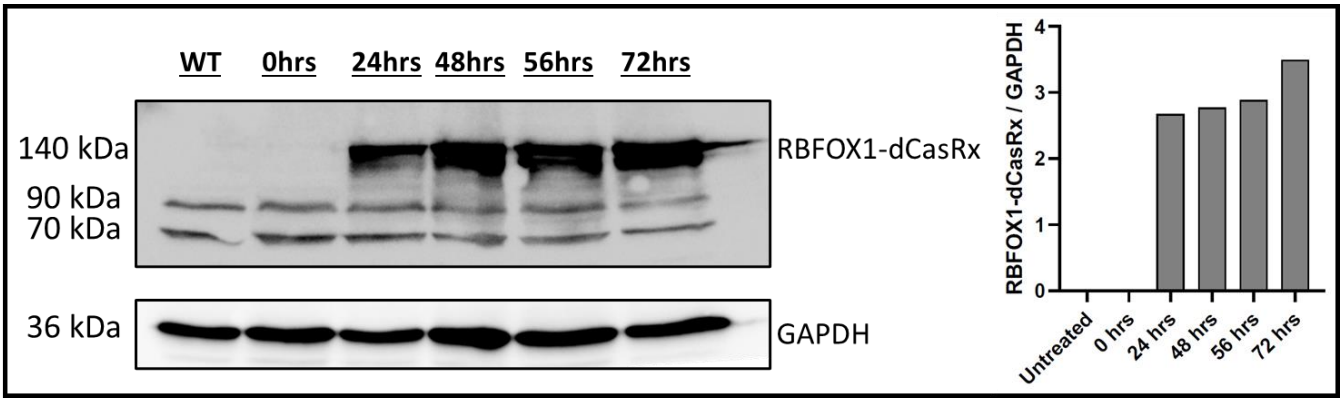

Figure 25. Western Blot for the validation of RBFOX1-dCasRx expression after the transfection of MCF-7 with the constructed CASFx plasmid, compared with wt MCF-7 (untransfected) in different time points: 0hrs, 24hrs, 48hrs, 56hrs, 72hrs. The expected band of the RBFOX1-dCasRx fusion is expected at ~140kDa. The results are compared to the expression of GAPDH. On the right is a quantitative representation of RBFOX1-dCasRx protein, compared to GAPDH expression levels.

4. Validation of splicing efficiency after CASFx

In order to get a first indication whether the implementation of the CASFx tool is functional in my model cell line, and additionally, to contrast the expression patterns of the exon 4 between MCF-7, MCF-12A and transfected cells, I performed PCR reactions. For this experiment different sets of primers were designed to be complementary either only on the targeted exon, or at the downstream exon. The design and naming of the produced fragments are represented below (Figure 26). The targeted exon 4 corresponds to the exon B of the scheme.

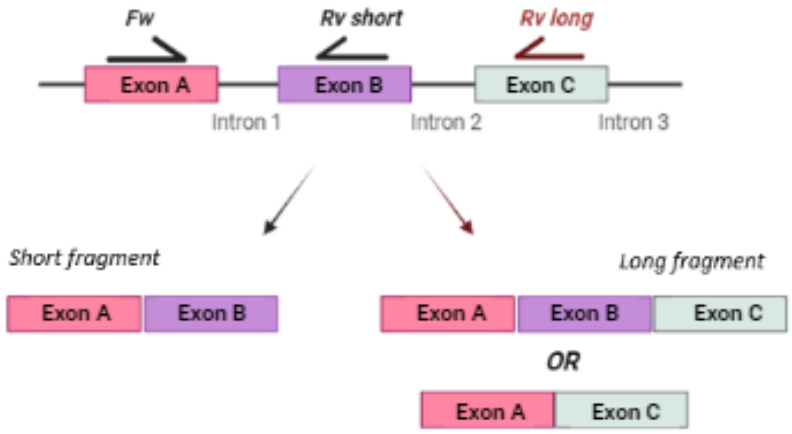

Figure 26. Schematic illustration of the design of primers, with their expected products, in order to validate the result after the enhancement of splicing with CASFx.

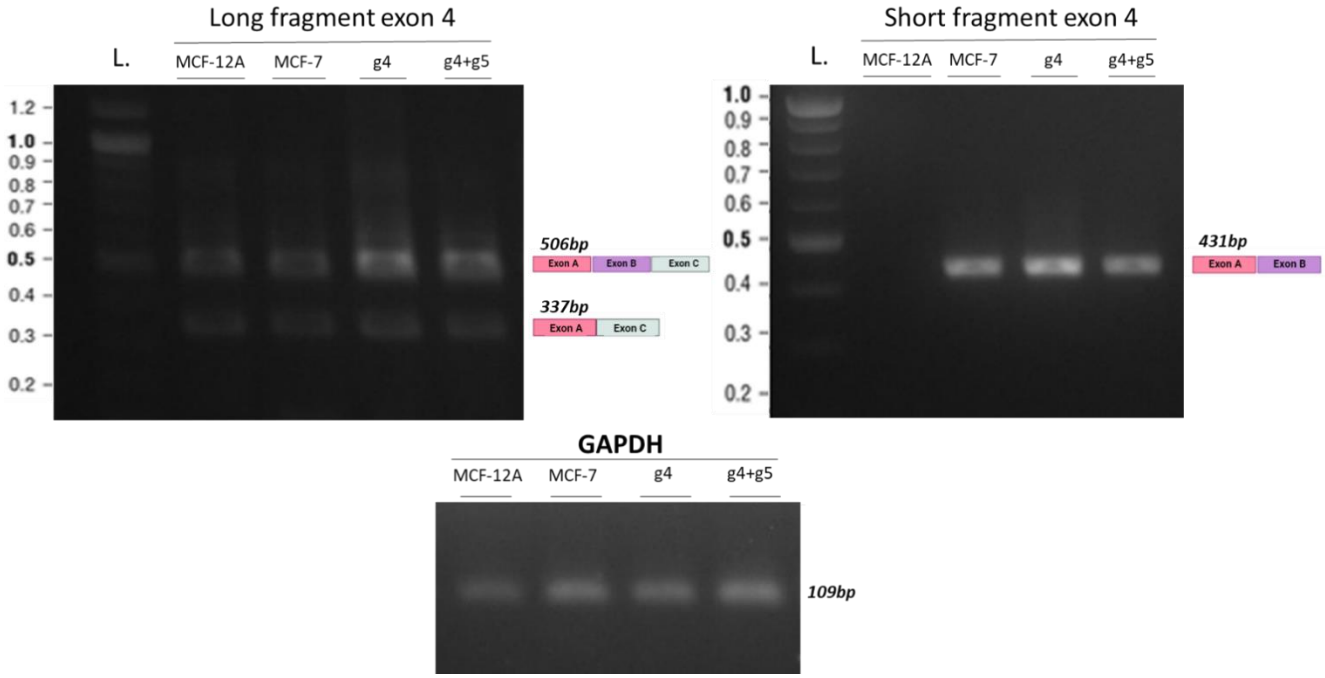

Figure 27. Gel Electrophoresis Analysis of PCR Products from MCF-12A, MCF-7 wt , MCF-7 transfected with gBlock4 and MCF-7 transfected with gBlock 4 and gBlock5. The figures include the long fragment product (left), the short fragment (right). The insensity of the bands were normalized to the insensity of GAPDH

The RNA samples were isolated only from the nucleus in order to enhance the signal, taking into consideration that lncRNAs are expressed at low levels compared to protein-coding mRNAs and also a large number of lncRNAs are localized in the nucleus to participate in regulatory functions, which has also been underlined for MCF-7 (19,21). For exon 4, the results indicated that, concerning the long fragment product (FW - RV long primer set), the upper band - representing a product including exon 3, along with both exon 4 and exon 5, exhibited increased intensity in MCF-7 cells transfected with gBlock 4, as well as in the case of double transfection with gBlock 4 and gBlock 5. In the case of the short fragment product (FW - RV short primer set), which corresponds only to exon 3 and exon 4, a slightly

higher band intensity is observed in MCF-7 cells transfected with gBlock 4 compared to other samples. However, no results were obtained for MCF-12A, possibly due to technical issues with the reaction. For all samples, the band intensities for both long and short products were compared to the intensity of GAPDH, used as a control (Figure 27). These findings suggest that the implementation of the CASFx tool might have the potential to enhance the inclusion of exon 4.

5. Quantitative PCR Analysis and RNA-seq data analysis of *MYC* gene expression

As mentioned earlier in the introduction, PVT1 exhibits a strong correlation with *MYC* expression, and various hypotheses suggest that it can impact *MYC* expression through multiple mechanisms, including interactions with miRNAs or its influence on the MYC promoter, thereby regulating *MYC* expression (90,91,94,95). Consequently, I wanted to investigate whether the attempted splicing modulation of PVT1 could influence *MYC* expression at the gene level. To achieve this, I conducted a qPCR experiment using GAPDH for the normalization. I measured MYC levels in cDNA extracted from nuclear RNA samples obtained from MCF-12A, unmodified MCF-7 (MCF-7 wt), MCF-7 cells transfected with gBlock4 or gBlock10, and cells double transfected with gBlock4 + gBlock5 or gBlock10 + gBlock11. The results of this analysis are presented below, with a comparison to *MYC* expression in MCF-7 wt (Figure 28).

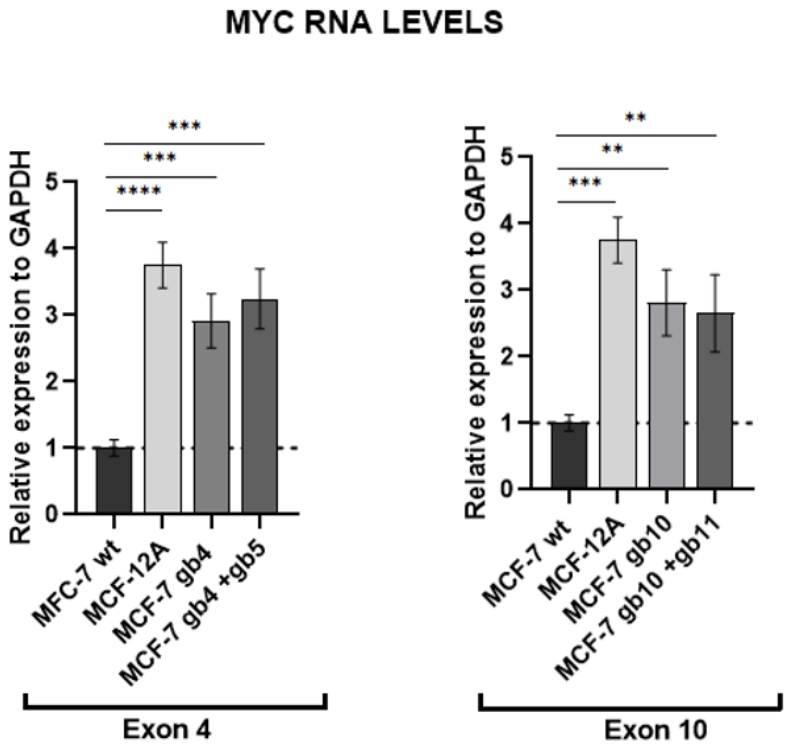

Figure 28. Quantitative PCR results for MYC expression. Left is the relative expression of *MYC* for MCF-7 and MCF-12A wild type cells, as well as after the enhancement of splicing at exon 4 in MCF-7 transfected with gBlock4 (MCF-7 gb4) and gBlock 4 and gBlock 5 (MCF-7 gb4 + gb5). On the right is relative expression of *MYC* for MCF-7 and MCF-12A wild type cells, as well as after the enhancement of splicing at exon 10 in MCF-7 transfected with gBlock10 (MCF-7 gb10) and gBlock 10 and gBlock 11 (MCF-7 gb10 + gb11). GAPDH expression levels were sued for the normalization of these results.

As depicted in the figure, *MYC* expression is significantly higher in MCF-12A compared to wild type MCF-7. Interestingly, in transfected MCF-7 cells, both for exon 4 and exon 10 an increase in RNA levels of *MYC* is clearly observed. Notably, the transfected MCF-7 cells appear to exhibit a gene expression profile more like MCF-12A than wild type MCF-7. To confirm our observations, we examined the *MYC* expression pattern between MCF-10A, which are alike MCF-12A and thus used again as a proxy, and MCF-7 using the published RNA-seq data. This analysis relied on read counts, that were obtained using *bedtool coverage* and were normalized to the total number of uniquely mapped reads to ensure consistency across libraries. The results from this analysis indicated that *MYC* expression is higher in MCF-10A compared to MCF-7, however without a statistical significance (Figure 29).

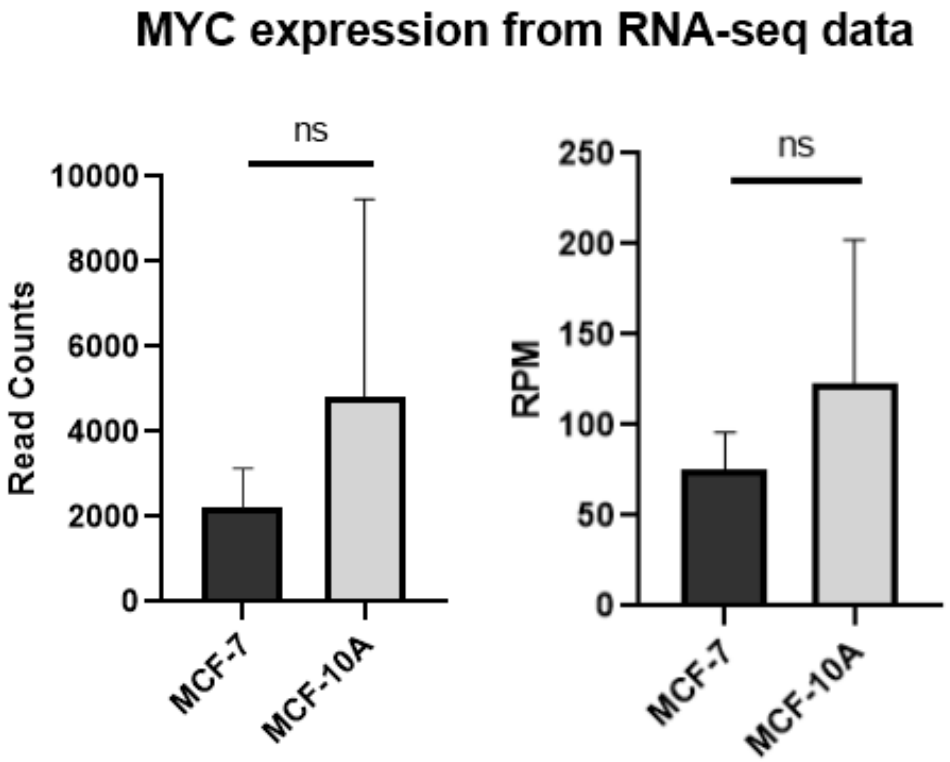

Figure 29. Representation of the expression levels of *MYC*, between MCF-7 and MCF-10A cell lines, acquired from published RNA-seq data. On the left is the read counts of *MYC* expression from 3 replicated for each cell line and on the right is the normalization of read counts to uniquely map reads per million (RPM).

Conclusions and Discussion

The study of Plasmacytoma Variant Translocation 1 (PVT1) holds significant importance in the field of cancer research and beyond. As a long non-coding RNA (lncRNA), PVT1 represents a fascinating subject for RNA biology studies, illuminating the roles of these molecules in cellular processes. *PVT1* is frequently overexpressed in various cancer types, making it a key player in cancer development. Investigating its role in different cancers can provide valuable insights into the mechanisms underlying tumorigenesis, potentially leading to novel therapeutic approaches. Moreover, *PVT1* expression levels have been proposed as diagnostic and prognostic markers, offering promise for more accurate cancer diagnosis and prognosis. Furthermore, understanding the normal biological functions of PVT1, which is expressed in non-cancerous tissues as well, contributes to our comprehensive understanding of gene regulation.

Building upon initial investigations that suggested a correlation between splicing efficiency of lncRNAs with their chromatin dissociation, as well as the categorization of PVT1 as a slowly released from chromatin lncRNA, that also contains introns with poor splicing efficiency, our objective is to explore the impact of PVT1 splicing enhancement in its localization dynamics related to chromatin and subsequently, the potential regulation of nearby genes, such us MYC. For this reason, CRISPR Artificial Splicing Factor (CASFx), which is a novel and promising tool is utilized to enhance the splicing of specific introns that are poorly spliced in PVT1. The MCF-7 cell line, that represents Luminal A breast cancer, was chosen as an experimental model for this study, as PVT1 is highly associated with breast cancer, and the results can be directly compared with MCF-10A and MCF-12A that are normal cell lines of breast tissue.

I successfully constructed a plasmid containing the catalytic fused protein of dCasRx and RBFOX1, along with introduced restriction sites for the insertion of any polycistronic gRNA sequence, previously mentioned as gBlock. As was indicated by the initial bioinformatic analysis of MCF-7 and MCF-10A RNA-seq data, there is an increased expression of PVT1 intronic region in MCF-7 compared to MCF-10A, which indicates decreased splicing efficiency. Thereby, after transfecting the constructed plasmid, and validated the expression of the splicing activator protein RBFOX1-dCasRX, I had to evaluate the functionality of the constructed CASFx plasmid.

The PCR results for the long fragment of exon 4 suggest that single transfection with gBlock 4 may have influenced exon retention, as evidenced by the slight increase of the intensity of the expected band. However, for the short fragment, a band was not obtained in the case of MCF-12A, possibly due to a technical issue during the reaction. Nevertheless, the results obtained in the case of gBlock 4 regarding the short fragment PCR experiment, support the observation in the long fragment case. Regarding gBlock 10, I was unable to obtain consistent results in my PCR profiles, which could be attributed problems with primer design or potential off-target effects of the tool. While the PCR results indicate the possibility of alternative splicing following the use of the tool and appear promising, these are only preliminary and further experiments are necessary to elucidate the tool's potential function and cross out the possibility of non-specific targeting. Specifically, Northern blot and 3'RACE experiments are planned to be conducted to assess the RNA transcripts produced after CASFx plasmid transfection and to compare them with those from the MCF-7 and MCF-12A cell lines.

Regarding MYC expression in MCF-7, MCF-12A, and transfected cells, previous studies suggested MYC overexpression in breast cancer (106), especially under *PVT1* overexpression, which stabilizes MYC expression (105). However, our experimental results did not align with this expectation. MYC expression can be influenced by various regulatory mechanisms, including transcription factors, chromatin accessibility, and epigenetic modifications. Variations in these regulatory elements between MCF-12A and MCF-7 may lead to differences in MYC expression levels. It's plausible that the interaction of MYC with other genes, proteins, and non-coding RNAs plays a role in these discrepancies. A possible hypothesis is that by affecting PVT1 splicing, it might dissociate from chromatin and potentially participate in complexes influencing MYC expression, probably via miRNA interactions, as observed in other genes (88,89). Another explanation is based on a previous study regarding the PVT1b isoform. With their research, Olivero et al. highlighted that PVT1b, which is activated by p53, has the capacity to suppress the transcription activation of *MYC* (85). In our study, by enhancing PVT1 splicing using the CASFx tool, we may potentially impact the stability or intracellular localization of PVT1b. This alteration could lead to a reduction in the inhibition of MYC, consequently resulting in increased levels of MYC expression, as indicated by the findings of qPCR presented in this thesis. It is essential to emphasize that these findings are still preliminary, and therefore, no definitive conclusions can be drawn at this point. Our initial focus should involve the exclusion of the possibility of non-specific targeting by the CASFx tools that are being designed and used, as such targeting could also influence MYC expression due to the close proximity of these genetic elements.

Looking ahead, a critical aspect that warrants further investigation is the tool's effect on chromatin association. Evaluating PVT1 chromatin binding post-plasmid transfection can be accomplished through subcellular fractionation techniques, allowing us to measure PVT1 levels in chromatin and nucleoplasmic fractions. Additionally, conducting FISH experiments will enable us to pinpoint the precise subnuclear location of PVT1 and determine whether CASFx has any impact on its subcellular positioning. Furthermore, considering PVT1 characterization as a competing endogenous RNA (ceRNA), investigating potential interactions in cancer after splicing enhancement could be insightful. Changes in splicing might affect the presence of miRNA recognition elements (MREs) in the transcript. This approach could be particularly relevant for the miR200 family, associated with breast cancer and highly interactive with PVT1, as indicated by computational analysis (72,87,88). Experimentation to confirm interactions between PVT1 and miR200 could involve RNA pull-down assays. Regarding the RNA pull-down assay, synthesized biotin-labeled PVT1 can be used in the cells and by using streptavidin-coated magnetic beads, the pull down the biotinylated RNA complexes can be achieved.

In summary, these preliminary findings provide a foundation for further research and optimization. I successfully constructed a plasmid housing the RBFOX1-dCasRx protein, incorporating a U6 promoter sequence, as well as the necessary gRNA insertion sites downstream of the U6 promoter. However, to enhance specificity and minimize off-target splicing, improvements in guide RNA design are necessary, including exploration of additional sites beyond the ones already designed. The regulatory mechanisms of PVT1 hold great promise, particularly in the context of its involvement in cancer

development, as indicated by previous studies. Further investigations are essential to unravel the precise molecular pathways in which PVT1 participates, shedding light on its role in cancer development.

This knowledge could pave the way for innovative breast cancer therapies, addressing a critical need given the resistance observed in many cases to existing treatments. Breast cancer remains a leading cause of cancer-related deaths among women. While current treatment options encompass hormone therapy, endocrine therapy, immunotherapy, chemotherapy, radiotherapy, and surgery, novel approaches to therapeutic splicing modulation, such as small molecules, Splice Switching Oligonucleotides (SSOs), and CRISPR-Cas9 gene editing, are under focused research (49). This innovative approach holds the potential to revolutionize breast cancer treatment and detection strategies, offering new avenues for combating this widespread disease.

Appendix

1. Table with the bioinformatic analysis

Chromosome	Start position	End position	Gene ID	p-value
chr8	127803147	127803151	*PVT1*-216;	0.0109786757893598
chr8	127816825	127816829	*PVT1*-218;	0.0302420655251375
chr8	127818897	127818901	*PVT1*-248;	0.0370864712135154
chr8	127819906	127819910	*PVT1*-248;	0.00867936610286304
chr8	127932462	127932466	*PVT1*-201;	0.0141714244885155
chr8	127939505	127939509	*PVT1*-203;	0.0142737367495261
chr8	127994880	127994884	*PVT1*-253;	0.0337892897463921
chr8	127996558	127996562	*PVT1*-204;	0.0177883860793165
chr8	127998197	127998201	*PVT1*-214;	0.0100010513052427
chr8	128009587	128009591	*PVT1*-214;	0.00912754752884235
chr8	128016632	128016636	*PVT1*-293;	0.00649465061050628
chr8	128031956	128031960	*PVT1*-226;	0.0234992114231621
chr8	128038914	128038918	*PVT1*-347;	0.0331246934841671
chr8	128048445	128048449	*PVT1*-202;	0.0134341846072167
chr8	128082750	128082754	*PVT1*-214;	0.0307577870951947
chr8	128090245	128090249	*PVT1*-251;	0.0116804993590901

2. MCF-7 transfection efficiency optimization

To enhance transfection efficiency in the MCF-7 cell line, various treatments were implemented and their corresponding outcomes were examined. According to Søndergaard JN et al., co-transfecting the cells with a smaller plasmid is recommended to improve transfection efficiency, as the smaller plasmid facilitates the integration of a larger one (108). For this purpose, pBluescript II SK(+) (kindly provided by Mr. Kostas Theodorakis) was employed for co-transfection, using a ratio of 1:1 between the larger and smaller plasmids. Furthermore, the impact of cell starvation was evaluated by subjecting the cells to different concentrations of FBS in DMEM and varying treatment durations. Serum starvation is known to enhance cell membrane permeability and promote the uptake and delivery of transfection reagents or nucleic acids into cells, thereby improving transfection efficiency (109).

In a 48-well plate, MCF-7 cells were cultured in70% confluency. In case A, three different conditions were set up: cells cultured in 5% FBS for 16 hours, 1% FBS for 16 hours, and 0.1% FBS for 16 hours. Each condition was further divided into two extra conditions, cells transfected with either 0.5µg GFP or a combination of 0.25µg GFP and 0.25µg pBluescript, in a 1:1 ratio, as suggested by Søndergaard JN et al (108). Following the 16-hour incubation at 37°C, the medium for all conditions was changed to 5% FBS, and the cells were transfected with JetPei. In case B, similar conditions were replicated, but with a

shorter starvation time. Initially, the cells were cultured in 5% FBS-DMEM, and after 16 hours, a 4-hour starvation treatment was initiated using either 5% FBS as a control, 1% FBS, or 0.1% FBS. For each condition, similarly to case A, it was evaluated the effect of pBluescript II SK(+) in the efficiency, with the cells being transfected with only GFP or GFP in combination with pBluescript II SK(+). The medium was changed to 5% FBS before transfection with JetPei. In case C, no starvation treatment was applied but allowed the cells to rest for 3 hours after splitting, before immediate transfection with only GFP plasmid or in combination with pBluescript II SK(+). After 24 hours of transfection, the medium was replaced to avoid any toxic effects, and the cells were observed under UV light to measure transfection efficiency.

The results indicated that pBluescript II SK(+) possibly hindered the penetration of GFP plasmid into the cells as there was no signal observed after UV exposure in all tested conditions. Moreover, in the case of 16 hours of starvation, the highest efficiency was observed in the 5% FBS-DMEM treatment, while in the 0.1% FBS-DMEM treatment, most cells died. For the 4-hour starvation, the best efficiency was similarly observed in the 5% FBS-DMEM condition, with no significant difference in the other two cases. However, the 3-hour incubation after splitting the cells also showed promising results in transfection efficiency.

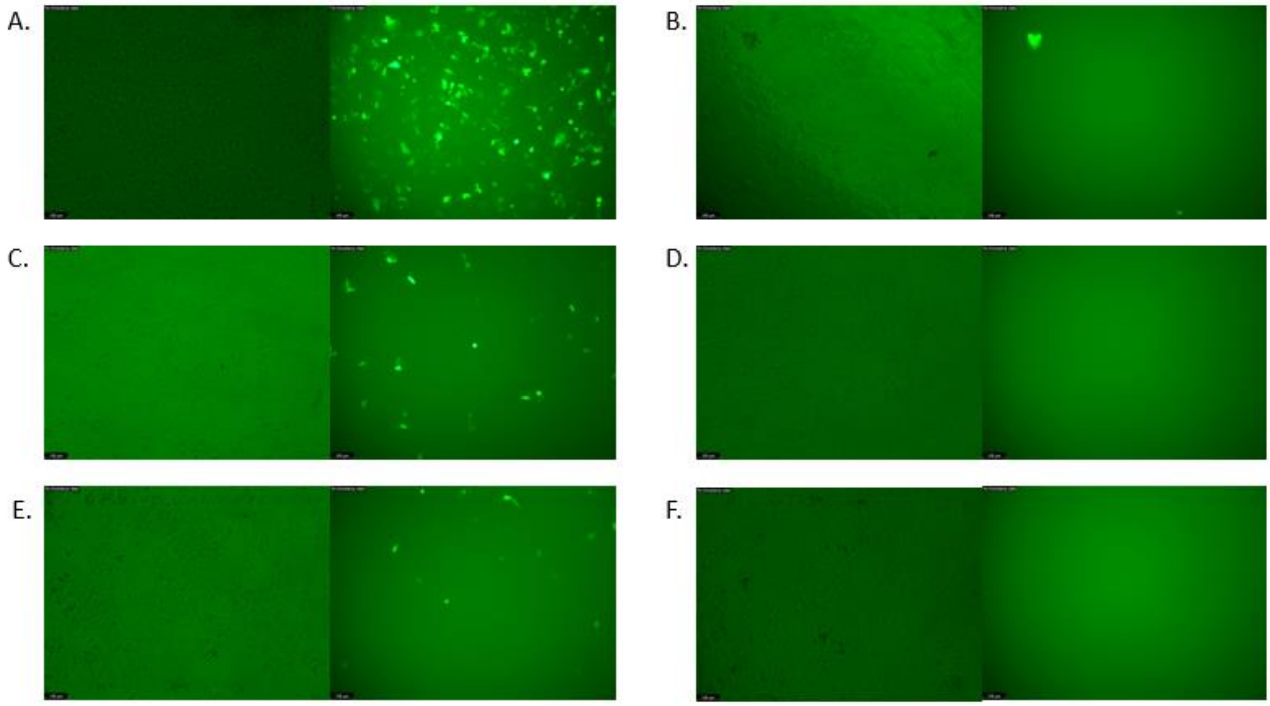

Figure 30. MCF-7 cell transfection efficiency with GFP plasmid was assessed under various conditions: A. Cells were cultured with 5% FBS for 16 hours after splitting, followed by transfection with the GFP plasmid (control), B. Cells were cultured with 5% FBS for 16 hours after splitting, followed by transfection with both the GFP and pBluescript SKIIa+ plasmids in a 1:1 ratio (control), C. Cells were cultured with 1% FBS for 16 hours after splitting, followed by transfection with the GFP plasmid, D. Cells were cultured with 1% FBS for 16 hours after splitting, followed by transfection with both the GFP and pBluescript SKIIa+ plasmids in a 1:1 ratio, E. Cells were cultured with 0.1% FBS for 16 hours after splitting, followed by transfection with the GFP plasmid, F. Cells were cultured with 0.1% FBS for 16 hours after splitting, followed by transfection with both the GFP and pBluescript SKIIa+ plasmids in a 1:1 ratio.

The conditions with the best-evaluated efficiencies, namely the overnight treatment with 5% FBS-DMEM and the 3-hour incubation in 5% FBS-DMEM after splitting, were further tested after a heat-

shock, as suggested by Piper et al. In the case of the 3-hour 5% FBS-DMEM treatment, the transfection efficiency slightly increased, whereas for the 16-hour 5% FBS-DMEM incubation, there was no significant improvement.

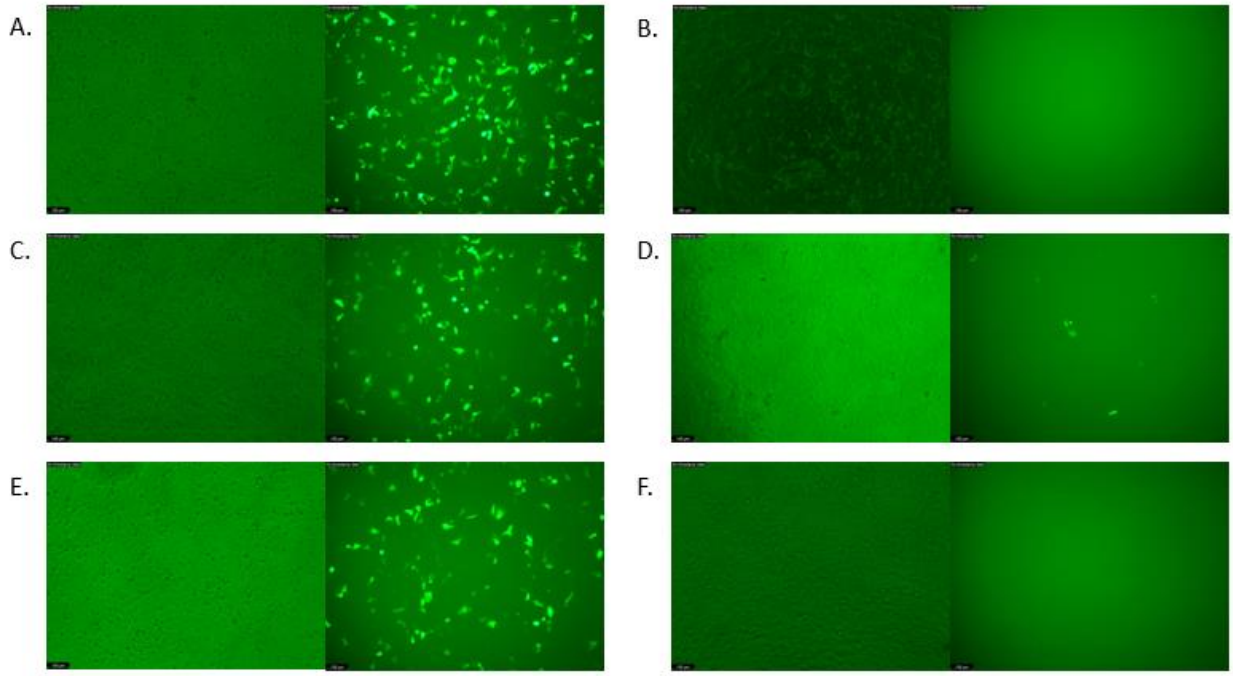

Figure 31. MCF-7 cell transfection efficiency with GFP plasmid was assessed under various conditions: A. Cells were initially cultured with 5% FBS for 16 hours after splitting, followed by an additional 4-hour incubation in fresh 5% FBS-DMED before transfection with the GFP plasmid, B. Cells were initially cultured with 5% FBS for 16 hours after splitting, followed by an additional 4-hour incubation in 5% FBS medium. Subsequently, they were transfected with both the GFP and pBluescript SKIIa+ plasmids in a 1:1 ratio, C. Cells were initially cultured with 5% FBS for 16 hours after splitting, followed by a 4-hour period of starvation with 1% FBS before transfection with the GFP plasmid, D. Cells were initially cultured with 5% FBS for 16 hours after splitting, followed by a 4-hour period of starvation with 1% FBS before transfection with both the GFP and pBluescript SKIIa+ plasmids in a 1:1 ratio, E. Cells were initially cultured with 5% FBS for 16 hours after splitting, followed by a 4-hour period of starvation with 0.1% FBS before transfection with the GFP plasmid, F. Cells were initially cultured with 5% FBS for 16 hours after splitting, followed by a 4-hour period of starvation with 0.1% FBS before transfection with both the GFP and pBluescript SKIIa+ plasmids in a 1:1 ratio.

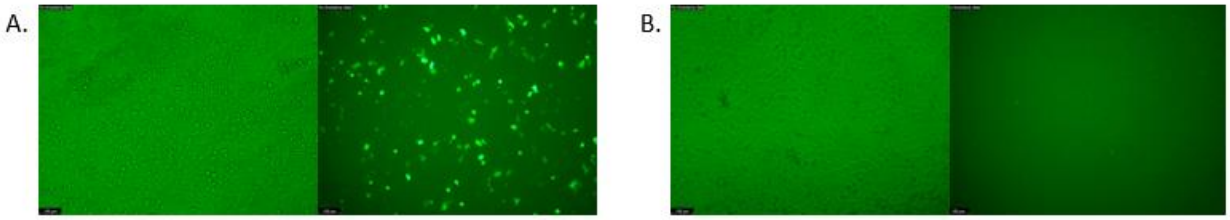

Figure 32. Transfection efficiency of MCF-7 with GFP plasmid A. Cells were allowed to settle for 4 hours after splitting in medium containing 5% FBS, after which they were transfected with the GFP plasmid, B. Cells were allowed to settle for 4 hours after splitting in medium containing 5% FBS, followed by an additional 4-hour incubation. After this, they were transfected with both the GFP and pBluescript SKIIa+ plasmids in a 1:1 ratio.

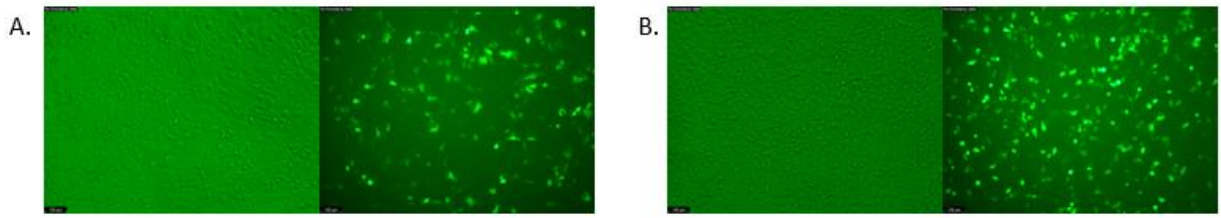

Figure 33. Transfection efficiency of MCF-7 with GFP plasmid. A. Cells were allowed to settle for 16 hours after splitting in medium containing 5% FBS. They were then transfected with the GFP plasmid and subjected to a heat shock treatment 4 hours after transfection, B. Cells were allowed to settle for 4 hours after splitting in medium containing 5% FBS. They were then transfected with both the GFP and pBluescript SKIIa+ plasmids in a 1:1 ratio and underwent a heat shock treatment 4 hours after transfection.

3. Validation of the correct incorporation of gBlock fragment in the CASFx vector.

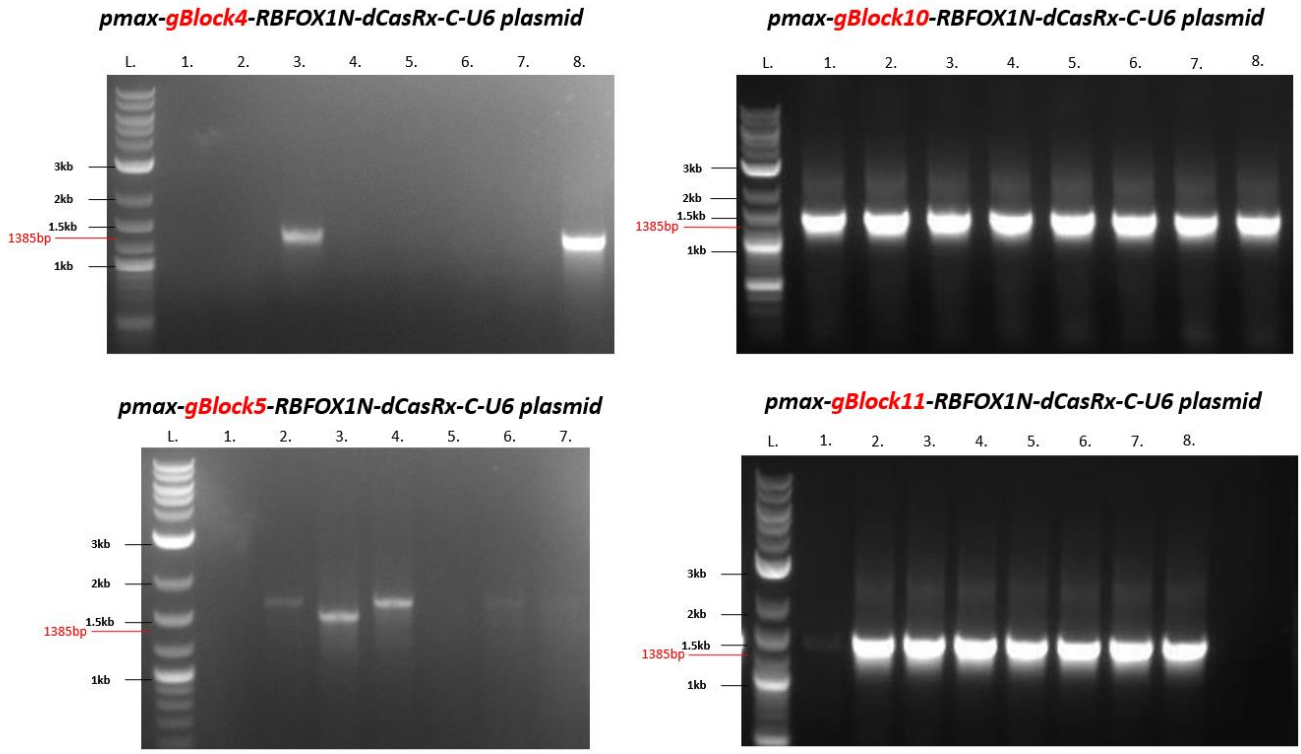

Figure 34. Figure of gel electrophoresis after colony PCR in 1% agarose gel, for the validation of the correct incorporation of gBlock 4, gBlock5, gBlock10 and gBlcok11 in the CASFx vector. Each lane represents one of the tested clones of Dh5a with the plasmid that includes both U6 sequence and gBlock.

References

1. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, et al. Landscape of transcription in human cells. Nature. 2012 Sep;489(7414):101–8.

2. Kopp F, Mendell JT. Functional Classification and Experimental Dissection of Long Noncoding RNAs. Cell. 2018 Jan 25;172(3):393–407.

3. Bhan A, Mandal SS. Long Noncoding RNAs: Emerging Stars in Gene Regulation, Epigenetics and Human Disease. ChemMedChem. 2014;9(9):1932–56.

4. Zhang P, Wu W, Chen Q, Chen M. Non-Coding RNAs and their Integrated Networks. J Integr Bioinforma. 2019 Jul 13;16(3):20190027.

5. Subramanian S. Competing endogenous RNAs (ceRNAs): new entrants to the intricacies of gene regulation. Front Genet [Internet]. 2014 [cited 2023 Aug 3];5. Available from: https://www.frontiersin.org/articles/10.3389/fgene.2014.00008

6. Yin Y, Shen X. Noncoding RNA-chromatin association: Functions and mechanisms. Fundam Res [Internet]. 2023 Apr 7 [cited 2023 Aug 22]; Available from: https://www.sciencedirect.com/science/article/pii/S2667325823000870

7. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature. 2009 Mar;458(7235):223–7.

8. Kazimierczyk M, Kasprowicz MK, Kasprzyk ME, Wrzesinski J. Human Long Noncoding RNA Interactome: Detection, Characterization and Function. Int J Mol Sci. 2020 Jan;21(3):1027.

9. Statello L, Guo CJ, Chen LL, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol. 2021 Feb;22(2):96–118.

10. Schlackow M, Nojima T, Gomes T, Dhir A, Carmo-Fonseca M, Proudfoot NJ. Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. Mol Cell. 2017 Jan 5;65(1):25–38.

11. Nadhan R, Isidoro C, Song YS, Dhanasekaran DN. Signaling by LncRNAs: Structure, Cellular Homeostasis, and Disease Pathology. Cells. 2022 Jan;11(16):2517.

12. Hermans-Beijnsberger S, van Bilsen M, Schroen B. Long non-coding RNAs in the failing heart and vasculature. Non-Coding RNA Res. 2018 Sep 1;3(3):118–30.

13. Wang KC, Chang HY. Molecular Mechanisms of Long Noncoding RNAs. Mol Cell. 2011 Sep 16;43(6):904–14.

14. Fang Y, Fullwood MJ. Roles, Functions, and Mechanisms of Long Non-coding RNAs in Cancer. Genomics Proteomics Bioinformatics. 2016 Feb;14(1):42–54.

15. Ahmad P, Bensaoud C, Mekki I, Rehman MU, Kotsyfakis M. Long Non-Coding RNAs and Their Potential Roles in the Vector–Host–Pathogen Triad. Life. 2021 Jan;11(1):56.

16. Clark MB, Johnston RL, Inostroza-Ponta M, Fox AH, Fortini E, Moscato P, et al. Genome-wide analysis of long noncoding RNA stability. Genome Res. 2012 May;22(5):885–98.

17. Wilusz JE, JnBaptiste CK, Lu LY, Kuhn CD, Joshua-Tor L, Sharp PA. A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails. Genes Dev. 2012 Nov 1;26(21):2392–407.

18. Kai Z, Yuting W, Yulin L, Jun L, Juanjuan H. An efficient simulated annealing algorithm for the RNA secondary structure prediction with Pseudoknots. BMC Genomics. 2019 Dec 27;20(13):979.

19. Sun Q, Hao Q, Prasanth KV. Nuclear long noncoding RNAs: key regulators of gene expression. Trends Genet TIG. 2018 Feb;34(2):142–57.

20. Cabili MN, Dunagin MC, McClanahan PD, Biaesch A, Padovan-Merhar O, Regev A, et al. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. Genome Biol. 2015 Jan 29;16(1):20.

21. Carlevaro-Fita J, Johnson R. Global Positioning System: Understanding Long Noncoding RNAs through Subcellular Localization. Mol Cell. 2019 Mar 7;73(5):869–83.

22. Sauvageau M. Diverging RNPs: Toward Understanding lncRNA-Protein Interactions and Functions. In: Oeffinger M, Zenklusen D, editors. The Biology of mRNA: Structure and Function [Internet]. Cham: Springer International Publishing; 2019 [cited 2023 Sep 21]. p. 285–312. (Advances in Experimental Medicine and Biology). Available from: https://doi.org/10.1007/978-3-030-31434-7_10

23. Maldonado R, Schwartz U, Silberhorn E, Längst G. Nucleosomes Stabilize ssRNA-dsDNA Triple Helices in Human Cells. Mol Cell. 2019 Mar 21;73(6):1243-1254.e6.

24. Ørom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, et al. Long Noncoding RNAs with Enhancer-like Function in Human Cells. Cell. 2010 Oct 1;143(1):46–58.

25. Yin QF, Yang L, Zhang Y, Xiang JF, Wu YW, Carmichael GG, et al. Long Noncoding RNAs with snoRNA Ends. Mol Cell. 2012 Oct 26;48(2):219–30.

26. Fernandes JCR, Acuña SM, Aoki JI, Floeter-Winter LM, Muxel SM. Long Non-Coding RNAs in the Regulation of Gene Expression: Physiology and Disease. Non-Coding RNA. 2019 Mar;5(1):17.

27. Derderian C, Orunmuyi AT, Olapade-Olaopa EO, Ogunwobi OO. PVT1 Signaling Is a Mediator of Cancer Progression. Front Oncol. 2019 Jun 12;9:502.

28. Gong C, Maquat LE. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. Nature. 2011 Feb;470(7333):284–8.

29. Chen Y, Li Z, Chen X, Zhang S. Long non-coding RNAs: From disease code to drug role. Acta Pharm Sin B. 2021 Feb 1;11(2):340–54.

30. Nishimoto Y, Nakagawa S, Hirose T, Okano HJ, Takao M, Shibata S, et al. The long non-coding RNA nuclear-enriched abundant transcript 1_2 induces paraspeckle formation in the motor neuron during the early phase of amyotrophic lateral sclerosis. Mol Brain. 2013 Jul 8;6(1):31.

31. Mus E, Hof PR, Tiedge H. Dendritic BC200 RNA in aging and in Alzheimer's disease. Proc Natl Acad Sci. 2007 Jun 19;104(25):10679–84.

32. Sparber P, Filatova A, Khantemirova M, Skoblov M. The role of long non-coding RNAs in the pathogenesis of hereditary diseases. BMC Med Genomics. 2019 Mar 13;12(2):42.

33. Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, et al. A large intergenic non-coding RNA induced by p53 mediates global gene repression in the p53 response. Cell. 2010 Aug 6;142(3):409–19.

34. Arun G, Diermeier SD, Spector DL. Therapeutic Targeting of Long Non-Coding RNAs in Cancer. Trends Mol Med. 2018 Mar 1;24(3):257–77.

35. Hannon GJ, Rossi JJ. Unlocking the potential of the human genome with RNA interference. Nature. 2004 Sep;431(7006):371–8.

36. Wittrup A, Lieberman J. Knocking down disease: a progress report on siRNA therapeutics. Nat Rev Genet. 2015 Sep;16(9):543–52.

37. Yoo BH, Bochkareva E, Bochkarev A, Mou TC, Gray DM. 2′-O-methyl-modified phosphorothioate antisense oligonucleotides have reduced non-specific effects in vitro. Nucleic Acids Res. 2004;32(6):2008–16.

38. Zhou T, Kim Y, MacLeod AR. Targeting Long Noncoding RNA with Antisense Oligonucleotide Technology as Cancer Therapeutics. In: Feng Y, Zhang L, editors. Long Non-Coding RNAs: Methods and Protocols [Internet]. New York, NY: Springer; 2016 [cited 2023 Sep 13]. p. 199–213. (Methods in Molecular Biology). Available from: https://doi.org/10.1007/978-1-4939-3378-5_16

39. Chavez M, Chen X, Finn PB, Qi LS. Advances in CRISPR therapeutics. Nat Rev Nephrol. 2023 Jan;19(1):9–22.

40. Winkle M, El-Daly SM, Fabbri M, Calin GA. Noncoding RNA therapeutics — challenges and potential solutions. Nat Rev Drug Discov. 2021 Aug;20(8):629–51.

41. Cyranoski D. CRISPR gene-editing tested in a person for the first time. Nature. 2016 Nov 1;539(7630):479–479.

42. Wilkinson ME, Charenton C, Nagai K. RNA Splicing by the Spliceosome. Annu Rev Biochem. 2020;89(1):359–88.

43. Wang Z, Burge CB. Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. RNA. 2008 May;14(5):802–13.

44. Park E, Pan Z, Zhang Z, Lin L, Xing Y. The Expanding Landscape of Alternative Splicing Variation in Human Populations. Am J Hum Genet. 2018 Jan 4;102(1):11–26.

45. Stanley RF, Abdel-Wahab O. Dysregulation and therapeutic targeting of RNA splicing in cancer. Nat Cancer. 2022 May;3(5):536–46.

46. Lai HC, Ho UY, James A, De Souza P, Roberts TL. RNA metabolism and links to inflammatory regulation and disease. Cell Mol Life Sci. 2021 Dec 31;79(1):21.

47. Tazi J, Bakkour N, Stamm S. Alternative splicing and disease. Biochim Biophys Acta BBA - Mol Basis Dis. 2009 Jan 1;1792(1):14–26.

48. Suñé-Pou M, Limeres MJ, Moreno-Castro C, Hernández-Munain C, Suñé-Negre JM, Cuestas ML, et al. Innovative Therapeutic and Delivery Approaches Using Nanotechnology to Correct Splicing Defects Underlying Disease. Front Genet [Internet]. 2020 [cited 2023 Sep 27];11. Available from: https://www.frontiersin.org/articles/10.3389/fgene.2020.00731

49. Zhang Y, Qian J, Gu C, Yang Y. Alternative splicing and cancer: a systematic review. Signal Transduct Target Ther. 2021 Feb 24;6(1):1–14.

50. Biamonti G, Catillo M, Pignataro D, Montecucco A, Ghigna C. The alternative splicing side of cancer. Semin Cell Dev Biol. 2014 Aug 1;32:30–6.

51. Ouyang J, Zhong Y, Zhang Y, Yang L, Wu P, Hou X, et al. Long non-coding RNAs are involved in alternative splicing and promote cancer progression. Br J Cancer. 2022 May;126(8):1113–24.

52. Liu S, Cheng C. Alternative RNA splicing and cancer. Wiley Interdiscip Rev RNA. 2013;4(5):547–66.

53. Singh B, Eyras E. The role of alternative splicing in cancer. Transcription. 2016 Dec 22;8(2):91–8.

54. Supek F, Miñana B, Valcárcel J, Gabaldón T, Lehner B. Synonymous Mutations Frequently Act as Driver Mutations in Human Cancers. Cell. 2014 Mar 13;156(6):1324–35.

55. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012 Aug 17;337(6096):816–21.

56. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nat Protoc. 2013 Nov;8(11):2281–308.

57. Du M, Jillette N, Zhu JJ, Li S, Cheng AW. CRISPR artificial splicing factors. Nat Commun. 2020 Jun 12;11(1):2973.

58. Cui M, You L, Ren X, Zhao W, Liao Q, Zhao Y. Long non-coding RNA PVT1 and cancer. Biochem Biophys Res Commun. 2016 Feb 26;471(1):10–4.

59. Jin K, Wang S, Zhang Y, Xia M, Mo Y, Li X, et al. Long non-coding RNA PVT1 interacts with MYC and its downstream molecules to synergistically promote tumorigenesis. Cell Mol Life Sci. 2019 Nov 1;76(21):4275–89.

60. Huppi K, Volfovsky N, Runfola T, Jones TL, Mackiewicz M, Martin SE, et al. The identification of microRNAs in a genomically unstable region of human chromosome 8q24. Mol Cancer Res. 2008;6(2):212–21.

61. Li H, Chen S, Liu J, Guo X, Xiang X, Dong T, et al. Long non-coding RNA PVT1-5 promotes cell proliferation by regulating miR-126/SLC7A5 axis in lung cancer. Biochem Biophys Res Commun. 2018 Jan 15;495(3):2350–5.

62. Huang T, Liu HW, Chen JQ, Wang SH, Hao LQ, Liu M, et al. The long noncoding RNA PVT1 functions as a competing endogenous RNA by sponging miR-186 in gastric cancer. Biomed Pharmacother. 2017 Apr 1;88:302–8.

63. Ping G, Xiong W, Zhang L, Li Y, Zhang Y, Zhao Y. Silencing long noncoding RNA PVT1 inhibits tumorigenesis and cisplatin resistance of colorectal cancer. Am J Transl Res. 2018 Jan 15;10(1):138–49.

64. Zhang Y, Wen D yue, Zhang R, Huang J cheng, Lin P, Ren FH, et al. A Preliminary Investigation of PVT1 on the Effect and Mechanisms of Hepatocellular Carcinoma: Evidence from Clinical Data, a Meta-Analysis of 840 Cases, and In Vivo Validation. Cell Physiol Biochem. 2018 Jul 5;47(6):2216–32.

65. Yang R, Shao T, Long M, Shi Y, Liu Q, Yang L, et al. Long noncoding RNA PVT1 promotes tumor growth and predicts poor prognosis in patients with diffuse large B-cell lymphoma. Cancer Commun. 2020 Jul 29;40(10):551–5.

66. Chen L, Ma D, Li Y, Li X, Zhao L, Zhang J, et al. Effect of long non-coding RNA PVT1 on cell proliferation and migration in melanoma. Int J Mol Med. 2018 Mar;41(3):1275–82.

67. You L, Wang H, Yang G, Zhao F, Zhang J, Liu Z, et al. Gemcitabine exhibits a suppressive effect on pancreatic cancer cell growth by regulating processing of PVT1 to miR1207. Mol Oncol. 2018 Dec;12(12):2147–64.

68. Yan C, Chen Y, Kong W, Fu L, Liu Y, Yao Q, et al. PVT1-derived miR-1207-5p promotes breast cancer cell growth by targeting STAT6. Cancer Sci. 2017 May;108(5):868–76.

69. Gao YL, Zhao ZS, Zhang MY, Han LJ, Dong YJ, Xu B. Long Noncoding RNA PVT1 Facilitates Cervical Cancer Progression via Negative Regulating of miR-424. Oncol Res. 2017 Sep 21;25(8):1391–8.

70. Riquelme E, Suraokar MB, Rodriguez J, Mino B, Lin HY, Rice DC, et al. Frequent co-amplification and co-operation between C-MYC and PVT1 oncogenes promote malignant pleural mesothelioma. J Thorac Oncol Off Publ Int Assoc Study Lung Cancer. 2014 Jul;9(7):998–1007.

71. Yang M, Zhang L, Wang X, Zhou Y, Wu S. Down-regulation of miR-203a by lncRNA PVT1 in multiple myeloma promotes cell proliferation. Arch Med Sci AMS. 2018 Oct;14(6):1333–9.

72. Yang T, Zhou H, Liu P, Yan L, Yao W, Chen K, et al. lncRNA PVT1 and its splicing variant function as competing endogenous RNA to regulate clear cell renal cell carcinoma progression. Oncotarget. 2017 Jul 31;8(49):85353–67.

73. Tian Z, Cao S, Li C, Xu M, Wei H, Yang H, et al. LncRNA PVT1 regulates growth, migration, and invasion of bladder cancer by miR-31/ CDK1. J Cell Physiol. 2019;234(4):4799–811.

74. Chang Z, Cui J, Song Y. Long noncoding RNA PVT1 promotes EMT via mediating microRNA-186 targeting of Twist1 in prostate cancer. Gene. 2018 May 15;654:36–42.

75. Xue W, Chen J, Liu X, Gong W, Zheng J, Guo X, et al. PVT1 regulates the malignant behaviors of human glioma cells by targeting miR-190a-5p and miR-488-3p. Biochim Biophys Acta BBA - Mol Basis Dis. 2018 May 1;1864(5, Part A):1783–94.

76. Zhou Q, Chen F, Zhao J, Li B, Liang Y, Pan W, et al. Long non-coding RNA PVT1 promotes osteosarcoma development by acting as a molecular sponge to regulate miR-195. Oncotarget. 2016 Nov 2;7(50):82620–33.

77. Lin QY, Qi QL, Hou S, Chen Z, Zhang L, Zhao HG, et al. LncRNA PVT1 Acts as a Tumor Promoter in Thyroid Cancer and Promotes Tumor Progression by Mediating miR-423-5p-PAK3. Cancer Manag Res. 2020 Dec 30;12:13403–13.

78. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012 Jun;486(7403):346–52.

79. Lu D, Luo P, Wang Q, Ye Y, Wang B. lncRNA PVT1 in cancer: A review and meta-analysis. Clin Chim Acta. 2017 Nov 1;474:1–7.

80. Kong R, Zhang E bao, Yin D dan, You L hui, Xu T peng, Chen W ming, et al. Long noncoding RNA PVT1 indicates a poor prognosis of gastric cancer and promotes cell proliferation through epigenetically regulating p15 and p16. Mol Cancer. 2015 Apr 12;14:82.

81. Jung H, Lee D, Lee J, Park D, Kim YJ, Park WY, et al. Intron retention is a widespread mechanism of tumor-suppressor inactivation. Nat Genet. 2015 Nov;47(11):1242–8.

82. Kahles A, Lehmann KV, Toussaint NC, Hüser M, Stark SG, Sachsenberg T, et al. Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. Cancer Cell. 2018 Aug 13;34(2):211-224.e6.

83. Ghetti M, Vannini I, Storlazzi CT, Martinelli G, Simonetti G. Linear and circular PVT1 in hematological malignancies and immune response: two faces of the same coin. Mol Cancer. 2020 Mar 30;19(1):69.

84. Pal G, Huaman J, Levine F, Orunmuyi A, Olapade-Olaopa EO, Onagoruwa OT, et al. Long Noncoding RNA from PVT1 Exon 9 Is Overexpressed in Prostate Cancer and Induces Malignant Transformation and Castration Resistance in Prostate Epithelial Cells. Genes. 2019 Dec;10(12):964.

85. Olivero C, Martínez-Terroba E, Zimmer J, Liao C, Tesfaye E, Hooshdaran N, et al. p53 activates the long noncoding RNA Pvt1b to inhibit Myc and suppress tumorigenesis. Mol Cell. 2020 Feb 20;77(4):761-774.e8.

86. Wu F, Zhu Y, Zhou C, Gui W, Li H, Lin X. Regulation mechanism and pathogenic role of lncRNA plasmacytoma variant translocation 1 (PVT1) in human diseases. Genes Dis. 2023 May 1;10(3):901–14.

87. Paci P, Colombo T, Farina L. Computational analysis identifies a sponge interaction network between long non-coding RNAs and messenger RNAs in human breast cancer. BMC Syst Biol. 2014 Jul 17;8(1):83.

88. Conte F, Fiscon G, Chiara M, Colombo T, Farina L, Paci P. Role of the long non-coding RNA PVT1 in the dysregulation of the ceRNA-ceRNA network in human breast cancer. PLOS ONE. 2017;12(2):e0171661.

89. Wang H, Huang Y, Yang Y. LncRNA PVT1 Regulates TRPS1 Expression in Breast Cancer by Sponging miR-543. Cancer Manag Res. 2020 Sep 4;12:7993–8004.

90. Tseng YY, Bagchi A. The PVT1-MYC duet in cancer. Mol Cell Oncol. 2015 Feb 11;2(2):e974467.

91. Onagoruwa OT, Pal G, Ochu C, Ogunwobi OO. Oncogenic Role of PVT1 and Therapeutic Implications. Front Oncol. 2020 Feb 4;10:17.

92. Dhanasekaran R, Deutzmann A, Mahauad-Fernandez WD, Hansen AS, Gouw AM, Felsher DW. The MYC oncogene — the grand orchestrator of cancer growth and immune evasion. Nat Rev Clin Oncol. 2022 Jan;19(1):23–36.

93. Kalkat M, De Melo J, Hickman KA, Lourenco C, Redel C, Resetca D, et al. MYC Deregulation in Primary Human Cancers. Genes. 2017 Jun;8(6):151.

94. Sarver AL, Murray CD, Temiz NA, Tseng YY, Bagchi A. MYC and PVT1 synergize to regulate RSPO1 levels in breast cancer. Cell Cycle. 2016 Feb 18;15(7):881–5.

95. Cho SW, Xu J, Sun R, Mumbach MR, Carter AC, Chen YG, et al. Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. Cell. 2018 May 31;173(6):1398-1412.e22.

96. Baljon KJ, Ramaiah P, Saleh EAM, Al-dolaimy F, Al-Dami FH, Gandla K, et al. LncRNA PVT1: as a therapeutic target for breast cancer. Pathol - Res Pract. 2023 Aug 1;248:154675.

97. Smolarz B, Nowak AZ, Romanowicz H. Breast Cancer—Epidemiology, Classification, Pathogenesis and Treatment (Review of Literature). Cancers. 2022 Jan;14(10):2569.

98. Onitilo AA, Engel JM, Greenlee RT, Mukesh BN. Breast Cancer Subtypes Based on ER/PR and Her2 Expression: Comparison of Clinicopathologic Features and Survival. Clin Med Res. 2009 Jun;7(1–2):4–13.

99. Ntini E, Louloupi A, Liz J, Muino JM, Marsico A, Ørom UAV. Long ncRNA A-ROD activates its target gene DKK1 at its release from chromatin. Nat Commun. 2018 Apr 24;9:1636.

100. Ntini E, Budach S, Ørom UAV, Marsico A. Predictive modeling of long non-coding RNA chromatin (dis-)association [Internet]. bioRxiv; 2020 [cited 2023 Oct 6]. p. 2020.12.15.422063. Available from: https://www.biorxiv.org/content/10.1101/2020.12.15.422063v3

101. Leclair NK, Brugiolo M, Urbanski L, Lawson SC, Thakar K, Yurieva M, et al. POISON EXON SPLICING REGULATES A COORDINATED NETWORK OF SR PROTEIN EXPRESSION DURING DIFFERENTIATION AND TUMORIGENESIS. Mol Cell. 2020 Nov 19;80(4):648-665.e9.

102. Louloupi A, Ntini E, Conrad T, Ørom UAV. Transient N-6-Methyladenosine Transcriptome Sequencing Reveals a Regulatory Role of m6A in Splicing Efficiency. Cell Rep. 2018 Jun 19;23(12):3429–37.

103. Baten A, Chang B, Halgamuge S, Li J. Splice site identification using probabilistic parameters and SVM classification. BMC Bioinformatics. 2006 Dec 18;7(5):S15.

104. Pandya-Jones A, Bhatt DM, Lin CH, Tong AJ, Smale ST, Black DL. Splicing kinetics and transcript release from the chromatin compartment limit the rate of Lipid A-induced gene expression. RNA. 2013 Jun;19(6):811–27.

105. Gao Z, Herrera-Carrillo E, Berkhout B. Delineation of the Exact Transcription Termination Signal for Type 3 Polymerase III. Mol Ther - Nucleic Acids. 2018 Mar 2;10:36–44.

106. Zhang X, Farrell AS, Daniel CJ, Arnold H, Scanlan C, Laraway BJ, et al. Mechanistic insight into Myc stabilization in breast cancer involving aberrant Axin1 expression. Proc Natl Acad Sci. 2012 Feb 21;109(8):2790–5.

107. Parolia A, Cieślik M, Chinnaiyan AM. Competing for enhancers: PVT1 fine-tunes MYC expression. Cell Res. 2018 Aug 1;28(8):785–6.

108. Søndergaard JN, Geng K, Sommerauer C, Atanasoai I, Yin X, Kutter C. Successful delivery of large-size CRISPR/Cas9 vectors in hard-to-transfect human cells using small plasmids. Commun Biol. 2020 Jun 19;3(1):1–6.

109. Rashid M, Coombs KM. Serum-reduced media impacts on cell viability and protein expression in human lung epithelial cells. J Cell Physiol. 2019 Jun;234(6):7718–24.