Splicing Regulation of Alternative Splicing Regulators as a Novel Therapeutic Approach in Breast Cancer

Mariam El-Esnawy

Professor Ihab Younis

Honors Thesis - Spring 2021

Biological Sciences Program

Carnegie Mellon University Qatar

Abstract

Ribonucleic acids (RNAs) in cells are associated with RNA-binding proteins (RBPs), which affect the processing and function of RNAs[1]. This project works with a specific type of RBPs, called heterogeneous nuclear ribonucleoproteins (hnRNPs), which contribute to multiple aspects of RNA metabolism, including alternative splicing[2]. Alternative splicing is the process during gene expression that results in a single gene coding for multiple proteins[3]. We investigated whether the genes that encode the hnRNPs are alternatively spliced such that the regulators of splicing (i.e. hnRNPs) are themselves regulated at the splicing level. Alternative splicing of a given hnRNP would generate different versions of it with various potentials for binding and regulating its downstream targets, which could lead to diseases, like cancer. This has not been extensively studied before, especially in the context of breast cancer. We focused on the alternative splicing of hnRNPA2B1 and hnRNPM in breast cancer, the most common cancer in Qatar. We predicted that there would be breast-cancer-subtype specific splicing of both hnRNPs. Indeed, we identified 2 isoforms of each hnNRP, where each isoform has different expression levels and phenotypic effects in different breast cancer subtypes. This could set the foundation for different therapeutic approaches, which is the ultimate aim of this project.

Introduction

Ribonucleic acids (RNAs) in cells are associated with RNA-binding proteins (RBPs) to form ribonucleoprotein (RNP) complexes. These RBPs affect the structure and interactions of the RNAs, and play important roles in their stability, function, and transport. They also function in every aspect of RNA biology, ranging from transcription, pre-messenger RNA (mRNA) splicing to RNA modification, localization, and translation[1]. In this project, we worked with a specific type of RBPs, called heterogeneous nuclear ribonucleoproteins (hnRNPs). hnRNPs represent a large family of RBPs that contribute to multiple aspects of RNA metabolism, including alternative splicing. It has been reported that the expression level of hnRNPs is altered in many types of cancer, suggesting their role in tumorigenesis. Therefore, additional knowledge about these hnRNPs will be beneficial for the development of new therapeutic approaches as the number of hnRNPs involved in cancer is rapidly increasing[2].

Alternative splicing is the process during gene expression where a single gene is expressed as multiple isoforms that could encode for different proteins. In other words, it is the process in which one pre-mRNA (of a single gene) is processed into various isoforms that encode for multiple proteins, and thus, it is the main reason for transcript variation and proteome diversity[3]. We know that hnRNPs play an important role in alternative splicing, meaning that they regulate the alternative splicing of hundreds of exons and introns in their downstream targets, but this project investigates whether the genes that encode the hnRNPs are alternatively spliced, such that the regulators of splicing (i.e. hnRNPs) are themselves regulated at the splicing level. We investigated this because we discovered that pre-mRNAs of hnRNPs themselves can be alternatively spliced, suggesting that their expression states are different depending on their own alternative splicing. Alternative splicing of a given hnRNP would generate different versions of that hnRNP, which could have different functions with regards to the alternative splicing of its downstream targets as each hnRNP regulates hundreds of different pre-mRNAs. Therefore, generating different isoforms with varying functions of a single hnRNP would have significant effects on the alternative splicing of hundreds of downstream effectors, which could contribute to diseases, such as cancer. So, as the regulators of alternative splicing are potentially alternatively spliced, we proposed that the regulators are regulated by the same process they regulate. However, the alternative splicing of these regulators, the hnRNPs, has not been extensively studied before, especially in the context of breast cancer. As the spliceosome (network of alternatively spliced transcripts) is significantly altered in disease states, such as cancer[3], it is important to uncover the mechanisms that lead to alternative splicing in cancer.

Breast cancer is a disease in which breast cells grow out of control, and acquire many features, such as their ability to metastasize to other tissue, evade growth suppression, resist cell death, avoid immune destruction, and more hallmarks of cancer that are acquired during the multistep progression of breast cancer. Proliferative breast disease is associated with a higher risk of breast cancer, where proliferative breast lesions confer only a small increased risk of developing breast cancer, almost 1.5-2 times of the general population. These include usual ductal hyperplasia, intraductal papillomas, sclerosing adenosis and fibroadenomas. Breast cancer has many risk factors, such as age, family history, and genetic predisposition[4]. Many breast cancer cases are a result of mutations in the *BRCA1* and *BRCA2* genes, which are usually inherited, but other genetic mutations can also lead to breast cancer. Cancers are a result of genetic mutations or epigenetic modifications. Breast cancer is the second leading cause of cancer mortality in women worldwide[4], and it is by far the most common cancer in Qatar. It consists of 31% of cancer cases in women. In addition, there is a 131% increase in breast cancer deaths in Qatar from 1990-2010[5].

The different breast cancer cell lines used in this project are MDAMB231, MDAMB468, T47D and MCF7. These are different breast cancer cell lines with different subtypes as the first two are triple negative cells (TNC), while T47D and MCF7 are ER^{+}, $PR^{+/-}$, and $HER2^{-}$. They also differ in their response to chemotherapy, where all these breast cancer cell lines are often chemotherapy responsive, except MDAMB231, which is the most aggressive breast cancer cell line, and has an intermediate response to chemotherapy, whereas the other three cell lines are chemotherapy responsive[6] (Table 3).

A computational screen done in the lab previously identified 12 hnRNPs whose pre-mRNAs could be alternatively spliced. Alternative splicing of these 12 hnRNPs was first studied using Reverse-Transcription Polymerase Chain Reaction (RT-PCR), and the results from several replicates were analyzed using a scoring system that takes into account if there was alternative splicing or not, if the hnRNP is differentially spliced and/or differentially expressed, and the functional consequences of skipping or including the alternatively spliced region. Two of the three top hnRNPs were hnRNPA2B1 and hnRNPM, which are what this project works with extensively.

hnRNPA2B1 plays a role in different processes, such as transcription, pre-mRNA processing, RNA export amongst others. More specifically, hnRNPA2B1, which forms particles with different hnRNPs, is involved in the transport of specific mRNAs to the cytoplasm[7], binds single-stranded telomeric DNA sequences, binds other RNA molecules, and acts as a regulator of efficiency of mRNA splicing[8]. It also acts as a nuclear reader of the m6A mark[8], and is involved in miRNA sorting into exosomes following sumoylation, possibly by binding (m6A)-containing pre-miRNAs[9]. Finally, it also plays a role in the activation of the innate immune response as it senses the presence of viral DNA in the nucleus, and gets translocated to the cytoplasm, where it activates the TBK1-IRF3 pathway, which leads to interferon alpha/ beta production[10]. This hnRNP is alternatively spliced at exon 2, leading to the production of two distinct mRNA isoforms: mRNA1 is characterized by skipping of exon 2, and mRNA2 has inclusion of exon 2. If mRNA1 is made, the encoded protein loses its NLS, a nuclear localization sequence or signal that is needed for the protein to enter the nucleus as the NLS is encoded by a big part of exon 2. An absent NLS means that hnRNPA2B1 cannot enter the nucleus to carry out any nucleic functions it has. It is now only localized to the cytoplasm.

hnRNPM also plays a role in different processes, mainly the regulation of transcription[11] and cell division[12]. It is also involved in the innate immune response, where it is an important regulator of the gene expression in the innate immune response of macrophages. It does this by repressing a group of transcripts in the infected macrophages. In addition, it induces some interleukins and some tumor necrosis factor alpha cytokines[13]. This hnRNP is alternatively spliced at a part of exon 6, meaning there could be two different mRNA isoforms: mRNA1 has skipping of a big portion of exon 6, and mRNA2 has inclusion of exon 6 entirely. The splicing factors (RBPs) binding on hnRNPM determine which exons are skipped, and which part of the exon is skipped. If mRNA1 is made, the encoded protein loses one of two RRMs, the RNA recognition motifs, which are important for protein-RNA interactions. An absent RRM means that hnRNPM might be unable to bind its targets, or it might bind a different set of targets.

This project utilizes many different techniques carried out on both hnRNPs, in addition to the RBPs that were found to bind on the alternatively spliced region for each hnRNP. These techniques include RT-PCR for the two hnRNPs in different breast cancer subtypes and under different conditions, immunofluorescence to test for hnRNPA2B1's localization, and RT-PCR for a known target for each hnRNP. For the RBPs that were found to bind on the alternatively spliced region of each hnRNP, the RBPs are overexpressed in different breast cancer subtypes, followed by RT-PCR for the hnRNP to check for splicing changes.

We hypothesize that there will be breast-cancer-subtype specific splicing of hnRNPA2B1 and hnRNPM, and we predict that the expression of various isoforms of these 2 hnRNPs will have different effects on breast cancer progression. As we test our hypothesis, we investigate if the overexpression or knockdown of an hnRNP isoform will be better for the cancer cells, allowing for better proliferation and aggression. We can also know how one can possibly change the expression of that hnRNP to be most beneficial to breast cancer patients. In addition, by doing the experiments on the RBPs associated with those hnRNPs, we hypothesize that we can identify splicing enhancer or silencer sequences on these hnRNPs, which could be blocked to also have a different therapeutic approach.

Therefore, additional knowledge about these hnRNPs, which this project aims at doing, will allow us to uncover breast cancer-specific alternative splicing of these hnRNPs. This could set the foundation for different therapeutic approaches, the ultimate aim of this project.

Methods

Cell Culture. MDAMB231, MDAMB468, and MCF7 cells purchased from ATCC (ATCC numbers: MDAMB231: HTB-26, MDAMB468: HTB-132, and MCF7: HTB-22) were seeded at a density of 3,000,000 cells in 10cm plates, and maintained in a culture medium consisting of Dulbecco's modified Eagles Medium (DMEM), 10% Fetal Bovine Serum (FBS), 1% (100 U/ml) Penicillin-Streptomycin mixture, and 1% non-essential amino acids. Seeded cells were allowed to grow in a humidified incubator with 5% carbon dioxide at a temperature of 37°C, where the culture medium was changed every 2-3 days as needed. At 80% confluence, cells were passaged in a 1:2 ratio, or as needed, after being treated with 700µl of 0.05% Trypsin, and incubated for 5 minutes at 37°C until cells were detached.

RT-PCR. RNA was first reverse-transcribed (RT) under the following conditions. The reaction volume was 20µl, containing 1µg RNA, 1µl $d(T)_{23}Vn$ (50uM), 1µl Random Primer Mix, and nuclease-free water to bring up the volume to 8µl. This was incubated for 5 minutes at 65°C, then 10µl 2X Reaction Mix and 2µl 10X Enzyme mix were added, and samples incubated as follows: 5 minutes at 25°C, 1 hour at 42°C, and 5 minutes at 80°C. The resulting cDNA was diluted in 180µl nuclease-free water, and used for PCR. Primers were designed and provided for PCR use. Primer sequences, expected sizes, and annealing temperatures are in Table 1. PCR was carried out using the Eppendorf Thermocycler under the following conditions. Reaction volume was 25µl, and it contained 12.5µl OneTaq (NEB, M0480L), 5µl cDNA, 5.5µl nuclease-free water, 1µl forward primer, and 1µl reverse primer. Cycling conditions were as follows: denaturing of DNA at 94°C for 30 seconds; 32 cycles of 30 seconds at 94°C for denaturation, 30 seconds at 50°C for annealing, and 30 seconds at 68°C for extension; and a final extension at 68°C for 5 minutes, and a hold at 4°C. Amplified products were electrophoresed in 2% TAE agarose gels, containing 1µg/ml ethidium bromide, for 30 minutes at 100V, and imaged in a UV box.

Cloning and Ligation. Cloning primers were designed and provided for use. Primer sequences are in Table 2. Cloning PCR was carried out using the Eppendorf Thermocycler under the following conditions. Reaction volume was 50µl, and it contained 25µl NEBNext® High-Fidelity 2X PCR Master Mix (NEB, M0541), 2.5µl forward primer, 2.5µl reverse primer, 5µl cDNA, and 15µl nuclease-free water. Cycling conditions were as follows: denaturing of DNA at 98°C for 30 seconds; 40 cycles of 30 seconds at 98°C for denaturation, 30 seconds at 50°C for annealing, and 2 minutes at 72°C for extension; and a final extension at 72°C for 2 minutes, and a hold at 4°C. Amplified products were electrophoresed in 1.2% TAE agarose gels, containing 1µg/ml ethidium bromide, for 30 minutes at 100V, and imaged in a UV box. The gel was ran to make sure the desired products were amplified, and following that, PCR products were purified using the QIAquick® PCR Purification Kit (QIAGEN, Cat No./ID: 28104), and eluted in 40µl of nuclease-free water. Restriction enzyme digestions were ran for the purified PCR products. Reaction volume was 25µl, and it contained 2.5µl of 10X buffer, 1µl of enzyme 1, 1µl of enzyme 2, and 20.5µl of DNA from the PCR purification. The buffers and enzymes were purchased from New England Biolabs (NEB). For hnRNPA2B1, the buffer was NEBuffer™ 2.1 (Cat #: B7202S), and the enzymes were KpnI (Cat #: R0142) and EcoRI (Cat #: R0101). For hnRNPM, the buffer was NEBuffer™ 2.1 (Cat #: B7202S), and the enzymes were KpnI (Cat #: R0142) and NotI (Cat #: R0189). The digested products were electrophoresed in 1.2% TAE agarose gels, containing 1µg/ml ethidium bromide, for 30 minutes at 100V, and imaged in a UV box. The desired bands on the gels were extracted using the QIAquick® Gel Extraction Kit (Cat No./ID: 28704), and the eluted DNA was then used for ligation. Reaction volume was 20µl, and followed NEB's Quick Ligation Protocol (M2200), which contained 10µl of Quick Ligase reaction buffer (2X), 50ng of vector DNA, 37.5ng of insert DNA, 1µl of Quick Ligase, and nuclease-free water to bring up the volume to 20µl. The reaction was mixed by pipetting up and down, and briefly centrifuging, then the reaction was incubated at 25°C for 5 minutes, and then placed on ice. The products were then transformed into competent cells.

Transformation. DNA was diluted to 10ng in TE buffer, and 1µl of diluted DNA was added to 50µl of DH5α competent cells. Reaction was incubated for 30 minutes on ice, then the heat shock took place, where the tubes were

placed at 42°C for 45 seconds. They were then incubated on ice for 2 minutes, and 450µl of pre-warmed Super Optimal Broth (SOC) medium (Invitrogen, catalog number: 15544034) was added at room temperature. Tubes were placed in an air shaker at 37°C for 1 hour. 100µl from tubes was spread plated on LB+ampicillin plates after shaking for an hour, and plates were placed in the 37°C incubator overnight. LB+ampicillin medium was prepared, and one colony from the plates was inoculated in 3ml LB+ampicillin, and placed in an air shaker at 37°C for several hours. The 3ml were then added to 250ml LB+ampicillin, and left to shake overnight at 37°C. The DNA was extracted from the bacterial culture using the QIAGEN plasmid midi kit (Cat No./ID: 12145).

Transfection. MDAMB231, MDAMB468, and MCF7 cells were plated at a density of 200,000 cells per well in 6-well plates, maintained in growth medium consisting of Dulbecco's modified Eagles Medium (DMEM), 10% Fetal Bovine Serum (FBS), 1% (100 U/ml) Penicillin-Streptomycin mixture, and 1% non-essential amino acids. Seeded cells were allowed to grow in a humidified incubator with 5% carbon dioxide at a temperature of 37°C overnight. 24 hours after plating, or when cells were 70-90% confluent, media was changed on the cells in preparation for transfection. Old media was aspirated, and growth medium without Penicillin-Streptomycin mixture was added. 5µg of DNA was added to 150µl Opti-MEM™ I Reduced Serum Medium (ThermoFisher, Catalog number: 31985062), then a mix of 155µl of Opti-MEM™ and Lipofectamine™ LTX Reagent (ThermoFisher, Catalog number: 15338500) (150µl Opti-MEM™ and 5µl Lipofectamine™ LTX Reagent) was added to DNA, and incubated at room temperature inside cell culture hood for 10 minutes. Entire mix was then added to the cells slowly, and allowed to grow in a humidified incubator with 5% carbon dioxide at a temperature of 37°C for 2 days. RNA and protein were harvested from the plates 48 hours following transfection.

RNA and Protein Extraction. RNA was harvested from live cells following transfection using QIAGEN's RNeasy Mini Kit (Cat No./ID: 74104) according to manufacturer's recommendations, eluted in 30µl RNAase-free water, and stored at -80°C. Protein was harvested from live cells following transfection as follows. 250µl of 1X Laemmli Sample Buffer and 2.5% 2-Mercaptoethanol (BME) were added to the wells after aspirating the media, and cells were scraped using a cell scraper. 250µl was aspirated from the wells following scraping, and transferred to 2ml Eppendorf tubes. Proteins were stored at -20°C.

Immunofluorescence. MDAMB231 and MDAMB468 cells were plated at a density of 150,000 cells per well in 8-well slide chambers, maintained in a culture medium consisting of Dulbecco's modified Eagles Medium (DMEM), 10% Fetal Bovine Serum (FBS), 1% (100 U/ml) Penicillin-Streptomycin mixture, and 1% non-essential amino acids. Seeded cells were allowed to grow in a humidified incubator with 5% carbon dioxide at a temperature of 37°C overnight. 24 hours after plating, media was decanted, and cells washed 3 times with 250µl Phosphate Buffer Saline (PBS). 250µl of paraformaldehyde (PFA) + 0.03% Triton was added to cells, and allowed to incubate for 30 minutes at room temperature. Cells were then washed 3X with PBS. 250µl of PMZ buffer (0.2% BSA, 50mM NH_4Cl in PBS) was then added to the cells, and allowed to incubate for 30 minutes at room temperature. PMZ buffer was removed, and cells washed 3X with PBS. 250µl of PBS were added to cells, and left overnight at 4°C. Anti-hnRNPA2B1 antibody (abcam, ab31645) was diluted 1:100 in PMZ. Alexa Fluor™ 488 Phalloidin (Invitrogen™, catalog number: A12379) was diluted 1:10 and used to stain actin filaments. Negative controls were PMZ buffer alone and secondary antibody alone. DAPI (1:10 dilution) was also added to stain the nuclei. These conditions were incubated for 30 minutes at room temperature. Cells were then washed 3X with PBS. Secondary antibody was prepared and diluted in PMZ as follows: Goat anti-Mouse IgG (H+L) Secondary Antibody, DyLight 488 (ThermoFisher, Catalog # 35502) diluted 1:1000. 50µl of the diluted secondary antibody was added to the cells, and allowed to incubate for 30 minutes at room temperature in the dark. Cells were then washed 3X with PBS, and 9µl mounting agent was added, and allowed to incubate for 1 hour at 70°C in the oven. Cells were then imaged under the EVOS cell imaging system (ThermoFisher) under the GFP and DAPI settings.

Bioinformatics. The ImageJ software (National Institutes of Health (NIH), Bethesda, MD, USA) was used to quantify the bands from all the agarose gels presented, and the quantifications from this software were then normalized to a housekeeping gene, GapDH, and plotted. Another computational tool was the UCSC genome browser's links specifically for hnRNPA2B1 and hnRNPM that show where different RBPs bind on the hnRNP (graciously provided by Muhammad Nahin Khan). This tool was used to view which RBPs bind close to or on the alternatively spliced regions of each hnRNP, and that list of RBPs is used to narrow down the RBPs to work on in the future. Finally, the last computational tool was the Human Protein Atlas by the Knut & Alice Wallenberg foundation, which allowed us to view immunohistochemistry images for hnRNPA2B1 and other hnRNPs in normal tissue and breast cancer.

Results and Analysis

Table 1: Primer sequences, annealing temperatures, and expected sizes for primers used for RT-PCR

Gene name	Location of primer	Forward (F) / Reverse (R)	Primer Sequence	Annealing temperature (°C)	Expected sizes with included exons (bp)
hnRNPA2B1	Exon 1	F	CAGCGGCAGTTCTCACTACA	53.8	254
	Exon 3	R	TAAGCTTTCCCCATTGTTCG	49.7	
hnRNPM	Exon 5	F	AGCTGCGGAAGTCCTAAACA	51.8	397
	Exon 7	R	TTGCACAGCTTCAATGGACT	49.7	
CD44	Exon V2	F	GCACAGACAGAATCCCTGCTACCA	60.8	1128
	Exon V10	R	CCTTCTTGACTCCCATGTGAGTG	57.4	
COX16	Exon 1	F	GCTATGGAGTCCCCATGTTG	55.5	240
	Exon 4	R	TTCTTCCTTGGAGGAGGTCA	55.2	

Table 2: Primer sequences and annealing temperatures for primers used for cloning

Gene name	Forward (F) / Reverse (R)	Primer Sequence	Annealing temperature (°C)
hnRNPA2B1	KpnI- F	TTTGGTACCATGGAGAAAACTTTA	52.5
	EcoRI- R	TTTGAATTCGTATCGGCTCCTCCC	58.9
hnRNPM	KpnI- F	TTTGGTACCATGGAAGAGAGCATG	57.0
	NotI- R	TTTCGGCCGAGCGTTTCTATCAAT	59.0

Table 3: Information on Breast Cancer cell lines used throughout project

Cell Line	Response to Hormonal Therapy			Response to Chemotherapy
	Estrogen	Progesterone	HER2	
MDAMB231	-	-	-	Most aggressive; intermediate response to chemotherapy
MDAMB468				Chemotherapy responsive
T47D	+	+/-	-	
MCF7				

Table 4: Information on non-breast cancer cell lines used throughout project

Cell Line	Origin
293T	Human embryonic kidney cells (non-cancerous)
U2OS	Human bone osteosarcoma epithelial cells (bone cancer)

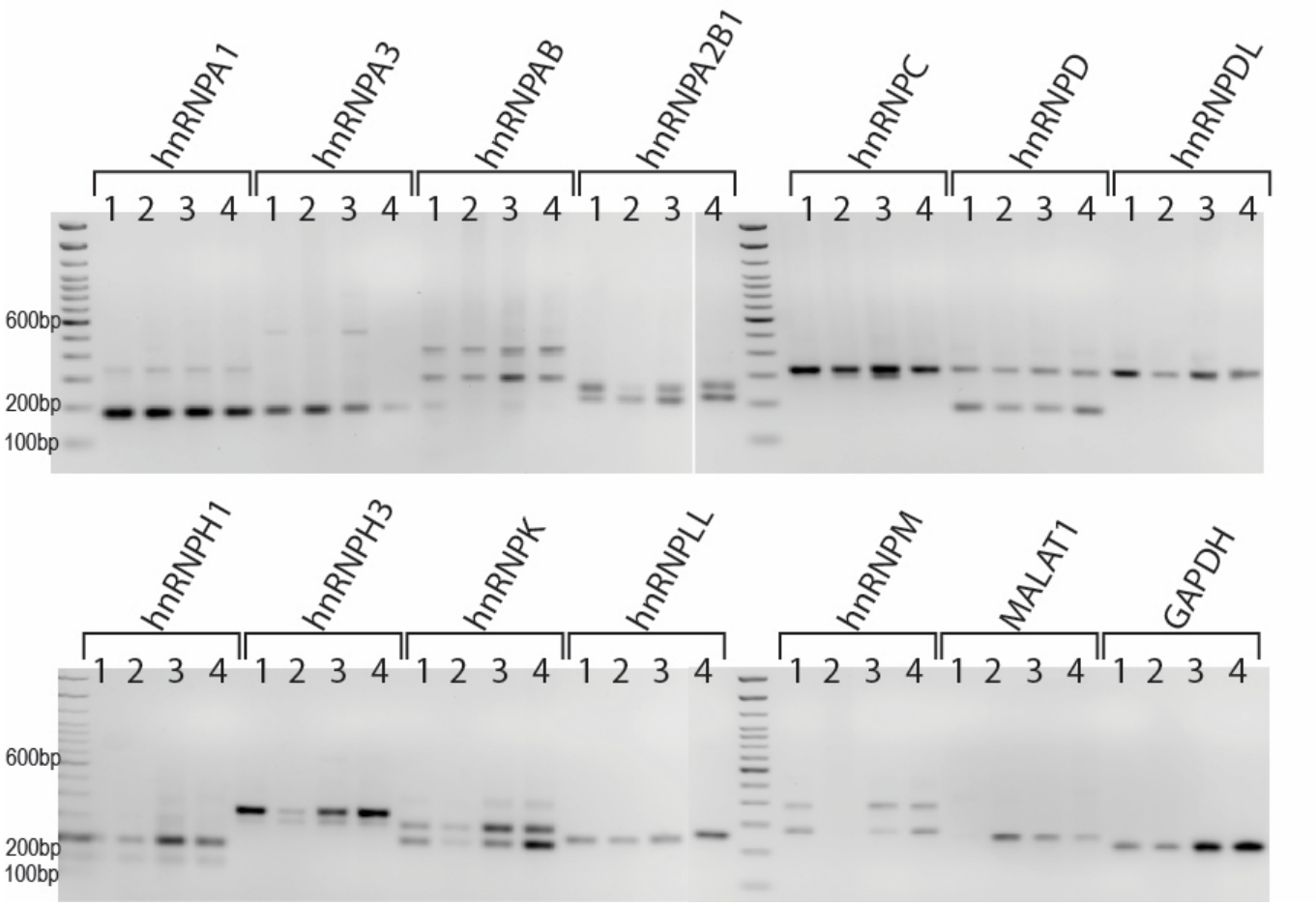


Figure 1: PCR Gels for 12 hnRNPs in MDAMB231, T47D, MCF7 and MDAMB468:
Loaded on these gels are the PCR products after amplifying the alternatively spliced regions of 12 hnRNPs. 1 = MDAMB231, 2 = T47D, 3 = MCF7, and 4 = MDAMB468. These are 2% agarose gels with 5µl ethidium bromide, and the gels immersed in TAE buffer, and ran for 1 hour at 95V. They were then imaged under UV light. The molecular weight marker used here is the 100bp ladder[a]. For this, 5µl of ladder was loaded, and 10µl of the PCR products. The PCR products were 25µl, and to that, 5µl of 6X loading dye was added.

a. '100bp DNA Ladder'. *ThermoFisher*. Retrieved from https://www.thermofisher.com/order/catalog/product/15628050#/15628050

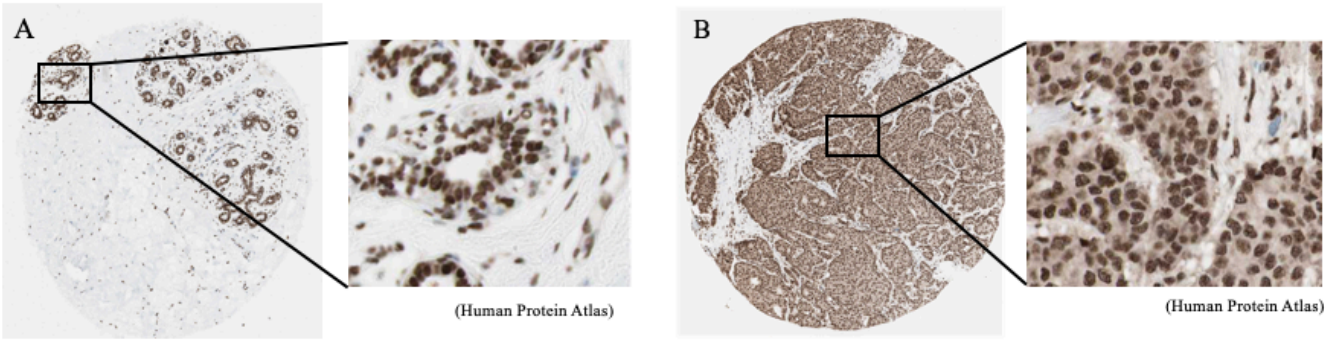


Figure 2: Localization images for hnRNPA2B1 provided by the *Human Protein Atlas*:

Immunohistochemistry images of a normal breast tissue of a 23-year-old female (panel A)[14] and a 75-year-old female patient with duct carcinoma (panel B)[15]. The brown staining is hnRNPA2B1, and panel A (normal tissue) is completely nuclear, while panel B (breast cancer) has more cytoplasmic localization. The localizations are labeled by the *Human Protein Atlas.*

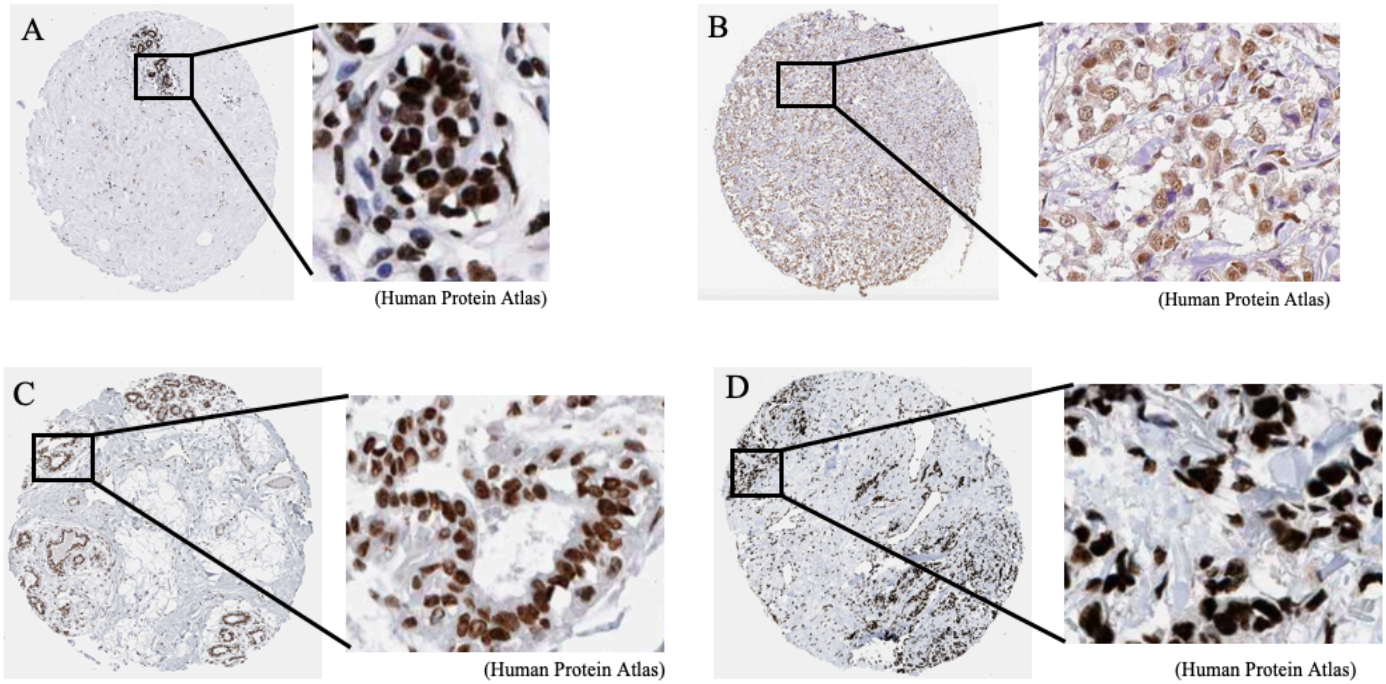


Figure 3: Localization images for more hnRNPs provided by the *Human Protein Atlas*:

Immunohistochemistry images for hnRNPA1 (panels A[16] and B[17]) and hnRNPM (panels C[18] and D[19]). The brown staining is the respective hnRNP, and panels A and C are normal tissue, while panels B and D are breast cancer. All the localizations are labeled nuclear, according to the *Human Protein Atlas*.

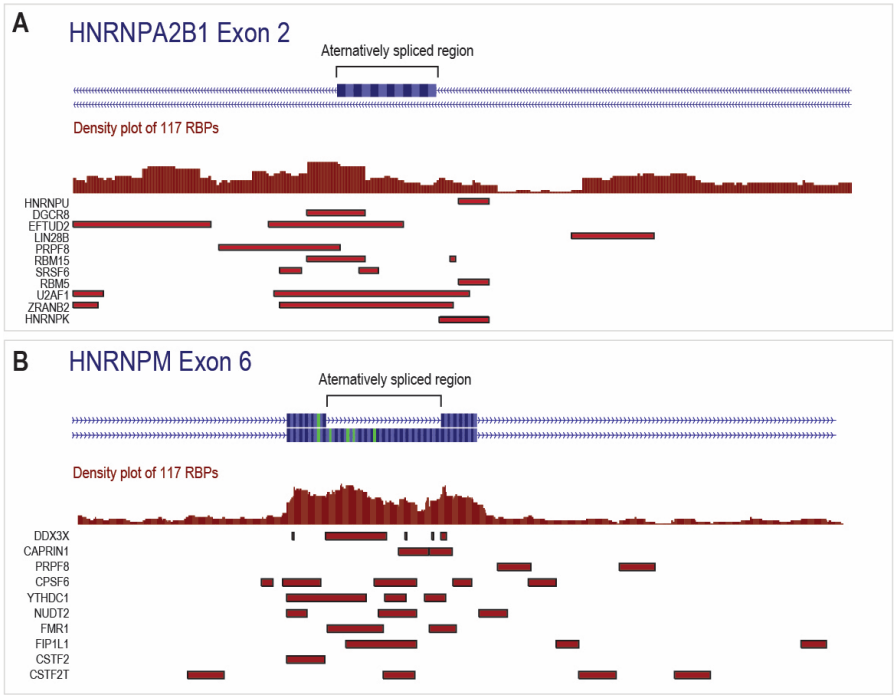


Figure 4: RBPs that bind on the alternatively spliced region of each hnRNP:

RBPs that bind on exon 2 of hnRNPA2B1 (panel A) and RBPs that bind on exon 6 of hnRNPM (panel B). These are from the UCSC genome browser, and the list narrowed down the RBPs to 3 per hnRNP, which are further studied in the lab.

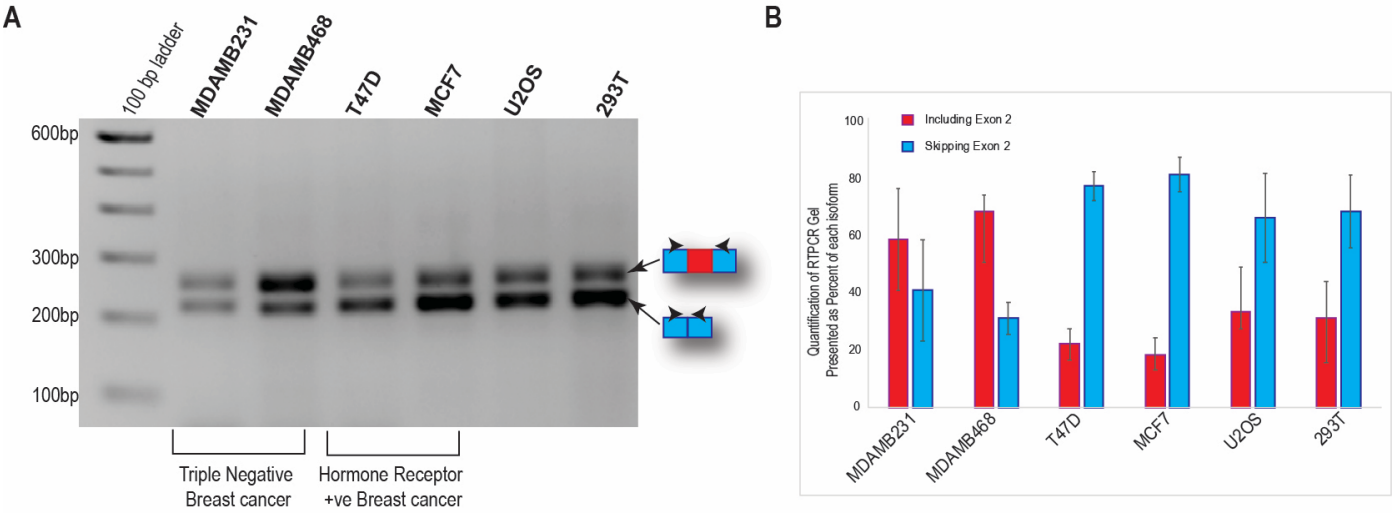


Figure 5: RT-PCR gel for hnRNPA2B1's exon 2 inclusion in different cell lines and the quantification:

(A) Loaded on these gels are the PCR products after amplifying the alternatively spliced region of hnRNPA2B1 on 6 different cell lines: MDAMB231, MDAMB468, T47D, MCF7, U2OS, and 293T. This is a 2% agarose gel with 5µl ethidium bromide, and the gel immersed in TAE buffer, and ran for 1 hour at 95V. It was then imaged under UV light. The molecular weight marker used here is the 100bp ladder. (B) The graph shows the average quantification of 3 gels that are different, biological repeats, where the PCR and gel electrophoresis were repeated three different times. The gel image quantifications reported in the graph were quantified by the ImageJ software. Values are plotted as percentages of each isoform $\pm$ standard deviation; n = 3. Percentage of each isoform is calculated by dividng each isoform's quantification by the total added quantification of both bands, and multiplying by 100.

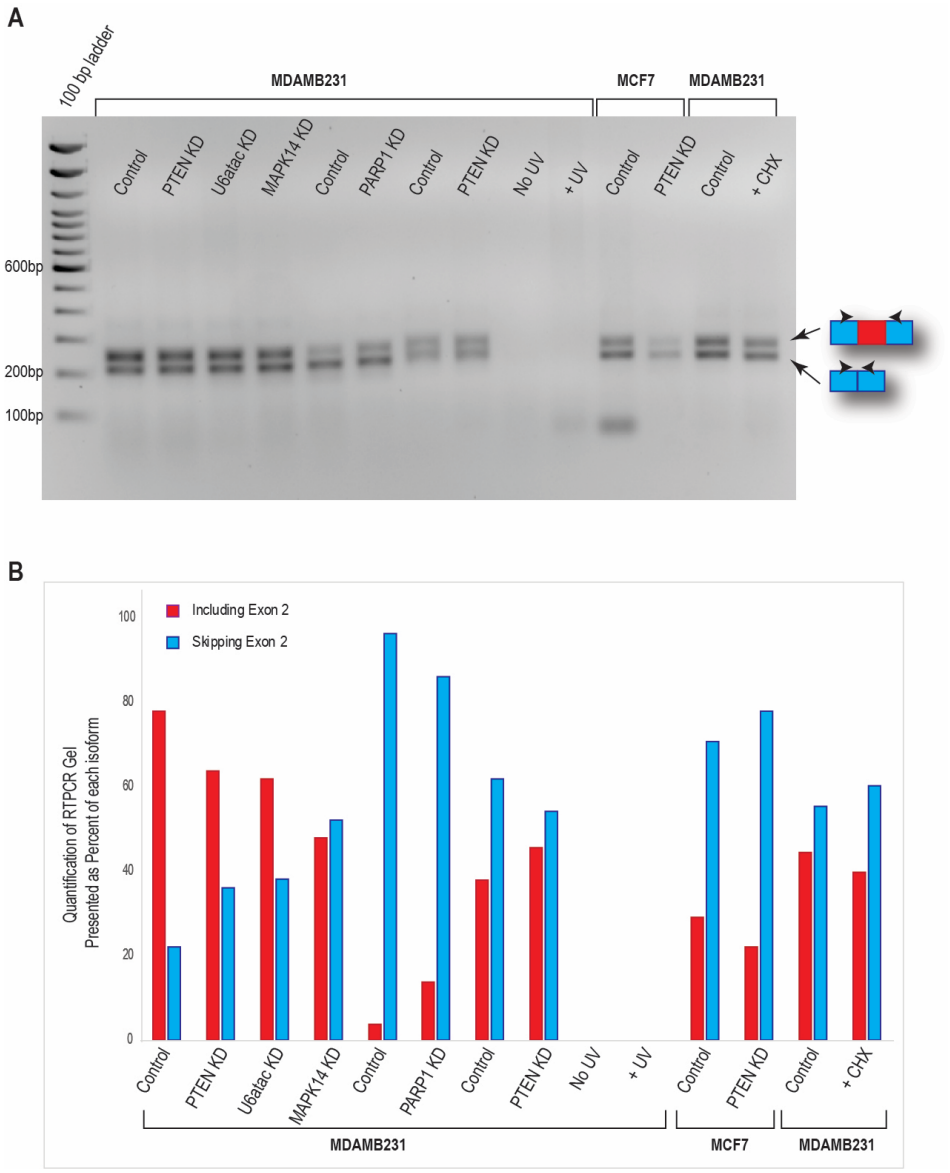


Figure 6: RT-PCR gel for hnRNPA2B1's exon 2 inclusion under different conditions and the quantification:

(A) Loaded on this gel are the PCR products after amplifying the alternatively spliced region of hnRNPA2B1 under different conditions. KD = knockdown, and CHX = cycloheximide. This is a 2% agarose gel with 5µl ethidium bromide, and the gel immersed in TAE buffer, and ran for 30 minutes at 100V. It was then imaged under UV light. The molecular weight marker used here is the 100bp ladder. (B) The graph shows the quantification of this gel. The gel image quantifications reported in the graph were quantified by the ImageJ software. Values are plotted as percentages of each isoform; n = 1.

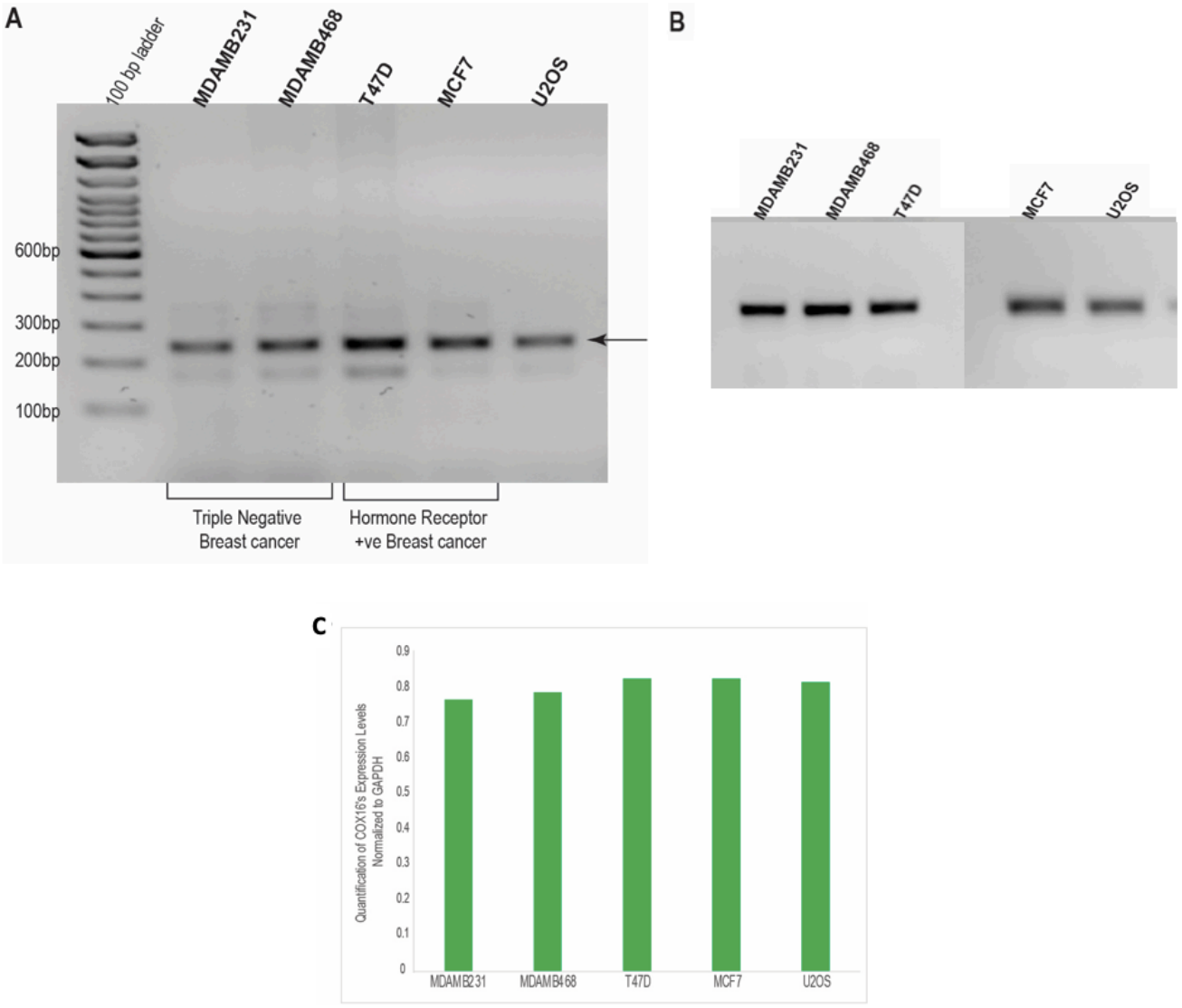


Figure 7: RT-PCR gel for hnRNPA2B1's known target, COX16, and the quantification:

(A) Loaded on this gel are the PCR products after amplifying COX16, hnRNPA2B1's known target, in different cell lines: MDAMB231, MDAMB468, T47D, MCF7 and U2OS. This is a 2% agarose gel with 5µl ethidium bromide, and the gel immersed in TAE buffer, and ran for 30 minutes at 100V. It was then imaged under UV light. The molecular weight marker used here is the 100bp ladder. (B) Loaded on this gel are the PCR products after amplifying GapDH, the housekeeping gene, used as a loading control, in different cell lines: MDAMB231, MDAMB468, T47D, MCF7 and U2OS. This is a 2% agarose gel with 5µl ethidium bromide, and the gel immersed in TAE buffer, and ran for 30 minutes at 100V. It was then imaged under UV light. (C) The graph shows the quantification of the COX16 gel. The gel image quantifications reported in the graph were quantified by the ImageJ software. Values are plotted as relative expression levels normalized to GapDH, the housekeeping gene used as a loading control; n = 1.

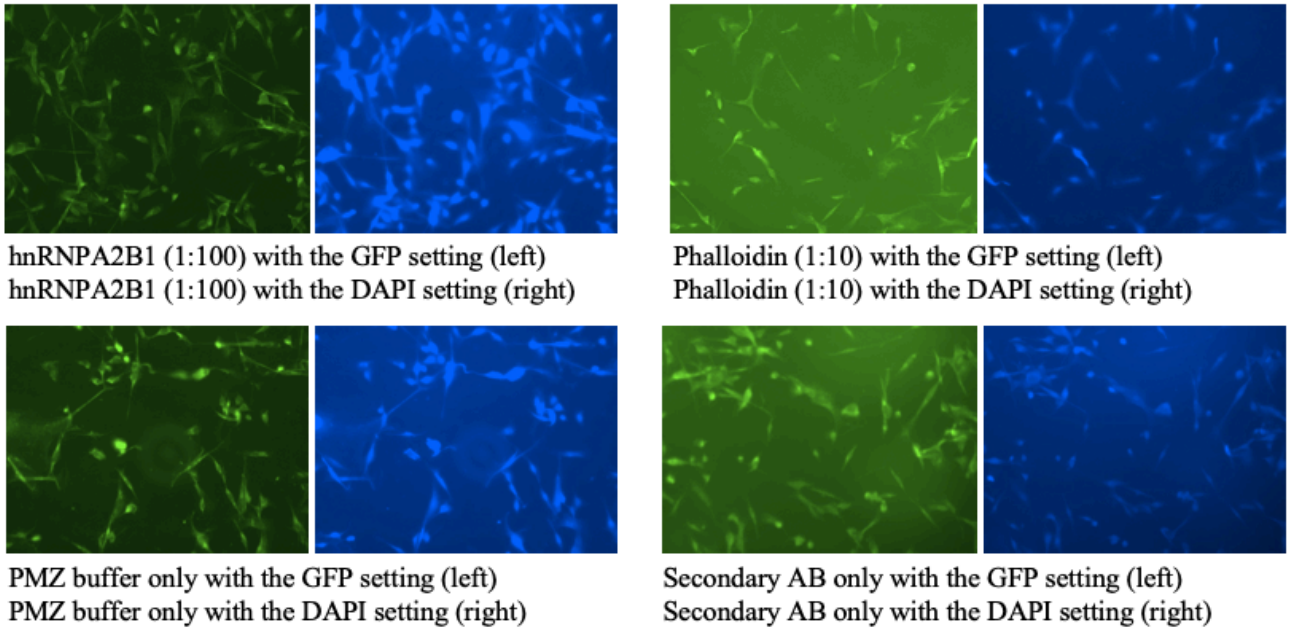
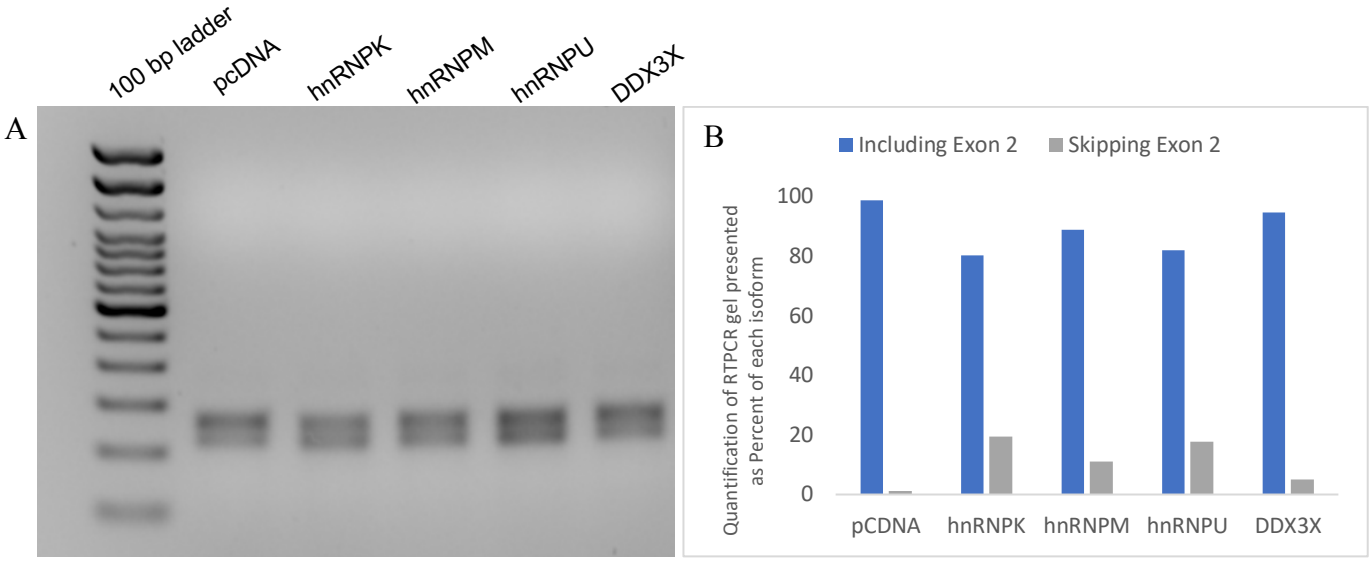


Figure 8: Immunofluorescence images for hnRNPA2B1 and phalloidin in MDAMB231 cells:

4 different conditions were studied: hnRNPA2B1, phalloidin (used as a control to view actin), PMZ buffer only, and secondary antibody only. All the conditions had DAPI (a dye that stains the nucleus). The cells are imaged under the EVOS imaging system after permeabilization, fixation, and incubation with primary and secondary antibodies. They are imaged under the GFP (Green Fluorescent Protein) setting and under the DAPI setting.



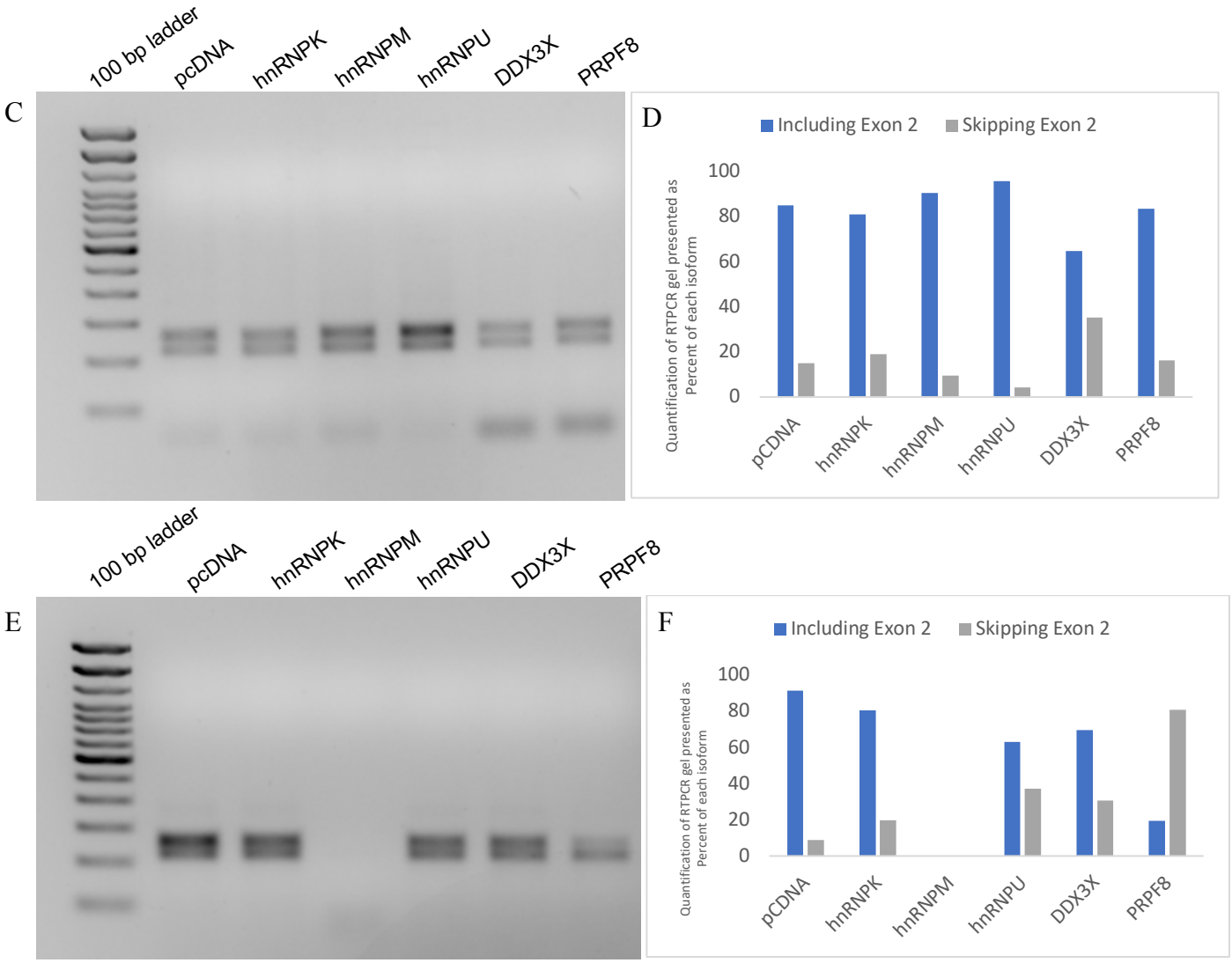


Figure 9: RT-PCR gels for hnRNPA2B1 following the overexpression of RBPs in MDAMB231, MDAMB468, and MCF7 cells:

Five different RBPs (hnRNPK, hnRNPM, hnRNPU, DDX3X and PRPF8) were overexpressed in MDAMB231 (Panels A and B), MDAMB468 (Panels C and D), and MCF7 (Panels E and F) cells.
(A, C and E) Loaded on the gels are the RT-PCR products carried out for hnRNPA2B1 after the overexpression. They are 2% agarose gels with 5µl ethidium bromide, immersed in TAE buffer, and ran for 30 minutes at 100V. They were then imaged under UV light. The molecular weight marker used here is the 100bp ladder.
(B, D and F) The graphs show the quantification of the gels. The gel image quantifications reported in the graphs were quantified by the ImageJ software. Values are plotted as percentages of each isoform; n = 1.

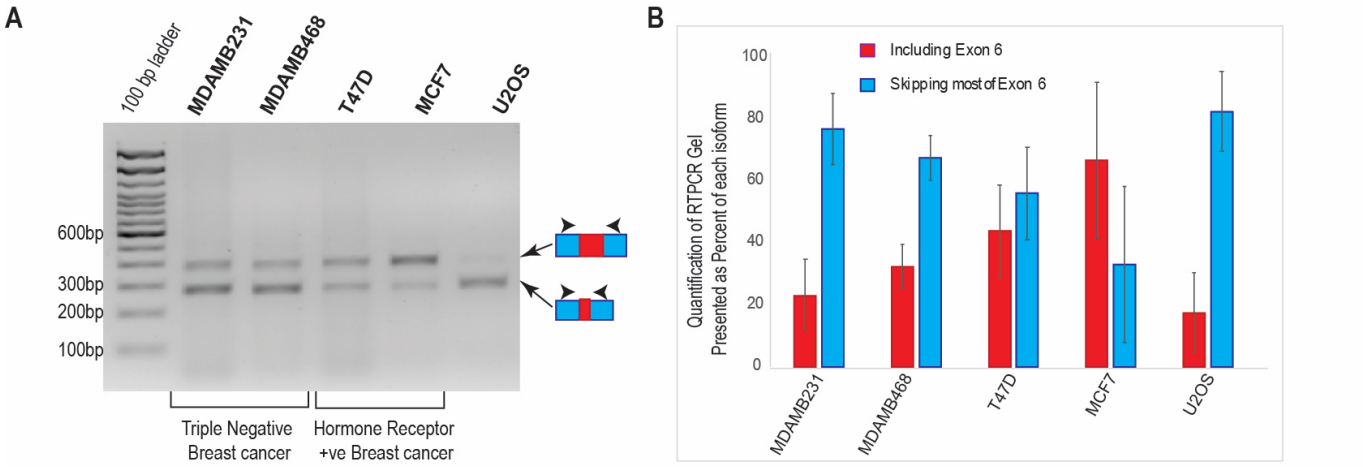


Figure 10: RT-PCR gel for hnRNPM's exon 6 inclusion in different cell lines and the quantification:

(A) Loaded on these gels are the PCR products after amplifying the alternatively spliced region of hnRNPM on 5 different cell lines: MDAMB231, MDAMB468, T47D, MCF7, and U2OS. This is a 2% agarose gel with 5µl ethidium bromide, and the gel immersed in TAE buffer, and ran for 1 hour at 95V. It was then imaged under UV light. The molecular weight marker used here is the 100bp ladder. (B) The graph shows the average quantification of 3 gels that are different, biological repeats, where the PCR and gel electrophoresis was repeated three times. The gel image quantifications reported in the graph were quantified by the ImageJ software. Values are plotted as percentages of each isoform $\pm$ standard deviation; n = 3.

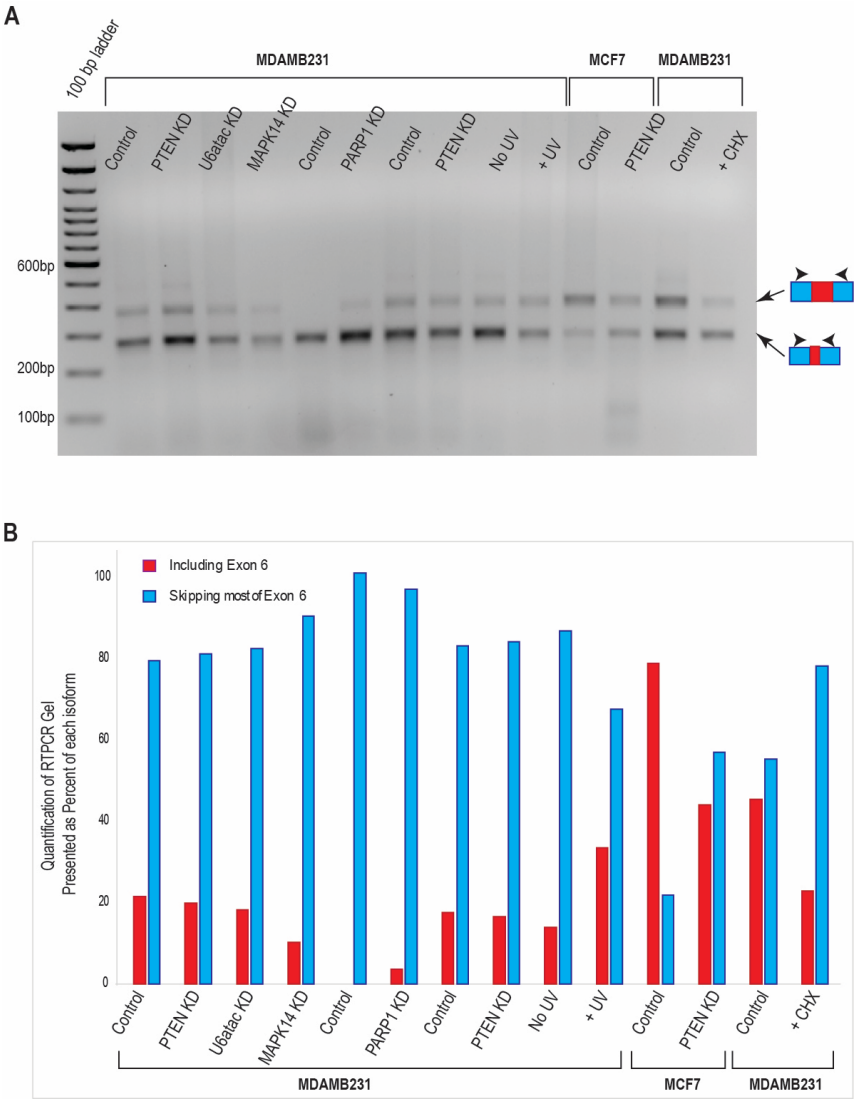


Figure 11: RT-PCR gel for hnRNPM's exon 6 inclusion under different conditions and the quantification:

(A) Loaded on this gel are the PCR products after amplifying the alternatively spliced region of hnRNPM under different conditions. KD = knockdown, and CHX = cycloheximide. This is a 2% agarose gel with 5µl ethidium bromide, and the gel immersed in TAE buffer, and ran for 1 hour at 95V. It was then imaged under UV light. The molecular weight marker used here is the 100bp ladder. (B) The graph shows the quantification of this gel. The gel image quantifications reported in the graph were quantified by the ImageJ software. Values are plotted as percentages of each isoform; n = 1.

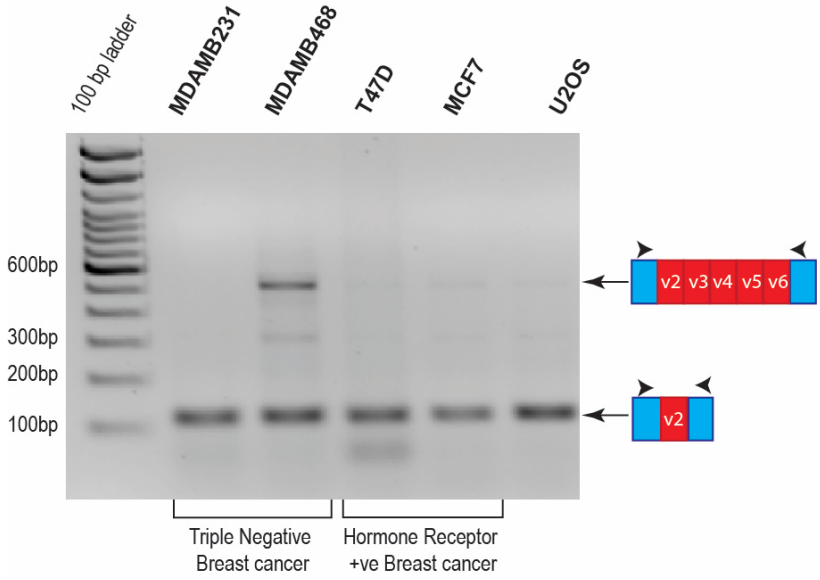
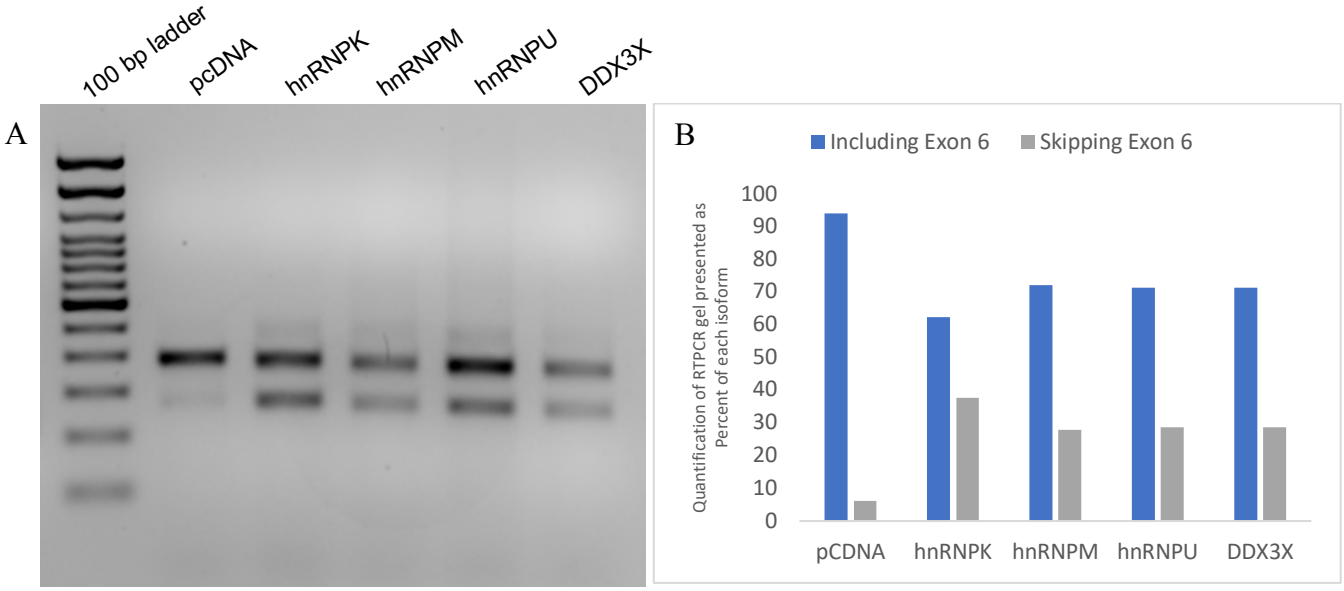


Figure 12: RT-PCR gel for hnRNPM's known target, CD44:

Loaded on this gel are the PCR products after amplifying CD44, hnRNPM's known target, in different cell lines: MDAMB231, MDAMB468, T47D, MCF7 and U2OS. This is a 2% agarose gel with 5µl ethidium bromide, and the gel immersed in TAE buffer, and ran for 30 minutes at 100V. It was then imaged under UV light. The molecular weight marker used here is the 100bp ladder.



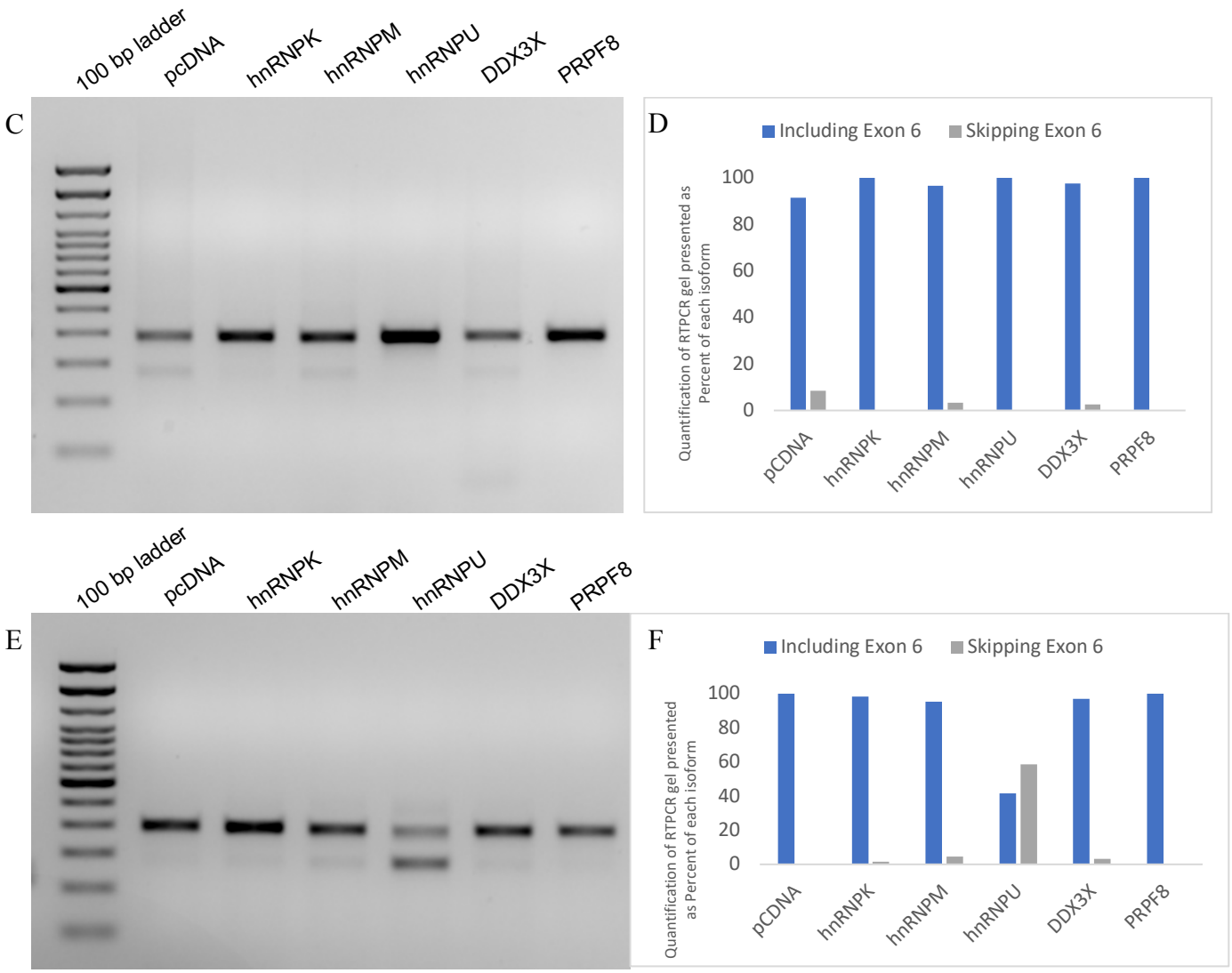


Figure 13: RT-PCR gels for hnRNPM following the overexpression of RBPs in MDAMB231, MDAMB468, and MCF7 cells:

Five different RBPs (hnRNPK, hnRNPM, hnRNPU, DDX3X and PRPF8) were overexpressed in MDAMB231 (Panels A and B), MDAMB468 (Panels C and D) and MCF7 (Panels E and F) cells.
(A, C and E) Loaded on the gels are the RT-PCR products carried out for hnRNPM after the overexpression. They are 2% agarose gels with 5µl ethidium bromide, immersed in TAE buffer, and ran for 30 minutes at 100V. They were then imaged under UV light. The molecular weight marker used here is the 100bp ladder.
(B, D and F) The graphs show the quantification of the gels. The gel image quantifications reported in the graphs were quantified by the ImageJ software. Values are plotted as percentages of each isoform; n = 1.

Below is a brief description of the results from this project.
Tables 1 and 2 summarize the primers used in this project for RT-PCR and for cloning the different isoforms of hnRNPA2B1 and hnRNPM. Tables 3 and 4 provide information about the cell lines used throughout this project, and these are the cell lines used in all the gel images that follow.

Figure 1 is the preliminary data for the project. Initially, a screen was done in the lab that identified 12 hnRNPs whose pre-mRNAs could be alternatively spliced. These 12 hnNRPs and their alternatively spliced regions were found computationally using RNASeq data. The alternative splicing of these 12 hnRNPs was first studied using RT-PCR, and the results from several replicates were analyzed using a scoring system, which narrowed down those 12 hnRNPs to hnRNPA2B1 and hnRNPM. The scoring system was further explained in the introduction.

Figure 2 is the localization images for hnRNPA2B1, which shows that hnRNPA2B1 is localized to the nucleus in normal tissue, but it is localized both to the nucleus and cytoplasm in breast cancer. The significance of this will be explained in the discussion. Figure 3 is also localization images, but for more hnRNPs, and it shows that for other hnRNPs, there is no difference in localization between normal and breast cancer tissue, suggesting that the localization patterns for hnRNPA2B1 can be unique, when compared to the small fraction of hnRNPs we study, which will also be further explained in the discussion.

Figure 4 is an in-depth computational analysis of the RBPs that bind on the alternatively spliced region of each hnRNP. Using this, the RBPs binding on exon 2 for hnRNPA2B1 and exon 6 for hnRNPM were studied to see if any of them might possibly regulate the alternative splicing of the hnRNPs. A scoring system was implemented for the list of RBPs, which led to the narrowing down of the RBPs to 3 RBPs per hnRNP. For hnRNPA2B1, these are hnRNPU, hnRNPK, and PRPF8, while for hnRNPM, these are DDX3X, PRPF8, and YTHDC1. These RBPs are further studied in the lab.

Figure 5 is the RT-PCR gel for hnRNPA2B1's exon 2 inclusion in the 6 different cell lines presented in tables 3 and 4, in addition to the average quantification of those gels. The plot shows us which of the 2 isoforms is more favored in each cell line. This experiment was repeated 3 times, and therefore the graph is a representation of the average, along with the standard deviations plotted as the error bars.

Figure 6 is also exon 2 inclusion for hnRNPA2B1, but under different conditions, where we used RNA from various experiments done previously by other students in the lab. As such, these conditions are not a comprehensive list of what can be done. Also, this experiment was repeated only once, and thus the controls and bands might look quite different. This experiment needs to be repeated, but it was only done as a quick screen (explained in the discussion). PTEN KD is listed twice as it was knocked down differently both times; once with an inhibitor, and once with an anti-sense morpholino oligonucleotide (AMO). The quantification also shows us which isoform is favored under each condition, and the significance of this will be further explained in the discussion.

Figure 7 is the RT-PCR for a known target of hnRNPA2B1, which is COX16. We wanted to study the alternative splicing patterns of a known target that hnRNPA2B1 regulates, which is COX16. This is because the authors of a published study[8] concluded that there are alternative splicing changes occurring to COX16 when there is a change to hNRNPA2B1's expression as they did the experiment when depleting or knocking down hnRNAP2B1. So, we designed primers for COX16, and did an RT-PCR experiment for it, which is presented in figure 7, along with the expression quantification normalized to GapDH (a loading control).

Figure 8 is a compilation of the immunofluorescence images done on MDAMB231 cells under 4 conditions: hnRNPA2B1, phalloidin, PMZ buffer only, and secondary antibody only. Phalloidin is added to stain actin, which should be present, and therefore is used as a positive control. The PMZ buffer condition is added as a negative control

because nothing should be viewed under that condition, and the secondary antibody condition should show non-specific binding. All the images are captured under the GFP and DAPI settings. The GFP setting was used because the secondary antibody was conjugated to a green fluorophore, and DAPI is used to stain the nuclei, and therefore its setting should show us the nuclei.

Finally, to conclude hnRNPA2B1's experiments, we overexpressed 5 different RBPs in MDAMB231, MDAMB468, and MCF7, and did RT-PCR for hnRNPA2B1, which is presented in figure 9, along with the quantification for each isoform of hnRNPA2B1 after the overexpression of each RBP in the different cell lines.

The results for hnRNPM are quite similar. Figure 10 depicts the RT-PCR gel for hnRNPM's exon 6 inclusion in the same cell lines, but 293T is not presented in figure 10A as we ran out of RNA to reverse-transcribe, but it was present in other gels, and that is why it is presented in figure 10B. Figure 10B is also the average quantification of 3 gels, in addition to the standard deviations. Figure 11 is hnRNPM's exon 6 inclusion under the same conditions for hnRNPA2B1, and the plot also shows us which isoform is favored under each condition. Figure 12 is the RT-PCR gel for hnRNPM's known target, CD44. CD44 has been identified as a key downstream target of hnRNPM[20], and therefore we carry out RT-PCR for it as well. CD44 has variable exons, which are labeled in figure 12A. Figure 13 is the same experiment as figure 9, but for hnRNPM. It is the RT-PCR for hnRNPM following the overexpression of 5 RBPs in MDAMB231, MDAMB468 and MCF7 cells. The significance of this is further explained in the discussion.

Discussion

This project started with computational analyses, which helped outline the experiments done in the lab. The results provided by the Human Protein Atlas inspired us to carry out localization experiments using immunofluorescence to test whether the localization is nuclear or cytoplasmic. The data from this analysis show that the hnRNPA2B1 protein has differences in localization in breast cancer when compared to normal tissue. In breast cancer, the localization is nuclear and cytoplasmic, while it is only nuclear in normal tissue. This suggests that the alternatively spliced isoforms, one with an intact NLS and one that lacks an NLS, could be present in different ratios in cancer or normal cells. Therefore, it was important to carry out localization experiments in different cell lines to study the localization of the hnRNP in breast cancer. In addition, to prove that the localization patterns (nuclear and cytoplasmic in breast cancer) are unique for hnRNPA2B1, and that it is not that the pathway for nuclear localization that is possibly defective in breast cancer, we analyzed more immunohistochemistry images for more hnRNPs (Figure 3), and it was observed that the localization in normal and breast cancer tissue is nuclear. This suggests that the localization of hnRNPA2B1 to the cytoplasm in breast cancer can be unique, and could be a product of the alternative splicing pattern that leads to losing the NLS.

Additional computational work (Figure 4) shows an in-depth analysis of the RBPs that bind on the alternatively spliced region of each hnRNP. The RBPs that bind on the alternatively spliced region of hnRNPA2B1 and hnRNPM are expected to regulate the splicing of these two hnRNPs, which themselves modulate the splicing of hundreds of introns and exons in their downstream targets. Therefore, these RBPs in Figure 4 regulate the alternative splicing of the regulators. For each list of RBPs on each hnRNP, a scoring system was implemented, which took into consideration the following: whether the UCSC results are computationally predicted or experimentally verified, the correlation of the RBP with the hnRNP and with the alternatively spliced exon, and the function of the RBP. This scoring system narrowed down the RBPs to 3 per hnRNP, which were discussed in the analysis. We overexpressed these RBPs, and then carried out RT-PCR for the hnRNPs to see if different alternative splicing patterns of the hnRNPs will be observed, which would mean that those RBPs are indeed regulating the alternative splicing of the regulators.

We next performed many laboratory-based experiments, which will be discussed here. First, to check for differential expression of the two alternatively spliced isoforms in different breast cancer subtypes, we performed RT-PCR for exon 2 of hnRNPA2B1 and exon 6 of hnRNPM. Firstly, the data for hnRNPA2B1 will be discussed here. As shown in Figure 5A, in all the cell lines, both isoforms (one with the NLS, and one lacking it) are present, but they differ in their relative expression levels in the different cell lines, which is quantitated in Figure 5B. The gel and the plot show that the triple negative breast cancer cells, namely MDAMB231 and MDAMB468, both make more of the isoform that includes exon 2, and therefore can enter the nucleus. On the other hand, the hormonal positive breast cancer cell lines (T47D and MCF7), the non-cancerous cell line (293T), and the bone cancer cell line (U2OS) all make more of the isoform that skips exon 2, and therefore should be defective in their ability to enter the nucleus. Thus, we conclude that only triple negative breast cancer cells favor the isoform including exon 2, and so it could be that there are nuclear functions for hnNRPA2B1 that make the cells more aggressive as the triple negative breast cancer cells are the only ones that make more of the isoform that can enter the nucleus. This is not inconsistent with the immunohistochemistry images for hnRNPA2B1's localization (figure 2) as the RT-PCR results (figure 5) are for different breast cancer subtypes, while the immunohistochemistry images are for normal and breast cancer tissue. In the different breast cancer subtypes, from figure 5, we see that we always have both isoforms, and it is a matter of which is expressed more than the other, and therefore both nuclear and cytoplasmic isoforms are present, which is also what we see in the immunohistochemistry images. For those images, we do not know which breast cancer subtype it is, and therefore we do not know which isoform is expressed more, but we could conclude that both isoforms are expressed, which is the case with all the breast cancer subtypes we tested. In addition, these expriments could be repeated with have normal breast cells (MCF10A), which we did not have in the lab, as they would allow us to also compare to the normal breast tissue.

We next wanted to do a quick screen for conditions that might alter the relative abundance of the two splicing isoforms. We used RNA from various experiments done previously by other students in the lab. As such, these conditions are by no means a comprehensive list of what can be done. As the purpose was to simply see if the isoforms abundance is changeable, we relied on this small set of conditions. This is because if there is a condition that changes the pattern, then the alternative splicing is indeed regulated, but if no conditions can change the alternative splicing whatsoever, then there is no regulation, and because we know there is regulation, there has to be a condition that switches the splicing. However, even if we did not see change with this small set of conditions, we cannot make the statement that the alternative splicing is not regulated as we would need to test more conditions. The conditions tested here were knocking down different genes, adding inhibitors, ultraviolet light and cycloheximide. As shown in figure 6, there are four conditions that change the classic alternative splicing in MDAMB231, which is favoring the isoform including exon 2. Here, MAPK14 knockdown, PARP1 knockdown, PTEN inhibition, and the addition of cycloheximide, which inhibits translation, result in the alternative splicing pattern of hnRNPA2B1 in MDAMB231 to be changed to favor the isoform skipping exon 2. These patterns open the way for more questions and research about these specific genes and their pathways, and why it is those specific genes that change the alternative splicing pattern of hnRNPA2B1, but here it can be concluded that there are at least 4 conditions, out of 8 studied, that change the alternative splicing pattern of hnRNPA2B1, suggesting that it is a highly regulatable splicing event.

One of the known target for hnRNPA2B1 is COX16[8] as it has been previously shown that hnRNPA2B1 binds to COX16 pre-mRNA and regulates its splicing, which is a nuclear function for hnRNPA2B1. *COX16* encodes a protein that is needed for the assembly of cytochrome c oxidase on the mitochondrial membrane. Thus, we used COX16 to test whether the differential expression of hnRNPA2B1 isoforms can indeed influence its nuclear vs. cytoplasmic functions. Figure 7A shows that COX16 has a consistent splicing and expression pattern with all the cell lines. The quantification of the bands, normalized to the loading control (Figure 7C), shows that all the cell lines have quite consistent expression levels, and that there are no clear differences between the different cell lines. We conclude that

COX16 expression is not a good reporter of hnRNPA2B1 activities as no differences were shown. Therefore, we ned to study more known targets for hnRNPA2B1 and possibly repeat this experiment.

To determine whether hnRNPA2B1 splicing isoforms can indeed shift the nuclear/cytoplasmic balance of the translated protein in breast cancer cells, we opted to perform an immunofluorescence assay. This is because we know that the alternatively spliced exon encodes the NLS, and so one isoform is able to enter the nucleus, while the other is not. Figure 8 presents the immunofluorescence images done on MDAMB231 cells after permeabilization and fixation. The secondary antibody was tagged with a green fluorophore, and so we viewed the images under the GFP setting, and under DAPI, which was added to stain the nuclei. As discussed above, for MDAMB231, the favored isoform is the one including exon 2, so the protein should be seen in the nucleus for hnRNPA2B1's condition. For phalloidin's condition, we should see Filament-actin, which is always present, while for the PMZ buffer condition, we should not see anything, and the secondary antibody condition shows non-specific binding. Unfortunately, we see that there is autofluorescence in the images. This is because for the PMZ buffer condition in Figure 8, the secondary antibody was not added, and so we expected no fluorescence, but we observed fluorescence in the absence of incubation with secondary antibody alone. Therefore, if there is fluorescence when it is unexpected, it is hard to conclude if the results for hnRNPA2B1 are reliable, or if they are also autofluorescence. This assay needs to be optimized to get rid of the autofluorescence. The autoflurescence could be caused by the dilution of the antibodies, which would be the first thing to optimize, then we can also optimize the amount of PFA+Triton added to permeabilize and fixate the cells, in addition to the amount of DAPI added, and the step at which the DAPI is added to stain the nuclei.

The second part of this project focused on the RBPs that are associated with each hnRNP, which were narrowed down to three per hnRNP. We overexpressed five of those RBPs, namely hnNRPK, hnRNPM, hnRNPU, DDX3X and PRPF8, in MDAMB231, MDAMB468, and MCF7 cells, followed by RNA extraction and RT-PCR for hnRNPA2B1 (Figure 9). The empty pcDNA plasmid is used as the control for the transfection, and the rest of the transfected plasmids that have the various RBPs cloned into them are compared to pcDNA. For MDAMB231 (Figures 9A and 9B), the results show the expected alternative splicing pattern in the pcDNA transfected cells, where the isoform including exon 2 is favored. This is what's observed in figure 5 under normal conditions, where MDAMB231 makes more of the isoform including exon 2 at a 3:2 ratio (including exon 2: skipping exon 2), while in figure 9, after the transfection with pcDNA, the cells now make more of the isoform including exon 2 at a 9:1 ratio, which is quite different. So, the effects of the transfection make it hard to make any conclusions for MDAMB231, especially that it was the first transfection done, before optimizing the protocol. But, even if the transfection worked and the RBPs were overexpressed, we did not observe a significant change in the splicing pattern following the overexpression of any of the RBPs. The same could be said for MDAMB468 (Figures 9C and 9D). Or, the effects of the transfection make it hard to make any conclusions for both the triple negative cell lines (MDAMB231 and MDAMB468), and therefore, the transfection needs to be repeated. For MCF7 (Figures 9E and 9F), the normal pattern, according to figure 5, is that the isoform skipping exon 2 is the favored one, but we see that the pcDNA (control) switches this trend to have more of the isoform including exon 2, and therefore it might be that the transfection is what switched the patterns. However, PRPF8 is the only RBP that had a noticeable difference from the rest of the RBPs as it favored skipping exon 2 in a high percent. From Figure 4, we see that PRPF8 binds at the junction between exon 2 and the intron, and therefore it could be inhibiting the splicing, which leads to a high percentage of the isoform skipping exon 2. The location where PRPF8 binds makes it a perfect candidate to regulate hnRNPA2B1's splicing. However, this needs to be repeated again, and in all the cell lines, before a conclusion can be made, especially that we are unsure whether or not the RBPs were actually overexpressed. We extracted proteins from the cells, in addition to the RNA, and so western blots can be done to ensure that the RBPs were indeed overexpressed. But, we can say that PRPF8 possibly regulates hnRNPA2B1's splicing, and this could be followed up on.

Our experimental approach for hnRNPM was very similar to that of hnRNPA2B1. Here is a brief discussion of hnRNPM data. First, RT-PCR of exon 6 of hnRNPM confirmed that two splicing isoforms are present, and indeed they differ in their relative expression levels based on the different cell lines (Figure 10). From the gel image and the quantifications, we show that the two triple negative cells, in addition to T47D, make more of the isoform skipping most of exon 6, and so excluding the RRM, suggesting that the protein-RNA interactions are altered, and hnRNPM could not bind, or binds a different set of targets. From all the 6 cell lines studied, only MCF7 makes more of the isoform including exon 6, which includes the RRM as the non-breast cancer cell lines have the same pattern as MDAMB231, MDAMB468, and T47D. It could be that when the RRM is missing for the triple negative cells and T47D, hnRNPM binds a different set of targets. However, the lack of a clear correlation to triple negative cells makes it hard to conclude that any of these isoforms is associated with an aggressive phenotype.

We next did the quick screen for conditions that might alter the relative abundance of the two splicing isoforms. Figure 11 presents the RT-PCR for hnRNPM's exon 6 inclusion under the same conditions used for hnRNPA2B1 to check whether any of conditions change the splicing pattern. For hnRNPM, we see that only one condition changes or reverses the alternative splicing pattern, which is the MCF7 PTEN knockdown. This is because MCF7, under normal conditions, as Figure 10B shows, favors the isoform including exon 6, but here, with the PTEN knockdown, the pattern is reversed. Therefore, we identified one condition for hnRNPM that reverses the alternative splicing, and we need to find more. This suggests that hnRNPM splicing is regulated in a PTEN-dependent manner. As PTEN is known to regulate the PI3K/AKT pathway, we speculate that this pathway might signal to regulate splicing of hnRNPM. This is an exciting and unexpected finding that needs to be confirmed and followed up on in the future.

The known target we opted to use to test if the splicing isoforms of hnRNPM correspond to a change in its protein activity is CD44. It has been shown that the inclusion of different variable variable exons in CD44 depends in part on hnRNPM activity[20]. RT-PCR of CD44 shows that in all the cell lines, the bottom band, which only includes the second variable exon, is present, while the top band, which includes the isoform from the second variable exon to the sixth, is unique to MDMAB468 (Figure 12). When the alternative splicing patterns of CD44 and hnRNPM are compared, we see that there is no clear correlation. For hnRNPM, as Figure 10B shows, MDAMB231, MDAMB468 and T47D all have the same pattern and favor the same isoform, whereas only MCF7 is the different one. But, for CD44, only MDAMB468 has a different pattern, and so we conclude that CD44 splicing is more complex than simply relying on hnRNPM and is not a good target to study, and we need to look into more targets.

Similar to our approach for hnRNPA2B1, we narrowed down potential upstream regulators to three. We overexpressed the same five RBPs from hnRNPA2B1 in MDAMB231, MDAMB468 and MCF7, followed by RNA extraction and RT-PCR for hnRNPM isoforms. The results are shown in Figure 13. In MDAMB231, the favored isoform for hnRNPM is the one skipping exon 6, but we see that this pattern is reversed for all the transfected samples (Figure 13B), including the pcDNA, which is the control. Therefore, we cannot conclude here that the RBPs being overexpressed is what changed the pattern for hnRNPM in MDAMB231, but it might be the transfection itself that reversed the alternative splicing patterns, as even the control with pcDNA changed, and so this needs to be repeated in MDAMB231. For MDAMB468 (Figures 13C and 13D), there is an obvious difference between these samples and the normal conditions. Under normal conditions, MDAMB468 favors the isoform skipping exon 6, but here with the overexpression of all the samples, including the pcDNA, the isoform favored, or the only isoform present in some cases, is the one including exon 6. This is a very interesting finding, but it needs to be repeated before conclusions are made as the transfection could have changed the patterns here, too. For MCF7 (Figures 13E and 13F), the isoform favored is indeed the one including exon 6, as the pcDNA shows, and the rest of the RBPs all show mostly the isoform including exon 6. The only RBP in figure 13E that includes both isoforms in an almost equal ratio is hnRNPU. When this is contrasted to MDAMB468, we see that in MDAMB468, only the isoform including exon 6 is included, but in MCF7, both isoforms are included in an almost equal ratio. Again, these are early findings, and need to be repeated, but we could possibly conclude here that hnRNPU regulates the splicing patterns of hnRNPM.

hnRNPU was not initially in the shortlisted list of RBPs that regulate hnRNPM as stated in the analysis, but this is because we only looked at the RBPs binding exon 6 of hnRNPM (Figure 4B), but once we looked into this again, we found that hnRNPU binds the preceding intron on hnRNPM, which could also regulate the splicing of the next exon. Therefore, hnRNPU might be a good regulator of hnRNPM that we can follow up on.

Conclusions and Future Work

In conclusion, we did find breast cancer subtype-specific alternative splicing patterns for both hnRNPA2B1 and hnRNPM, and we conclude that there are different patterns depending on the status of hormone receptors of the breast cancer cell line. We also concluded that the alternative splicing patterns of each hnRNP are indeed regulated as there are conditions that change those patterns. Finally, we concluded that there is at least one RBP per hnRNP that could regulate the splicing of each hnRNP.

The findings from this project have paved the way for more research to be done on hnRNPs. First and foremost, some of these experiments were done once or twice, and so they need to be repeated for statistical significance. We saw that the RT-PCR for the conditions, especially for hnRNPA2B1's exon 2 inclusion under different conditions, gave us four conditions that switched the splicing, and therefore following up on that, and studying why those specific conditions, and why the knockdown of those genes specifically alters the splicing could be useful. In other words, we can find the pathways for those genes to further understand why they changed hnRNPA2B1's splicing, in addition to finding more conditions that could change hnRNPM's alternative splicing as well, and following up on that. One of the more important future work is cloning and overexpressing the two different isoforms of the hnRNPs, one including the alternatively spliced region, and one lacking it in breast cancer cells. Once we overexpress those 2 isoforms in breast cancer cells, we can do several phenotypic assays to test how differently breast cancer cells act in the presence of the different isoforms. Examples of those assays can be cell proliferation, wound healing, cell cycle, and markers of epithelial–mesenchymal transition (EMT). Another important assay that we were unable to get reliable data from in this project is the immunofluorescence, and therefore some time needs to be spent optimizing this assay till good images that actually represent hnRNPA2B1's localization can be obtained, and then the assay can be done for different breast cancer cell lines. For example, figure 5 showed us that MDAMB468 and MCF7 have an opposing difference when it comes to which isoform is expressed, and therefore in MDAMB468, we expect most of hnRNPA2B1 to be in the nucleus, while for MCF7, we expect it to be in the cytoplasm. Therefore, optimizing the immunofluorescence assay and doing it for those two cell lines will allow us to conclude, with confidence, that indeed the differences in isoforms correlate with a difference in localization. Finally, another area where lots of research can be done is for the second part of the project associated with the RBPs for each hnRNP. Here, we saw that there is at least 1 RBP per hnRNP that regulates the hnRNP, namely PRPF8 for hnRNPA2B1 and hnRNPU for hnRNPM, and therefore this first needs to be repeated and confirmed, then we can follow up on these RBPs, and study them more. We can do this by knocking the RBPs down, using siRNA, to also monitor the effects on the cells and the splicing changes on the hnRNPs when the RBPs' expression is diminished, and compare the results to when they are overexpressed. We can propose blocking the splicing enhancer/ silencer sequences on those RBPs to have a different therapeutic approach, in addition to widening our list of RBPs to try to find more RBPs that regulate the hnRNPs.

Acknowledgments

I would like to express my thanks to Professor Younis for his mentorship, advice, help, and support, Rim El-Ghandour for her immense help and assistance in the lab, Reem Al-Sayed for helping me start the project, and Muhammad Nahin Khan for providing me with the necessary data from the UCSC genome browser to study the RBPs. I would also like to thank my Thesis Advisory Panel, Professor Rule and Professor Affara, for their important help and guidance. Finally, I would like to thank Professor Bouaouina and Weilin Li for their great help with the immunofluorescence assay.

References

(1) Glisovic, T., Bachorik, J., Yong, J., Dreyfuss, G. (2008). 'RNA-binding proteins and post-transcriptional gene regulation.' *HHMI*. DOI: 18; 582(14): 1977-1986. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2858862/

(2) Geuenes, T., Bouhy, D., Timmerman, V. (2016). 'The hnRNP family: insights into their role in health and disease.' *Crossmark*. DOI: 10.1007/s00439-016-1683-5. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4947485/

(3) El Marabti, E., Younis, I. (2018). 'The Cancer Spliceome: Reprogramming of Alternative Splicing in Cancer.' *Frontiers in Molecular Biosciences*. DOI: 10.3389/fmolb.2018.00080. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6137424/

(4) Shah, R., Rosso, K., Nathanson, S. (2014). 'Pathogenesis, prevention, diagnosis and treatment of breast cancer.' *World Journal of Clinical Oncology*. DOI: 10.5306/wjco.v5.i3.283. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4127601/

(5) 'Breast Cancer Awareness Campaign.' n.d. Retrieved from https://www.hamad.qa/EN/your%20health/bcac/Pages/default.aspx

(6) Holliday, D., Speirs, V. (2011). 'Choosing the right cell line for breast cancer research.' *PMC*. doi: 10.1186/bcr2889. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3236329/

(7) Munro, T., Magee, R., Kidd, G., Carson, J., Barbarese, E., Smith, L., Smith, R. (1999). 'Mutational analysis of a heterogeneous nuclear ribonucleoprotein A2 response element for RNA trafficking.' *The Journal of Biological Chemistry*. DOI: 10.1074/jbc.274.48.34389. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/10567417

(8) Alarcón, A., Goodarzi, H., Lee, H., Liu, X., Tavazoie, S., Tavazoie, S. (2015). 'HNRNPA2B1 Is a Mediator of m(6)A-Dependent Nuclear RNA Processing Events.' *Cell.* DOI: 10.1016/j.cell.2015.08.011. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4673968/

(9) Villarroya-Beltri, C., Gutiérrez-Vázquez, C., Sánchez-Cabo, F., Pérez-Hernández, D., Vázquez, J., Martin-Cofreces, N., Martinez-Herrera, DJ., Pascual-Montano, A., Mittelbrunn, M., Sánchez-Madrid, F. (2013). 'Sumoylated hnRNPA2B1 controls the sorting of miRNAs into exosomes through binding to specific motifs.' *Nature Communications*. DOI: 10.1038/ncomms3980. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/24356509

(10) Wang, L., Wen, M., Cao, X. (2019). 'Nuclear hnRNPA2B1 initiates and amplifies the innate immune response to DNA viruses.' *Science.* DOI: 10.1126/science.aav0758. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/31320558

(11) Zhang, F., Cao, J., Xie, H., Yang, L., Shao, Z., Li, D. (2018). 'Cancer-Associated MORC2-Mutant M276I Regulates an hnRNPM-Mediated CD44 Splicing Switch to Promote Invasion and Metastasis in Triple-Negative Breast Cancer.' *Cancer Research.* DOI: 10.1158/0008-5472.CAN-17-1394. Retrieved from https://cancerres.aacrjournals.org/content/78/20/5780.long

(12) Llères, D., Denegri, M., Biggiogera, M., Ajuh, P., Lamond, A. (2010). 'Direct interaction between hnRNP-M and CDC5L/PLRG1 proteins affects alternative splice site choice.' *EMBO Reports.* DOI: 10.1038/embor.2010.64. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2892320/

(13) West, K., Scott, H., Torres-Odio, S., West, A., Patrick, K., Watson, R. (2019). 'The Splicing Factor hnRNP M Is a Critical Regulator of Innate Immune Gene Expression in Macrophages.' *Cell*. 29, 1594-1609. Retrieved from https://www.cell.com/cell-reports/pdf/S2211-1247(19)31282-3.pdf

(14) CAB012403 / hnRNPA2B1/ 2773 from v20.1. *Human Protein Atlas.* Retrieved from https://www.proteinatlas.org/ENSG00000122566-HNRNPA2B1/tissue/breast#img

(15) CAB012403 / hnRNPA2B1 / 2428 from v20.1. *Human Protein Atlas.* Retrieved from https://www.proteinatlas.org/ENSG00000122566-HNRNPA2B1/pathology/breast+cancer#img

(16) CAB010894 / hnRNPA1 / 2773 from v20.1. *Human Protein Atlas*. Retrieved from https://www.proteinatlas.org/ENSG00000135486-HNRNPA1/tissue/breast#img
(17) HPA001609 / hnRNPA1 / 1874 from v20.1. *Human Protein Atlas*. Retrieved from https://www.proteinatlas.org/ENSG00000135486-HNRNPA1/pathology/breast+cancer#img
(18) HPA024344 / hnRNPM / 3544 from v20.1. *Human Protein Atlas*. Retrieved from https://www.proteinatlas.org/ENSG00000099783-HNRNPM/tissue/breast#img
(19) CAB016113 / hnRNPM / 1874 from v20.1. *Human Protein Atlas*. Retrieved from https://www.proteinatlas.org/ENSG00000099783-HNRNPM/pathology/breast+cancer#img
(20) Xu, Y., Gao, X., Lee, J., Huang, H., Tan, H., Ahn, J., Reinke, L., Peter, M., Feng, Y., Siziopikou, K., Peng, J., Xiao, X., Cheng, C. (2014). 'Cell type-restricted activity of hnRNPM promotes breast cancer metastasis via regulating alternative splicing.' *Genes & Development*. DOI: 10.1101/gad.241968.114. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4052765/