USE OF AN INTRONIC POLYADENYLATION SITE IN BREAST CANCERS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

GÖZDE KÖKSAL BIÇAKCI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
MOLECULAR BIOLOGY AND GENETICS

JANUARY 2020

Approval of the thesis:

USE OF AN INTRONIC POLYADENYLATION SITE IN BREAST CANCERS

submitted by **GÖZDE KÖKSAL BIÇAKCI** in partial fulfillment of the requirements for the degree of **Master of Science in Molecular Biology and Genetics Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Ayşe Gül Gözen
Head of Department, **Biology** _____

Prof. Dr. A. Elif Erson Bensan
Supervisor, **Biology, METU** _____

Examining Committee Members:

Prof. Dr. Tolga Can
Computer Engineering, METU _____

Prof. Dr. A. Elif Erson Bensan
Biology, METU _____

Assoc. Prof. Dr. Urartu Özgür Şafak Şeker
UNAM, Bilkent University _____

Date: 14.01.2020

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Gözde Köksal Bıçakcı

Signature:

ABSTRACT

USE OF AN INTRONIC POLYADENYLATION SITE IN BREAST CANCERS

Köksal Bıçakcı, Gözde
Master of Science, Molecular Biology and Genetics
Supervisor: Prof. Dr. A. Elif Erson Bensan

January 2020, 92 pages

Alternative polyadenylation is an important mRNA processing step in which the length and/or the sequence of 3'UTR of mRNAs are altered. This change in 3'UTRs may have significant effects on mRNA stability, localization and translational efficiency. One of the APA types, called intronic polyadenylation (IPA), occurs on intronic sites rather than 3'UTRs. Recent evidence shows that almost 20% of human genes have a poly(A) site in at least one intron. In fact, some introns even may have more than one poly(A) site. Since generation of IPA isoforms could potentially have serious consequences on both protein levels and function, the significance of intronic APA and its occurrence is of interest. Here, we describe such an IPA site in a eukaryotic initiation factor 2B-3 (*EIF2B3*). Using in silico, and in vitro tools, we studied this IPA site usage in response to estrogen in ER+ breast cancer cells. We confirmed expression of *EIF2B3* IPA isoform in MCF7, ER+ breast cancer cells, as well as other breast cancer cell lines and patient samples. We confirmed the usage of this intronic poly(A) site by 3'RACE, cloning and sequencing. Moreover, we showed the expression of a protein from this IPA isoform. Overall, we provide confirmatory data on an in-silico prediction of an IPA case in breast cancers. Future experiments will show the consequence of this IPA event and provide insight into similar IPA events that may happen in cancer cells.

Keywords: Intronic Polyadenylation, EIF2B3, Alternative Polyadenylation, Breast Cancer

ÖZ

BİR İNTRONİK POLİADENİLASYON BÖLGESİNİN MEME KANSERLERİNDE KULLANIMI

Köksal Bıçakcı, Gözde
Yüksek Lisans, Moleküler Biyoloji ve Genetik
Tez Danışmanı: Prof. Dr. A. Elif Erson Bensan

Ocak 2020, 92 sayfa

Alternatif poliadenilasyon 3'UTR uzunluğunu ve/veya sekansını etkileyen önemli bir mRNA işleme mekanizmasıdır. 3'UTR'deki bu değişimler mRNA stabilitesi, lokasyonu ve translasyonel verimliliği üzerinde önemli ölçüde etkilidir. Alternatif poliadenilasyon tiplerinden biri olan intronik poliadenilasyon (İPA), 3'UTR yerine intronik poli(A) bölgelerinde gerçekleşmektedir. Son bulgular, insan genlerinin yaklaşık 20%'sinin en az bir intronda poli(A) bölgesinin bulunduğunu, hatta bazı intronların birden çok poli(A) bölgesi içerdiğini göstermiştir. İntronik poliadenilasyonun üzerine yoğunlaşılmasının başlıca nedeni, farklı İPA izoformlarının oluşmasının hem protein ekspresyonunu hem de fonksiyonunu etkileyebilecek potansiyelinin olmasıdır. Bu araştırmada, östrojen reseptör pozitif (ER+) meme kanseri hücrelerinde, *in vitro* ve *in siliko* yöntemlerle *EIF2B3* geninin intronik poli(A) bölgesinin östrojene bağlı kullanımı üzerine çalışılmıştır. Söz konusu intronik izoformun ekspresyonu MCF7 ER+ meme kanseri hücresinde, diğer meme kanseri hücrelerinde ve hasta örneklerinde gösterilmiştir. Ayrıca, bu intronik poli(A) bölgesinin kullanımı 3'RACE, klonlama ve sekanslama çalışmaları ile de desteklenmiş olup, *EIF2B3* intronik izoformun bir protein ifadesinin olduğu da ortaya çıkarılmıştır. Sonuç olarak bu çalışmada, *EIF2B3* geninin intronik poli(A) bölgesinin meme kanseri hücrelerinde kullanıldığı *in siliko* olarak öngörülmüş ve bunu

doğrulayan *in vitro* bulgular da sunulmuştur. İleride yapılması planlanan deneyler de, söz konusu genin intronik poli(A) bölgesi seçiliminin hücredeki sonuçlarını gösterecek ve kanser hücrelerindeki buna benzer intronik poliadenilasyon olaylarının kavranmasına ışık tutacaktır.

Anahtar Kelimeler: İntronik Poliadenilasyon, Alternatif Poliadenilasyon, EIF2B3, Meme Kanseri

To my beloved family,

ACKNOWLEDGEMENTS

Firstly, I would like to thank and express my sincere gratitude to my supervisor Prof. Dr. A. Elif Erson-Bensan for her peerless teaching, motivation and for all the guidance that she provided throughout my study. Her feedbacks, comments and inspiring encouragement made it possible to finalize this thesis successfully.

Besides, I would like to express my gratitute to Prof. Dr. Tolga Can for his precious guidance in bioinformatics. Also, I would like to thank my thesis committee member, Assoc. Prof. Dr. Urartu Özgür Şafak Şeker for his valuable comments.

I am grateful to all my fellow lab mates; Merve Öyken, Ayça Çırçır Hatıl, Harun Cingöz, İbrahim Özgül, Murat Erdem, Didem Naz Döken, Irmak Gürcüoğlu, Esra Çiçek, Mustafa Çiçek and Melda Ercan for all stimulating discussions, feedbacks, and for all the fun we have had in the last three years. Special thanks to my Damla, who is the best lab partner ever, for her precious friendship and invaluable support all the time.

I am mostly thankful to my beloved husband, Ozan Bıçakcı, for his constant support, endless patience and encouragement throughout the whole process from very beginning till the end. I am also grateful to my precious dog, Lily, for all the barks and love. Special thanks to Serap Demirel, for being the sister I never had.

Lastly, I wish to express my deepest gratitude to my father, Tekin Köksal, and my mother, Neslihan Köksal whose support, love and guidance were with me all the time.

TABLE OF CONTENTS

ABSTRACT v

ÖZ vii

ACKNOWLEDGEMENTS x

TABLE OF CONTENTS xi

LIST OF TABLES xiv

LIST OF FIGURES xv

LIST OF ABBREVIATIONS xviii

ABBREVIATIONS xviii

LIST OF SYMBOLS xix

SYMBOLS xix

CHAPTERS

1. INTRODUCTION 1

1.1. Alternative Polyadenylation 1

1.1.1. 3'-end formation of mRNAs 1

1.1.2. Polyadenylation Mechanism 1

1.1.3. Types of Polyadenylation 3

1.1.3.1. Consequences of UTR-APA 5

1.1.3.2. Consequences of CR-APA 7

1.1.4. APA Regulating Mechanisms 10

1.2. Eukaryotic Initiation Factors 13

1.2.1. Role of Eukaryotic Initiation Factors in Translation Initiation 13

1.2.2. The Role of eIF2B in Translation Initiation 14

1.2.3. Structure of eIF2B 15
1.3. Aim of Study 17
2. MATERIALS AND METHODS 19
2.1. Databases 19
2.2. RNA-Seq Datasets 19
2.3. Probe Screening on Affymetrix 19
2.4. APADetect and SAM 20
2.5. Analysis of RNA-Sequencing Datasets 20
2.6. Cell Lines and Cell Culture 21
2.7. E2 Treatment 22
2.8. Actinomycin D Treatment 22
2.9. RNA Isolation and DNase Treatment 22
2.10. cDNA Synthesis 24
2.11. RT-qPCR 25
2.12. 3'RACE 26
2.13. Cloning of 3'RACE products to pGEM-T Vectors 28
2.14. Cloning of 3'UTR Isoforms into pMIR-Report Luciferase Vectors 28
2.15. Dual-Luciferase Assay 29
2.16. Transfections 29
2.17. Cloning of Coding Sequences into pCDNA 3.1 (-) Vectors 30
2.18. Protein Isolation 30
2.19. Western Blotting 31
3. RESULTS AND DISCUSSION 33
3.1. *EIF2B3* and Probe Level Screen 33

3.2. *EIF2B3* Gene Structure and *EIF2B3* Transcript Isoforms 36
3.3. Experimental Confirmation of *EIF2B3* in E2 Treated MCF7 Samples 39
3.3.1. RT-qPCR Quantification 40
3.3.2. 3' Rapid Amplification of cDNA Ends (3' RACE).....41
3.4. Cloning of 3'UTR Isoforms into Reporter Plasmid 48
3.5. Dual-Luciferase Assay 50
3.6. Measuring the mRNA Stabilities 51
3.7. Detection of Protein Expressions 52
3.8. Expression of Intronic *EIF2B3* in Breast Cell Lines.....54
3.9. Expression of the Intronic Isoform in Different Tissues and Patient Samples55
4. CONCLUSION.....63
REFERENCES.....65
APPENDICES 75

LIST OF TABLES

TABLES

Table 2.1. List of primers used in the experiments..... 23
Table 2.2. Amounts of components for the synthesis of cDNAs 25
Table 2.3. Expected product sizes for the first round of 3'RACE..... 27
Table 2.4. Expected product sizes for the second round of 3'RACE 27
Table 3.1. APA candidates found by the comparison of control and 6 hours of 100 nM Estrogen treated MCF7 Cells by SAM. 34
Table 3.2. APA candidates found by the comparison of control, and 12 hours of 100 nM Estrogen treated MCF7 Cells by SAM. 34
Table 3.3. Ratios of the means of proximal and distal probe intensities for treated and control samples. 36

LIST OF FIGURES

FIGURES

Figure 1.1. Sequence elements and the poly(A) machinery.....2
Figure 1.2. Types of polyadenylation 4
Figure 1.3. 3'UTR-APA.....5
Figure 1.4. CR-APA.....8
Figure 1.5. Domain differences between full-legth and IpA Isoform of IgM.....9
Figure 1.6. Sequestering of intronic poly(A) site by the U1 snRNP 12
Figure 1.7. Translation Initiation Pathway.....14
Figure 1.8. The relationship between eIF2B and the ternary complex formation. 15
Figure 3.1. Gene schema for *EIF2B3* and the positions of the probes.....35
Figure 3.2. Transcript variants of EIF2B3 gene.....37
Figure 3.3. Alignment of the amino acids encoded by different isoforms of EIF2B3 gene. 37
Figure 3.4. Protein structure predictions of two isoforms (SWISS-MODEL).....38
Figure 3.5. Conserved domains of the first isoform. 38
Figure 3.6. RT-qPCR Showing the TFF1 Expression on E2/EtOH Treated MCF7 cDNAs.....40
Figure 3.7. Relative Quantification of EIF2B3 intronic and long isoforms in E2 and EtOH Treated MCF7 Cells. 41
Figure 3.8. Target sequence locations of F1&F2 primers and predicted product sizes. 42
Figure 3.9. First-round 3'RACE PCR products on an agarose gel. 43
Figure 3.10. Second round 3'RACE PCR products on agarose gel validate the usage of intronic poly(A) site.....44
Figure 3.11. Colony PCR verifying the presence of pGEMT-3'UTR insertion. 44

Figure 3.12. Colony PCR verifying the presence of pGEMT- intronic 3'UTR insertion. 45
Figure 3.13.Plasmid sequencing result confirming the usage of intronic poly(A) site. 46
Figure 3.14. Plasmid sequencing result confirming the usage of Hs. 533549.1.4. poly(A) site. 47
Figure 3.15. Cloning of Different 3'UTR Isoforms. 49
Figure 3.16. Colony PCR for the validation of 3'UTR isoforms in pMIR vector..... 49
Figure 3.17. Colony PCR for the validation of 3'UTR isoforms in pMIR vector..... 50
Figure 3.18. Dual-luciferase reporter assay results. 51
Figure 3.19. Actinomycin D treatment showing mRNA Stability differences between intronic and long isoform..... 52
Figure 3.20. Western Blotting for protein products of intronic and full transcripts.. 53
Figure 3.21. RT-qPCR showing the intronic transcript expression on different cell lines..... 54
Figure 3.22. Expression Profile for Both Full-Sized and Intronic Transcripts in Different Tissues..... 55
Figure 3.23. IGV Snapshots for ER+ Breast Cancer Cells Taken from the Patients. 57
Figure 3.24. IGV Snapshots for ER+ and HER2+ Breast Cancer Cells Taken from the Patients..... 59
Figure 3.25. IGV Snapshots for HER2+ Breast Cancer Cells Taken from the Patients. 60
Figure 3.26. IGV Snapshots for TNBC Cells Taken from the Patients..... 61
Figure 0.1. Map of pCDNA 3.1(-) vector 75
Figure 0.2. Map of pMIR-REPORT^{TM} Luciferase vector 76
Figure 0.3. Map of pGEM-T Easy Vector 77
Figure 0.4. PageRuler Prestained Protein Ladder..... 79
Figure 0.5. GeneRuler^{TM} 100bp Plus DNA Ladder 79
Figure 0.6. PCR with GAPDH primers to check DNA contamination. 85
Figure 0.7. Sequencing result for pMIR-Hs.533549.1.2.-3'UTR vector..... 87

Figure 0.8. Sequencing result for pMIR-Hs.533549.1.4.-3'UTR vector. 88
Figure 0.9. Sequencing result for pMIR-Hs.533549.1.4.-3'UTR vector. 88
Figure 0.10. Sequencing result for pCDNA-Full-HA C terminus vector. 89
Figure 0.11. Sequencing result for pCDNA-Full-HA N terminus vector. 90
Figure 0.12. Sequencing result for pCDNA-intronic-HA C terminus vector. 91
Figure 0.13. Sequencing result for pCDNA-intronic-HA N terminus vector. 92

LIST OF ABBREVIATIONS

ABBREVIATIONS

RT-qPCR: Real Time Quantitative Polymerase Chain Reaction

cDNA: Complementary DNA

Poly(A): Polyadenylation

3'UTR: The Three Prime Untranslated Region

APA: Alternative Polyadenylation

CR-APA: Coding Region Alternative Polyadenylation

UTR-APA: Untranslated Region Alternative Polyadenylation

IPA: Intronic Polyadenylation

PAS: Polyadenylation Signal

RBP: RNA-binding Protein

ScRNA-Seq: Single Cell RNA Sequencing

EST: Expressed Sequence Tags

CTD: Carboxy Terminal Domain

LIST OF SYMBOLS

SYMBOLS

EIF2B3: Eukaryotic Initiation Factor 2B-3(gamma)

EIF2Bγ: Eukaryotic Initiation Factor 2B-gamma(3)

$eIF2B_{cat}$: eIF2B catalytic subunit

$eIF2B_{reg}$: eIF2B regulatory subunit

GAPDH: Glyceraldehyde-3-Phosphate Dehydrogenase

RPLP0: Ribosomal Protein Lateral Stalk Subunit P0

CHAPTER 1

INTRODUCTION

1.1. Alternative Polyadenylation

1.1.1. 3'-end formation of mRNAs

Regulation of the pre-mRNA maturation process is one of the key factors affecting gene expression. In eukaryotes, three major processing events are required to have a mature mRNA, which is ready for the translation process. These major processing events encompass the 5' m^7G capping, splicing and polyadenylation (Proudfoot, Furger, & Dye, 2016). Polyadenylation occurs in two steps in which the pre-mRNA is endonucleolytically cleaved at polyadenylation (poly(A)) site and then the long stretch of adenosine residues namely the "poly(A) tail" is added to the 3' end of the cleaved mRNA by poly(A) polymerase (PAP) (Elkon, Ugalde, & Agami, 2013). These cleavage and polyadenylation processes are coupled reactions, which are together called as "polyadenylation". In eukaryotes, all protein-encoding mRNAs are polyadenylated except for replication-dependent histone genes (Marzluff, W. F., Wagner, E. J., & Duronio, 2008; Proudfoot et al., 2016).

1.1.2. Polyadenylation Mechanism

Polyadenylation is coordinated by cis-elements on the pre-mRNA recognized by the polyadenylation machinery. Cleavage stimulatory factor (CSTF), cleavage and polyadenylation specificity factor (CPSF), cleavage factor Im and IIm (CFIm and CFIIm), poly(A) polymerase (PAP), Poly(A)-binding protein (PABPN1), scaffold protein symplekin and RNA Polymerase II C-Terminal Domain (CTD) are the major factors generating the polyadenylation machinery (Chan, Choi, & Shi, 2011; Tian & Manley, 2016).

For polyadenylation factors to bind, certain cis-elements are required. These cis-elements are present on both upstream and downstream of the polyadenylation site. Highly conserved A[A/U]UAAA hexamer which is the polyadenylation signal (PAS), and less highly conserved UGUA and U-rich elements are presented as upstream elements(USE), while some U- or GU-rich elements are positioned as downstream elements(DSE)(Tian & Manley, 2016). The sequence pattern of USE-AAUAAA-DSE is highly conserved in the eukaryotes (Proudfoot, 2011). The actual cleavage site presents itself between these upstream and downstream elements and generally, a CA sequence is found at the cleavage site's 5'foot (Proudfoot et al., 2016). These core sequence elements and the protein factors can be seen in Figure 1.1.

Figure 1.1. Sequence elements and the poly(A) machinery

Upstream and downstream sequence elements and poly(A) machinery including CPSF, CSTF, PAP, PABPN1, CFIm, CFIIm, Symplekin and PolII are shown in the figure (Tian & Manley, 2016).

CPSF complex makes direct contact with the A[A/U]UAAA polyadenylation signal in a sequence-specific manner through the CPSF30 and WDR33 subunits (Chan et al., 2014). At the same time, CstF complex recognizes the downstream U- and GU- rich

elements. CPSF binding to USE and CstF binding to DSE occurs cooperatively so that one's target binding enhances the efficiency of the other ones. (Colgan & Manley, 1997) The other two elements of the machinery, CF Im and CF IIm are required for the cleavage process, which occurs by the coordinative works of CstFs, CPSFs, and the PAP. (Zhao & Hyman, 1999) After the cleavage event, PAP is needed for the poly(A) tail addition, and its' recruitment to the poly(A) site is made possible by the interactions of PAP and CPSF together with CF Im (Proudfoot et al., 2016). Poly(A) tail formation is started with the addition of adenosine residues by PAP; however, the tail length is controlled and stabilized by PABP (PAB1P) (Amrani et al., 1997). In the end, tail length reaches up to ~300 residues in mammals (Keller et al., 2000). Lastly, as mentioned above, RNA Pol II is also accepted as one of the factors of polyadenylation machinery since there is a dynamic interplay between the transcription termination and the poly(A) site recognition (Proudfoot, 2011). To illustrate, one of the CF IIm complex members, namely Pcf11 is known to be one of the key factors ensuring the termination event (West & Proudfoot, 2008).

1.1.3. Types of Polyadenylation

RNA transcripts may have multiple poly(A) sites, and their differential usage basically forms different mRNA isoforms (Tian & Manley, 2016). This phenomenon is called the "alternative polyadenylation (APA)". APA has long been known as a common mechanism among all eukaryotes, as a matter of fact, almost 70% of human genes have multiple APA sites, and around 50% of them have three or more different sites (Derti et al., 2012). In the beginning, it was thought that the APA is the major marker for only the 3' UTR ends; however, for decades, it is known that APA occurs not only on the 3'UTRs, but also in other parts of the genes in eukaryotes, for instance, the introns and coding sequences (Guo, Spinelli, Liu, Li, & Liang, 2016).

As it can be seen in Figure 1.2., there are four different types of alternative polyadenylation events. The former one is the tandem 3' UTR APA that is the most common one, which is changing the 3'end of the transcript. As a result of distal or

proximal poly(A) site choice, 3'UTR of a transcript may undergo either shortening or lengthening. The second one, alternative terminal exon APA is also called as skipped terminal exon APA, which occurs when an internal exon, bearing a poly(A) site, is selected by splicing mechanism to be a 3'terminal exon (Tian & Manley, 2016). The third one is intronic APA, which occurs when a poly(A) site is located on an upstream intron, which is normally spliced out in the other transcripts, is chosen. In this case, the coding sequence is extended with the inclusion of a part of that intronic sequence due to the inhibition of a 5'splice site (Tian, Pan, & Ju, 2007). And the last one is the internal exon APA, that is also termed as composite terminal exon APA in which the premature polyadenylation at an internal coding exon occurs (Elkon, Ugalde, & Agami, 2013; Tian & Manley, 2016). As mentioned above, the splicing process is very crucial in these different APA events considering that splicing and APA are coupled processes (Proudfoot et al. 2002).

Figure 1.2. Types of polyadenylation

These distinct types includes Tandem 3'UTR APA, Alternative Terminal Exon APA, Intronic APA and Internal Exon APA (Elkon, Ugalde, & Agami, 2013).

Throughout these different APA events, mRNA isoforms with truncated coding sequences or altered 3'UTR lengths can form.

1.1.3.1. Consequences of UTR-APA

As mentioned above, if the APA takes place on the 3'UTR, this might lead to either shortening or lengthening of the 3'UTR (Figure 1.3). As a result of that, stability, localization, and translational efficiency of mRNA could change since those specific sequences on UTR might be used as docking sites for regulatory elements (Elkon et al., 2013b). The length of the 3'UTRs directly affects the binding of both RNA-binding proteins (RBPs) and microRNAs (miRNAs), which are the small 18-22nt RNAs, altering both the stability and the translational efficiency of their targets (Bartel D. P., 2009). Therefore, shortening of 3'UTR might result in loss of cis-elements, on which miRNA and RBPs bind, leading to reduced numbers of protein products.

Figure 1.3. 3'UTR-APA

This figure showing the formation of mRNA isoforms with different 3'UTRs, was adapted from (Hardy & Norbury, 2016).

One of the important factors which are affected by APA is the stability of mRNA. Generally, isoforms having long 3′ UTRs would be less stable than the ones having short 3'UTR since destabilizing elements will most likely have their place to bind on long 3'UTRs. Wiestner et al. (2007) showed that Cyclin D1 gene has major point mutations that bring premature poly(A) signals into existence in mantle cell lymphoma (MCL). In these cells, usage of those premature signals creates Cyclin D1 mRNAs losing most of their 3'UTRs. These short Cyclin D1 transcripts lose their destabilization elements so that they become more and more stable in contrast to their wild-type counterparts. Significantly increased Cyclin D1 mRNA levels are thought to have a major contribution to the strong proliferative abilities of MCL tumors (Wiestner et al., 2007).

The 3'UTR length might affect the localization of mRNAs, as well as the stability. The old but gold example for the APA events affecting the mRNA localization would be the two transcripts of brain-derived neurotrophic factor (BDNF) having distinct 3'UTR lengths. An et al. (2008) revealed that those transcripts with shorter 3'UTRs are basically restricted to the cell body; on the other hand, transcripts having long 3'UTRs are generally found in the dendrites. Indeed, this change in the localization could be exploited by the cells to achieve localized translation events, which results in the enrichment of some proteins at distinct locations (An et al., 2008).

Lastly, changes in translational efficiency is another result of 3'UTR APA. With the help of Transcript Isoforms in Polysomes Sequencing (TrIP-seq), Floor and Doudna (2016) have found that the long mRNA isoforms are globally associated with lower translation rates in comparison to their short counterparts. Moreover, they demonstrated that the regulatory regions on those 3' untranslated regions control the output of translation (Floor & Doudna, 2016). Furthermore, there are many RNA binding proteins that directly affect the translation process by preventing the closed-loop formation and recruitment of ribosome (Szostak & Gebauer, 2013). For example, Cho et al. (2005) have shown that translation of caudal (cad) mRNA is prevented by Bicoid (Bcd) protein, which binds to 3'UTR of cad in *Drosophila* embryo (Cho et al.,

2005). In this study, it has been revealed that Bicoid binding to 3'UTR leads to recruitment of d4EHP (eIF4E-related cap-binding protein) to 5'cap structure and its binding to the 5'cap greatly reduces the affinity for eIF4G causing inhibition of closed-loop structure, and eventually the translation process (Cho et al., 2005). Another example could be the interferon- γ -activated inhibitor of translation (GAIT) complex binding to 3'UTR of ceruloplasmin mRNA causing the translation repression during inflammation. Here, the ribosomal protein L13a, which is one of the four subunits of GAIT complex, associates with eIF4G and prevents its interaction with eIF3 leading to the interference of 43S ribosomal complex recruitment, and translation at the end (Kapasi et al., 2007).

1.1.3.2. Consequences of CR-APA

If the APA event takes place in the coding region, not only the amount but also the functions of the protein products might be affected. As a result of this process, either unfunctional or functional but truncated proteins will be introduced into the proteome (Figure 1.4.). Since the consequences of IPA events will cover those of CR-APA events, the following examples will be based on the outcomes of intronic polyadenylation.

Intronic polyadenylation (IPA) once have been thought as a rare mechanism; however, the evidence coming from both cDNA/EST and genomic sequences have shown that almost 20% of human genes have a poly(A) site in at least one intron, in fact, some introns even may have more than one poly(A) site (Tian et al., 2007). With the advancements in PAS-sequencing methods, the significance of intronic APA and its' occurrence have been of interest (Kamieniarz-Gdula & Proudfoot, 2019).

Figure 1.4. CR-APA

This figure showing the formation of mRNA isoforms with alternating C-terminal sequences and 3'UTRs, was adapted from (Hardy & Norbury, 2016).

In some cases, the end result of an intronic polyadenylation site usage might be the formation of transcripts, which are very unstable and prone to rapid degradation (Kamieniarz-Gdula & Proudfoot, 2019). To illustrate, the mammalian CSTF77 gene (a.k.a CSTF3) is known to have a conserved intronic poly(A) site, and the usage of this PAS basically produces truncated, a non-functional protein that undergoes degradation (Pan et al., 2006). It might be considered as a superfluous event; however, cells might use this route as a feedback mechanism so that when the CSTF77 levels are high, IPA is induced in order to work as a negative feedback loop (Luo et al., 2013).

Intronic polyadenylation would lead to stable transcripts as well as the unstable ones. Furthermore, the proteins encoded by those stable transcripts might have either altered functions or localizations as compared to full-length counterparts. One of the most well-known examples of IPA event that is changing both the localization and the function of the protein is the immunoglobulin M (IgM). When the distal poly(A) site

located on the 3'UTR of the IgM mRNA is chosen, membrane-bound IgM is produced; however, if the poly(A) site choice is shifted towards the intronic site, two terminal exons of the full-sized transcript are not translated so that the transmembrane domain is lost resulting with the generation of soluble IgM (Elkon et al., 2013). The indicated mechanism here is cell-type specific since in mature B cells distal poly(A) site usage is more prevalent while the IPA of IgM occurs mostly in plasma cells (Singh et al., 2018) (Figure 1.5.).

Figure 1.5. Domain differences between full-legth and IpA Isoform of IgM

Cell type specific IpA occurrence is shown in the figure. In plasma cells, transmembrane domain is lost because of IpA, and soluble IgM is produced (Singh et al., 2018).

What's more is that IPA is certainly a common mechanism among the receptor tyrosine kinases (RTKs), having both anchoring transmembrane and kinase domains in their full-length isoforms (Vorlová et al., 2011). Here, the usage of an intronic poly(A) site creates the soluble isoform of RTKs (e.g., EGFR, HER2, VEGFR1/2, and PDGFR-α/-β) which does have neither of the two domains, but still have the ligand-binding ability. By this means, it might serve as a dominant-negative regulator to turn off a signalling pathway (Lemmon & Schlessinger, 2010).

Another striking example would be the role of IPA in evoking tumorigenesis in chronic lymphocytic leukemia (CLL) cells. Studies suggest that IPA in those cells

mostly occurs on the genes having tumour suppressor functions which results with the formation of truncated proteins either lacking those suppressive functions by diminishing the functional ones' expressions (e.g. DICER) or gaining oncogenic functions by acting as a dominant-negative manner (e.g. CARD11 and MGA) (Lee et al., 2018).

Based on these observations, IPA events broaden the diversity of not only the transcriptome but also the proteome.

1.1.4. APA Regulating Mechanisms

There are many factors regulating the APA mechanism and the core polyadenylation factors are the first things that come to mind. That is, when the expressions of core polyadenylation factors are either upregulated or downregulated, this might directly alter the polyadenylation site choice. To illustrate, at the time of B cell maturation, CSTF64 upregulation may lead to increased expression of whole CSTF complex, which results in more proximal PAS usage (Tian & Manley, 2016; Xia et al., 2014). Another good example might be CFI heterodimer, which is composed of CFI25 together with either CFI68 or CFI59. Studies suggest that those CFI subunits might promote distal PAS usage, because of the fact that when one of the CFI25 or CFI68 subunits is knocked-down, proximal PAS choice is remarkably enhanced (Gruber, Martin, Keller, & Zavolan, 2012).

Other than the core polyadenylation machinery proteins, many of the remaining RNA binding proteins (RBPs) also regulate the APA mechanism by either supporting the core proteins with or preventing them from binding to their targets (Erson-Bensan, 2016). One of the most studied examples of RBPs is the embryonic-lethal abnormal visual (ELAV), which is known to be responsible for neuron-specific suppression of proximal poly(A) site usage resulting in the 3'UTR lengthening in Drosophila (Hilgers et al, 2012). On the contrary, cytoplasmic polyadenylation element-binding protein 1 (CPEB1) supports the weak proximal poly(A) signal sites with enhanced CPSF

binding so that the PAS choice shifted towards the proximal site resulting with 3'UTR shortening.

Transcription process itself could also be accounted as one of the regulators of APA. Especially, RNA Polymerase II (Pol II) elongation rate have a voice in the PAS choice owing to the fact that decreased rate of Pol II elongation would enhance the proximal poly(A) site usage by making more time for the polyadenylation machinery (Pinto et al, 2011).

In addition, signalling pathways affecting both proliferation and differentiation of cells are also known to regulate APA process. To illustrate, one of the most essential regulators for cell proliferation, which is mammalian target of rapamycin (mTOR) pathway, is shown to enhance the proximal PAS choice leading to prevailing 3'UTR shortening along the transcriptome (Chang et al., 2015). For instance, a proto-oncogene (ARG2) is shown to undergo 3'UTR shortening through direct induction of mTOR signaling activated by PI3K/AKT pathway (Matoulkova, Sommerova, Pastorek, Vojtesek, & Hrstka, 2017). Another important proliferative signal, namely, estrogen receptor signaling pathway, may induce APA. Our group showed that 17 b-estradiol (E2) induction causes an increase in proximal poly(A) site usage of a proto-oncogene (CDC6), in ER+ breast cancer cells, resulting with the upregulation of a shorter 3'UTR isoform (Akman, Can, & Erson-Bensan, 2012).

In addition to the important regulators mentioned above, splicing factors are known to have regulatory roles in PAS choice. One of the fundamental elements of the spliceosome, U1 small nuclear ribonucleoparticle (U1 snRNP), that is responsible for the recognition of 5' splice site may have a critical role in the APA regulation. U1 snRNP might be responsible for the repression of proximal PAS usage due to the fact that when the U1 snRNP is inhibited, usage of proximal and especially the intronic poly(A) sites are significantly enhanced (Berg et al., 2012). Therefore, U1 snRNP sequestering could be an important regulator for intronic PAS choices which is illustrated in Figure 1.6.

Figure 1.6. Sequestering of intronic poly(A) site by the U1 snRNP

U1 snRNP playing an essential role in repression of intronic pol(A) site is shown above. Inhibition of U1 snRNP results with active intronic PAS (Desterro, Bak-Gordon, & Carmo-Fonseca, 2019).

Lastly, there is a dynamic interplay between splicing and polyadenylation, and this is more understandable when the intronic polyadenylation cases are considered. According to literature, occurrence of IPA in large introns having weak 5' splice sites is more common than the other sites, since more time is needed to splice those large introns out, and polyadenylation machinery would have more time than the spliceosome has (Tian & Manley, 2016; Tian et al., 2007). Therefore, together with the factors mentioned above, the size of the intron and the weakness of the 5' splice sites could also be accounted as the regulators of IPA.

Use of intronic poly(A) sites in cancer cells is just beginning to be understood. Generation of IPA isoforms could potentially alter the efficiency of translation and/or function of the normal isoform in a dominant negative manner (Lemmon & Schlessinger, 2010). Therefore, discovery of such isoforms could have interesting implications. In this thesis, we describe activation of such an IPA site for the *EIF2B3* mRNA that is important in ribosome assembly.

1.2. Eukaryotic Initiation Factors

Eukaryotic initiation factors (eIF1-6) support ribosomal assembly process. Six different eIFs are known to participate in the translation initiation process. Except for the last factor, eIF6, all of these initiation factors are multi-subunit complexes whose expression alterations may result in oncogenic activity (Ali, Ur Rahman, Jia, & Jiang, 2017).

1.2.1. Role of Eukaryotic Initiation Factors in Translation Initiation

The translation initiation process has three major steps, which include the formation of 43S pre-initiation complex, then 48S pre-initiation complex, and lastly, the 80S ribosome. All the way through these steps, protein factors play vital roles. The first initiation factor starting the sequence of events is eIF2, which forms the ternary complex by binding through Met-$tRNA_i$ and work as a GTPase. Only when Met-$tRNA_i$-eIF2-GTP ternary complex is formed, Met-$tRNA_i$ could be delivered to 40S ribosomal subunit, and binding of the ternary complex to ribosomal subunit is promoted by eIF1, eIF1A, and eIF3 (Klann & Dever, 2004; Pavitt, 2005). By the participation of those three factors and eIF5, 43S pre-initiation complex is formed (Asano, 2000). For mRNA to be associated with this pre-initiation complex, eIF4F, which is composed of eIF4E, eIF4G, and eIF4A should be bound to m^7G cap structure of mRNA at 5' end (Richter & Sonenberg, 2005). mRNA binds to 43S pre-initiation complex and forms 48S pre-initiation complex. Through the scanning of mRNA, AUG codon is found and recognized by Met-$tRNA_i$ at P site, which leads to the hydrolysis of GTP on eIF2 by the catalysis of eIF5, that is a GTPase activating protein (GAP). This hydrolysis causes eIF2-GDP and other initiation factors to be released from the ribosome. Discharge of the factors make 40S ribosomal subunit be able to join with 60S subunit by the help of eIF5B and eIF1A, forming the 80S ribosome ready for the elongation (Pavitt, 2005). A simplified illustration could be seen in Figure 1.7.

Figure 1.7. Translation Initiation Pathway

Overview of the translation initiation beginning from the 40S ribosomal subunit, and going through the formation of 80S ribosome is shown above (Klann & Dever, 2004).

1.2.2. The Role of eIF2B in Translation Initiation

eIF2(α-γ) is a heterotrimeric GTPase that forms the ternary complex with Met-tRNA$_i$ when it is in GTP bound state. When the GTP is hydrolysed by eIF5, inactive eIF2-GDP is dissociated from the ribosome. Interaction between the Methionine and eIF2 is basically disrupted when GDP is bound to eIF2, even if the tRNA interaction would be intact (Kapp & Lorsch, 2004). Therefore, in its inactive GDP bound state, eIF2 cannot go through the new rounds of translation initiation unless there is a switch from GTP to GDP (Mohammad-Qureshi, Jennings, & Pavitt, 2008). At this very moment that eIF2B comes into the scene by its guanine nucleotide exchange factor (GEF) activity. eIF2B acts to re-establish the active form of eIF2 by exchanging GDP for

GTP, which is a tightly controlled, rate-limiting step that could be accepted as a key point for regulation of translation, and eventually, the protein expression (Mohammad-Qureshi et al., 2008). The illustration shows the relationship between eIF2B and the ternary complex (Figure 1.8).

Figure 1.8. The relationship between eIF2B and the ternary complex formation.

GEF activity of eIF2B provides the exchange of GDP on eIF2 for GTP, providing the formation of ternary complex leading to translation initiation and eventually, the elongation.

1.2.3. Structure of eIF2B

Eukaryotic initiation factor 2B (eIF2B) is composed of five non-identical subunits (α-Ɛ), which together form ~600 kDa decameric protein that has a much more complex and large structure in comparison with the other Guanine-nucleotide-exchange factors (Bogorad, Lin, & Marintchev, 2017)**.** This large protein complex has two catalytic subcomplexes composed of eIF2Bγ and eIF2Bε ($eIF2B_{cat}$), and a regulatory hexameric subcomplex having eIF2Bα, eIF2Bβ, and eIF2Bδ subunits in dimerized manner ($eIF2B_{reg}$). These catalytic subcomplexes assemble to both sides of the regulatory

subcomplex, which is placed right at the center and together, they form a large decameric eIF2B complex (Hinnebusch & Yang, 1996; Kashiwagi et al., 2016).

As the name implies, the regulatory subcomplex has a major role in the regulation of eIF2B function since eIF2α phosphorylation on Ser 51 makes eIF2(α-P)-GDP has a higher affinity for $eIF2B_{reg}$, and when the phosphorylated version of eIF2 binds to the regulatory complex, it basically locks eIF2B. Here, the phosphorylation of eIF2α is induced by different types of cell stress and resulting phosphorylated eIF2α-GDP act as a competitive inhibitor for eIF2B (Bogorad, Lin, & Marintchev, 2018) The other part of the complex; the catalytic subunit is composed of two dimers of eIF2Bγ-ε. From these subunits, eIF2Bε has a main catalytic part, called $\varepsilon_{cat,}$ at its C-terminal domain on which the GTP exchange occurs. Some studies suggest that GEF activity would be catalysed even if the only subunit present is eIF2Bε *in vitro;* however, the efficiency will be very low without the presence of eIF2Bγ (Gomez, Mohammad, & Pavitt, 2002). On the other hand, Gordiyenko et al. (2014) showed that γ and ε subunits together function to exchange GDP for GTP. Moreover, they suggest that the pyrophosphorylase-like (PL) domain of γ subunit is the main responsible part for GTP binding, while the left-handed b-helix (LbH) domain of the same subunit mostly provides the regulatory complex with binding to the catalytic core (Gordiyenko et al., 2014).

All of these subunits (eIF2Bα-ε) have a vital role in the formation of the proper eIF2B complex and its functioning. In fact, mutations in any one of the genes encoding these five subunits might result in pathologies including leukoencephalopathy, which is called vanishing white matter disease (Bugiani, Boor, Powers, Scheper, & Knaap, 2010).

1.3. Aim of Study

Intronic polyadenylation of mRNAs is a critical mechanism that might significantly affect both protein expression and function. Even though the evidence of intronic polyadenylation is shown in immune cells and leukaemias, it is not clear whether such APA events are seen in breast cancers. Here, we tried to understand whether estradiol (E2), which is a robust proliferative signal in breast cells, induces intronic polyadenylation in ER+ breast cancer cells and the potential effect of such an IPA event.

CHAPTER 2

MATERIALS AND METHODS

2.1. Databases

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly was used in order to find the location of the exons and introns on which the poly(A)sites are placed. Database is available at https://genome.ucsc.edu/.

PolyA_DB 2 source was used for finding the previously identified poly(A) sites on the mRNAs. Web source is available at http://exon.umdnj.edu/polya_db/v2/

GTEx Portal (Analysis Release V8) was used to get the gene expressions of different isoforms on various normal tissues (dbGaP Accession phs000424.v8.p2).

2.2. RNA-Seq Datasets

Gene Expression Omnibus (GEO) is a database containing both array and sequence-based data, was used in order to get the various RNA-sequencing datasets used in this study. The following datasets were analyzed in this thesis:

GSE11324 is the RNA-sequencing dataset for 100 nM E2 treated MCF7 samples at different time-points (0, 3, 6 and 12 hours). This dataset was used in the APADetect analysis.

GSE75688 is the scRNA-sequencing dataset for primary breast cancer. The dataset includes 515 samples from total of 11 patients having distinct subtypes of the breast cancer.

2.3. Probe Screening on Affymetrix

NetAffx™ Analysis Center has been used for the probe set search. As a gene chip array, Human Genome U133 Plus 2.0 Array has been selected and the resulted 54,675

probe sets were filtered by the gene name. The current genomic location of the probe set belongs to *EIF2B3* gene (probe set ID: 218488_at) was visualized on UCSC Genome Browser in order to comprehend the probe positions compared to different polyA site locations.

2.4. APADetect and SAM

APADetect tool is a probe-based analysis tool which was developed by Prof. Dr. Tolga Can. This tool uses available probe sets for each gene on HGU133A (GPL96) and HGU133Plus2(GPL570) platforms and analyzes the mean signal intensities of both proximal and distal probes according to location of the poly(A) site. To do this, Poly(A) site locations are obtained from PolyA_DB2, and total of 11 probes are divided as proximal and distal according to those poly(A) site locations. The mean signal intensities of proximal probes divided by those of distal probes gives the short/long (SLR) ratio.

The dataset with an accession number of GSE11324, was used in APADetect analysis. The output of the APADetect was loaded into Multiple Experiment Viewer (MeV) software for the statistical analysis. One of the statistical analysis tools in the software, Significance Analysis of Microarrays (SAM) was used for the analysis, since it was a statistical method for microarrays in particular. SAM detected statistically significant genes by using t tests for each gene. Sample/control SLR values were undergone log normalization. Resulting significant genes were categorized as positive and negative significance showing shortening and lengthening, respectively.

2.5. Analysis of RNA-Sequencing Datasets

All RNA-Sequencing analyses were performed on The Cancer Genomics Cloud (CGC), powered by Seven Bridges Genomics (available at https://cgc.sbgenomics.com/). Firstly, from the Gene Expression Omnibus database, Sequence Read Archives (SRA) accession number belonging to that specific sample was gathered following the NCBI-SRA links. Then, these SRA accession numbers were given as inputs for SRA Toolkit fastq-dump app on CGC. If the sample is paired,

from the app settings, “Read ids” must be selected as “true”, and “Split reads” must be selected as “split_files”. Otherwise, both sections must be selected as “no_paired_end”. By using this app, sample files having SRA format were converted into FASTQ format in the end. Next, FASTQ files were loaded as input files for Trim Galore app. This app was used for adapter trimming from the sequencing files. Results of Trim Galore tool could be presented as “Trimmed reads” files. These Trimmed reads files having “_val_1.fq” file extension were presented to RNA-seq alignment-STAR 2.5.4b tool as input files. As a reference genome, “human_g1k_v37_decoy.phiX174” was selected. For the gene annotation file, “Homo_sapiens.GRCh37.75.gtf” file was selected from the Public files section. Lastly, from the app settings section, output format was selected as “BAM” file. Then, aligned reads, having bam extension were loaded onto The Integrative Genomics Viewer (IGV) for visualization. IGV is available at https://igv.org/.

2.6. Cell Lines and Cell Culture

MCF7 (Michigan Cancer Foundation-7) cells were cultered in Dulbecco's Modified Eagle's Medium containing High Glucose (4500 mg/L Glucose), 1 mM sodium pyruvate, 1.5 g/L sodium bicarbonate, 2 mM L-glutamine, 1 % penicillin-streptomycin and lastly, 10 % Fetal Bovine Serum (FBS). All of the supplements for the cell culture were obtained from Biowest (Riverside, USA). For elimination of mycoplasma contamination, cells were treated with Plasmocin (Invitrogen, CAT# ant-mmp). Cells were cultured as monolayers and incubated in CO2 incubators having 5% CO2 and 95% humidified air, and arranged to optimum temperature, 37 ºC. For the long-term storage of the cells, cell pellets were taken when they got to 70-80% confluency, and they were resuspended in MCF7 medium with 5% DMSO (dimethyl sulfoxide) (Sigma, Cat# 154938). Then, the cells are preserved in cryovials in a liquid nitrogen.

2.7. E2 Treatment

For the E2 treatment, MCF7 cells were grown up to 50-60% confluency in T-75 cell culture flasks. Next, the cells were washed with PBS to eliminate the growth medium completely. Then, the cells were grown in starvation medium (phenol red-free) including 10% dextran coated charcoal stripped FBS, 1% P/S and 1% L-glutamine for 72 hours. After 72 hours of starvation, MCF7 cells were treated with 100 nM of 17 β-Estradiol (Sigma-Aldrich, CAT#E2257) and 100 nM of Ethanol as vehicle control for 12 hours. At the end of 12 hours of treatment, cell pellets were gathered and used for the RNA isolation. The success of the treatment was checked by looking at the TFF1 expression differences between E2 and EtOH samples.

2.8. Actinomycin D Treatment

For the Actinomycin D treatment, MCF7 cells were grown up to 50-60% confluency in 6-well cell culture plates. Next, the cells were treated with either 2 µg/mL Actinomycin D (Tocris Bioscience, CAT#1229) or DMSO as a vehicle control for 0,2, and 12 hours. This treatment was performed by İbrahim Özgül, in our lab.

2.9. RNA Isolation and DNase Treatment

For total RNA isolations from the cells, High Pure RNA Isolation Kit (Roche, CAT#11828665001) was used. Next, overnight DNase treatment of isolated RNAs were done by using DNase I Recombinant Enzyme (Thermo Fisher Scientific, CAT#EN0521) and by following the manufacturer's instructions. After DNase treatment, DNA contamination was checked by PCR using GAPDH primers whose sequence is given in Table 2.1. After checking the contamination, RNA concentrations were measured by BioDrop Duo (Isogen Life Science). Besides the concentrations, sample purities were also checked by the same device. A260/A280 ratios between 1.8 and 2.0, and A260/A230 ratios around 2.0 were accepted as pure sample.

Table 2.1. *List of primers used in the experiments*

Primer Name	Primer Sequence (5' to 3')	Experiment that is used
GAPDH_F	GGGAGCCAAAAGGGTCATCA	PCR
GAPDH_R	TTTCTAGACGGCAGGTCAGGT	PCR
TFF1_F	CCATGGAGAACAAGGTGATCTGC	PCR
TFF1_R	GTCAATCTGTGTTGTGAGCCGAG	PCR
TFF1_F	TTGTGGTTTTCCTGGTGTCA	RT-qPCR
TFF1_R	CCGAGCTCTGGGACTAATCA	RT-qPCR
RPLP0_F	GGAGAAACTGCTGCCTCATA	RT-qPCR
RPLP0_R	GGAAAAAGGAGGTCTTCTCG	RT-qPCR
EIF2B3 Long_F	GTGATCGAGAAGGGTGCAGAC	RT-qPCR
EIF2B3 Long_R	GTCATTCCCCACGATCACCTC	RT-qPCR
EIF2B3 3'RACE F1	GTCCATTCGTCAGCCCAGAT	3'RACE
EIF2B3 3'RACE F2	GTCAGCCCAGATTGTCAGCA	3'RACE
EIF2B3 intronic_F	GGCCAGAGACACAGATTGGA	RT-qPCR
EIF2B3 intronic_R	ACAGGGGGAAACATACCCTTC	RT-qPCR
EIF2B3 pMIR 3'UTR_Hs.533549.1.2_F	ACGCGTGTTCTGAGCAAGTCAGACTCC	pMIR Cloning
EIF2B3 pMIR 3'UTR_Hs.533549.1.2_R	AAGCTTTTGCTCCAGATGATCTTTACCACAT	pMIR Cloning
EIF2B3 pMIR 3'UTR_Hs.533549.1.3_F	ACGCGTGTTCTGAGCAAGTCAGACTCCTT	pMIR Cloning
EIF2B3 pMIR 3'UTR_Hs.533549.1.3_R	AAGCTTACTGTGTACACCTGGAGCCC	pMIR Cloning
EIF2B3 pMIR 3'UTR_Hs.533549.1.4_R	AAGCTTGCCTTTGGAAGCCCTTCTTT	pMIR Cloning
EIF2B3 pMIR 3'UTR_Hs.533549.1.10_F	ACGCGTTGGAAAAATGCCTTCAAAATTC	pMIR Cloning
EIF2B3 pMIR 3'UTR_Hs.533549.1.10_R	AAGCTTTTTTGATATCTGAAAATGTTTTTATTC	pMIR Cloning
EIF2B3 coding HA_F	CGCATTCTAGAAAAATGTACCCATACGATGTTCCAGATTACGCTGAATTTCAAGCAGTAGTGATG	pCDNA Cloning
EIF2B3 full coding_R	CGCATAAGCTTTCAGATCTCCATGAGCTGGTC	pCDNA Cloning

EIF2B3 coding HA Forward	CGCATTCTAGAAAAATGTACCCATACGATGTTCCAGATTACGCTGAATTTCAAGCAGTAGTGATG	pCDNA Cloning
EIF2B3 coding Forward	CGCATTCTAGAAAAATGGAATTTCAAGCAGTAGTGATG	pCDNA Cloning
EIF2B3 coding intronic HA Reverse	CGCATAAGCTTTTAAGCGTAATCTGGAACATCGTATGGGTATCTTTGCCTCAAGTGGGT	pCDNA Cloning
EIF2B3 coding intronic Reverse	CGCATAAGCTTTTATCTTTGCCTCAAGTGGGT	pCDNA Cloning
EIF2B3 full coding HA Reverse	CGCATAAGCTTTCAAGCGTAATCTGGAACATCGTATGGGTAGATCTCCATGAGCTGGTC	pCDNA Cloning
EIF2B3 full coding Reverse	CGCATAAGCTTTCAGATCTCCATGAGCTGGTC	pCDNA Cloning
RACE_OligodT	GACCACGCGTATCGATGTCGACTTTTTTTTTTTTTTTV	3’RACE
RACE_Anchor_R	GACCACGCGTATCGATGTCGAC	3’RACE
T7	TAATACGACTCACTATAGGG	Sequencing
BGH\pCDNA3.1_R	TAGAAGGCACAGTCGAGG	Sequencing
pMIR Sequencing_F	AGGCGATTAAGTTGGGTA	Sequencing
pMIR Sequencing_R	GGAAAGTCCAAATTGCTC	Sequencing
EIF2B3 sequencing_F	CAGACTTGGATGAAGAGCTGGT	Sequencing

2.10. cDNA Synthesis

For the synthesis of cDNAs, RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, CAT# EP0441) was used. The kit recommendations for the amounts of the components were shown in Table 2.2.

Table 2.2. *Amounts of components for the synthesis of cDNAs*

Components	Amounts
RNA	1 µg
Oligo(dT) primer (100 µM)	1 µL
Nuclease-free water	up to 12 µl
*5 minutes of incubation at 70 °C, after brief spin-down. *Incubate on ice for 1 min.	
5X Reaction Buffer (250 mM Tris-HCl (pH 8.3), 250 mM KCl, 20 mM MgCl2, 50 mM DTT)	4 µl
dNTP Mix (10 mM)	2 µl
RiboLock RNase Inhibitor (20 U/µL)	1 µl
RevertAid Reverse Transcriptase (200 U/µL)	1 µl
*60 minutes of incubation at 42°C, then the reaction is stopped by heating up to 70°C for 5 minutes.	

The success of cDNA synthesis was validated by PCR using GAPDH primers. The resulting cDNAs were further used in RT-qPCR experiments.

2.11. RT-qPCR

Real-Time Quantitative PCR (RT-qPCR) experiments were performed by using Bio-RAD CFX-Connect detection system. For the reactions, BioRAD SYBR Green Supermix (CAT#172-5270) including both reverse transcriptase and hot-start DNA Polymerase was used. As primers, 0.4-0.5 µM of forward and reverse primers that are shown in Table 2.1 were used together with 3µL of cDNAs. When the reactions were completed, Ct values were calculated by using relative standard curves. For each reaction, normalizations were made by using the Ct values coming from the reference house-keeping genes like RPLPO. Then, all these Ct values were used to calculate ΔΔCq fold change (Livak & Schmittgen, 2001).

Moreover, for each RT-qPCR experiment, melt peaks were checked to see if there was any non-specific product, and everything was done by following MIQE Guidelines (Bustin et al., 2009).

EIF2B3 intronic transcripts were amplified with EIF2B3 intronic_F 5'-GGCCAGAGACACAGATTGGA-3', R: 5'- ACAGGGGGAAACATACCCTTC -3' (product size: 143 bp, annealing temperature: 59.5˚C) and EIF2B3 Long_F: 5'-GTGATCGAGAAGGGTGCAGAC-3', R: 5'- GTCATTCCCCACGATCACCTC -3' (product size: 102 bp, annealing temperature: 64˚C).

RPLP0 was amplified by using RPLP0_F: 5'-GGAGAAACTGCTGCCTCATA-3', RPLP0_R: 5'- GGAAAAAGGAGGTCTTCTCG-3' (product size: 191 bp, annealing temperature: 60˚C).

For checking the success of E2 treatment, TFF1 expressions were calculated. TFF1 was amplified by using TFF1_F: 5'- TTGTGGTTTTCCTGGTGTCA -3' and TFF1_R: 5'-CCGAGCTCTGGGACTAATCA -3' (product size: 209 bp, annealing temperature: 56˚C).

2.12. 3'RACE

RACE specific cDNAs were synthesized by using RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, CAT# EP0441) with 5 µg total RNA (DNase treated) from both starved and E2/EtOH treated MCF7 cells, and the oligo dT-anchor primer (5'-GACCACGCGTATCGATGTCGACTTTTTTTTTTTTTTTTV-3'). For the nested PCRs, reverse primer specific for the anchor sequence was used (Anchor_R: 5'-GACCACGCGTATCGATGTCGAC-3'). For the first round of 3'RACE, F1 forward primer was used (EIF2B3 3'RACE_F1 5'-GTCCATTCGTCAGCCCAGAT-3') together with the reverse anchor primer in the following conditions: 95°C for 3 minutes, 34 cycles of 95°C for 30 seconds, 58°C for 30 seconds, 72°C for 30 seconds. Expected product sizes for the first round of 3'RACE were given in Table 2.3.

Table 2.3. *Expected product sizes for the first round of 3'RACE*

PolyA Site ID	*Expected Product Size*
Hs.533549.1.2	*593bp*
Hs.533549.1.3	*553bp*
Hs.533549.1.4	*491bp*
Hs.533549.1.10	*369bp*
Hs.533549.1.14	*188bp*

For the second round, Forward 2 (F2) primers (EIF2B3 3'RACE_F2 5'-GTCAGCCCAGATTGTCAGCA-3') together with the anchor reverse primer were used. As a template, 1/20 diluted PCR products of the first round 3'RACE were used. Touchdown PCR was performed with the following conditions: 95°C for 3 minutes, 1 cycles of 95°C for 30 seconds, 64°C for 30 seconds, 72°C for 30 seconds, 2 cycles of 95°C for 30 seconds, 62°C for 30 seconds, 72°C for 30 seconds, and 29 cycles of 95°C for 30 seconds, 60°C for 30 seconds, 72°C for 30 seconds, and lastly 72°C for 5 minutes. Expected product sizes for the second round of 3'RACE were given in Table 2.4. After the completion of PCRs, products were loaded onto 1.5% agarose gel and then the desired bands were extracted from the gel by using Zymoclean™ Gel DNA Recovery Kit (CAT#D4008). The purity and the concentration of the DNA samples were measured by BioDrop Duo (Isogen Life Science).

Table 2.4. *Expected product sizes for the second round of 3'RACE*

PolyA Site ID	*Expected Product Size*
Hs.533549.1.2	*585bp*
Hs.533549.1.3	*545bp*
Hs.533549.1.4	*483bp*
Hs.533549.1.10	*361bp*
Hs.533549.1.14	*180bp*

2.13. Cloning of 3'RACE products to pGEM-T Vectors

Two of the 3'RACE products showing the usage of intronic poly(A) site and Hs.533549.1.4 poly(A) site were extracted from the agarose gel by using DNA recovery kit mentioned above. After measuring the concentrations of corresponding 3'RACE products, these products were inserted into 50 ng pGEM®-T Easy Vector (Promega, CAT# A1360) and ligated by T4 DNA ligase (Promega, CAT# A1360) at 4˚C for 16h. For the ligation, the following calculations were made:

$$\text{Insert amount (ng)} = \frac{\text{Vector amount (ng) x MW of insert (kb)}}{\text{Molecular Weight of the vector (kb)}} x \frac{3(insert)}{1(vector)}$$

$$\text{Volume of insert (µl)} = \frac{\text{Amount of insert (ng)}}{\text{Measured concentration of extraction product}(\frac{\text{ng}}{\mu\text{l}})}$$

Next, ligated vector was transformed into *E.coli* cells and colony PCR was performed by taking single colonies from the bacteria plate. Insertion of the intronic 3'RACE product was confirmed by colony PCR using intronic RT-qPCR primers. On the other hand, presence of Hs.533549.1.4-3'RACE-pGEMT product was confirmed by the colony PCR using F2 and reverse anchor primers. One of the positive colonies from each case was selected and isolated plasmids were sent for sequencing. Sequencing result was shown in Figure 3.13, and Figure 3.14.

2.14. Cloning of 3'UTR Isoforms into pMIR-Report Luciferase Vectors

For the cloning of 3'UTR isoforms, specific primers having restriction enzyme digestion cut sites were designed for Hs.533549.1.2, Hs.533549.1.3, Hs.533549.1.4, and Hs.533549.1.10 poly(A) sites. MluI (Forward) and HindIII (Reverse) cut sites were incorporated into specific primers which were shown in Figure 2.1. Firstly, PCR reactions having 50 µl volume in total were performed with each specific primer pairs using MCF7 cDNA as a template. Then, these PCR products were loaded into agarose gel and desired products were extracted from the gel. By using the formula in above section, PCR products were ligated into pGEM-T vectors. Next, ligation products were transformed into bacteria cells and colony PCR was performed for the resulting

colonies. The selected positive colonies were taken for plasmid isolation and these plasmids were double digested by using MluI and HindIII enzymes at 37°C, for 3 hours. Later, both digested products and the empty pMIR-Report Luciferase vectors were loaded onto agarose gel and the digested parts were extracted from the gel by the gel recovery kit. Then, digested products were ligated into digested pMIR empty vectors by T4 DNA Ligase enzyme and by using the same equation shown above. After the ligation, ligated vectors were transformed into bacteria cells and verified by colony PCR. Plasmid isolation was performed for one of the positive colonies in each case, and the plasmids were send for sequencing to confirm the presence of the inserts.

2.15. Dual-Luciferase Assay

After the confirmation of 3'UTR isoform insertions into pMIR-Report Luciferase vectors by sequencing results, MCF7 cells were co-transfected with phRL-TK (Renilla Luciferase) (125 ng) and pMIR-Report Luciferase (Firefly) (375 ng) in 24-well cell culture plates by using Turbofect Transfection Reagent. After 24 hours of transfection, cells were lysed and manufacturer's instructions for Dual-Luciferase® Reporter Assay System (Promega, CAT# E1910) were followed. Luciferase activities were measured by using Modulus Microplate Luminometer (Turner Biosystems). After getting the measurements, luciferase activity of Firefly luciferase was normalized to those of Renilla luciferase.

2.16. Transfections

All transfections were performed by using Turbofect Transfection Reagent (Thermo Fisher Scientific, CAT#R0531) following the manufacturer's instructions. For the transfection of the cells with pCDNA vectors, MCF7 cells were grown up to 50-60% confluency in 6-well cell culture plates. 2µg/mL plasmids were mixed with 200µL of serum-free DMEM and 4µL of transfection reagent. Mixture was incubated for 30 minutes, and added onto each wells drop by drop. Next, the cells were incubated at 37°C for 24 hours at CO_2 incubator.

2.17. Cloning of Coding Sequences into pCDNA 3.1 (-) Vectors

For the cloning of coding sequences of both full and intronic transcripts, specific primers having restriction enzyme digestion cut site and HA-sequence tags were designed. XbaI enzyme cut site was incorporated into forward primers, while the HindIII cut site was incorporated into reverse primers. Moreover, primers were designed to have HA- sequence on either N-terminus or C-terminus of both transcripts (Table 2.1.). All cloning steps were same with the cloning of 3'UTR isoforms into pMIR vectors, besides the initial PCR reaction which had the Phusion Polymerase enzyme (Thermo Fisher Scientific, CAT# F530S) instead of Taq Polymerase (Thermo Fisher Scientific, CAT# EP0401). Since Phusion Polymerase does not add -A nucleotides at the end of the PCR products as Taq Polymerase do, an additional -A tailing procedure was performed for insertion of PCR products into pGEM-T vectors. -A tailing was performed by using 10 mM dATP and the nucleotides were added by Taq Polymerase. Resulting -A tailed PCR products were inserted into pGEM-T vectors by the same calculation above. Then, the vectors with insertion were transformed into bacteria and verified by colony PCR. Positive colonies were selected for plasmid isolation. Then, both the plasmids(5µg) and pCDNA 3.1(-) empty vectors were double digested with XbaI and HindIII enzymes at 37°C for 3 hours. Next, the digestion products were loaded into 1% agarose gel and the desired products were extracted from the gel. Later, digested plasmids were ligated into double digested pCDNA vectors by T4 DNA Ligase enzyme at 16°C for overnight. After ligation, ligation products were transformed into bacteria cells and ligation is validated by colony PCR using vector specific T7 and BgH reverse primers. One of the positive colonies for each case was selected and isolated plasmids of those colonies were sent for sequencing to confirm the insertion.

2.18. Protein Isolation

Total proteins of the transfected MCF7 cells were isolated by using M-PER Mammalian Protein Extraction Reagent (Thermo Fisher Scientific, CAT# 78501).

10X phosSTOP (Roche, CAT# 04906837001) and 25X protease inhibitor (PI) (Roche, CAT# 1187350001) were mixed with M-PER. Then, the isolation procedure was performed by following the manufacturer's manual. Next, the protein concentrations were measured by using Pierce BCA Protein Assay Kit (Thermo Fisher Scientific, CAT# 23227) according to manufacturer's instructions.

2.19. Western Blotting

50 µg of total proteins were denatured by using 6X Laemmli buffer at 100°C for 5 minutes. Then, the denatured proteins were run on 12% seperating and 5% stacking polyacrylamide gel and transferred to PVDF Western Blotting Membrane (Roche, CAT# 03010040001) by using wet-transfer system at 100V for 75 minutes. Next, the membranes were blocked with 5% non-fat milk in 0.1% TBS-T (Tris Buffer Saline-Tween) for 1 hour at room temperature with continious shaking. Later, the membrane was incubated with 1:2000 Anti-HA tag primary antibody (Abcam, CAT# ab9110) for overnight at 4°C. After overnight incubation, the membrane was washed with TBS-T for 10 minutes, 3 times. Washing steps were followed by 1:2000 secondary goat anti-rabbit HRP antibody (Advansta, CAT# R-05072-500) for 1 hour at room temperature. After incubation with secondary antibody, the membrane was washed with TBS-T for 10 minutes, 3 times. For the visualization, WesternBright ECL kit (Advansta, CAT# K-12045-D50) was used. As a loading control, ACTB protein levels were measured. For the blocking, 3% BSA in 1% TBS-T was used and it was followed by 1:2000 primary ACTB antibody (Santa-Cruz Biotechnology, CAT# sc-47778) incubation for overnight. Then, the membrane was washed with TBS-T for 10 minutes, 3 times. Next, the membrane was incubated with 1:2000 secondary anti-mouse antibody (Santa-Cruz Biotechnology) at room temperature for 1 hour. After 3 times washing, the blots were visualized. As a second loading control, GAPDH protein expression was checked. For the blocking, 5% Skim Milk in 1% TBS-T was used and it was followed by 1:2000 GAPDH Antibody (Santa-Cruz Biotechnology, CAT # sc-25778) incubation for overnight at +4°C. As a secondary antibody, 1:2000 anti-rabbit antibody (Advansta, CAT# R-05072-500) was used.

CHAPTER 3

RESULTS AND DISCUSSION

3.1. *EIF2B3* and Probe Level Screen

Estradiol (E2) has long been known to be a robust proliferative signal in breast cancer cells (Prall, Rogan, & Sutherland, 1998). To understand whether E2 induces APA in ER+ breast cancers, an *in-silico* approach was used. APA regulated transcripts were identified by APADetect Java program, which was developed by Prof. Dr. Tolga Can (Department of Computer Engineering, METU, Turkey) (Akman et al.,2012). By using APADetect on publicly available microarray datasets, alterations of probe levels between the proximal and distal poly(A) sites were identified.

In this study, we have used expression profiling dataset of 100 nM E2 treated MCF7 samples at different time points (0, 3, 6, and 12 hours), which is available at Gene Expression Omnibus (GEO) database with the accession number of GSE11324. The average intensities of proximal and distal probes in E2 treated samples at different time points were compared with that of control (zero timepoint) to understand the differences between the proximal/distal ratios at different time points.

For testing the statistically significant genes undergoing APA events, Significance Analysis of Microarrays (SAM) was used. When the fold changes representing the proximal/distal ratios of these significant genes were analysed, one gene caught our attention as a common APA candidate; *EIF2B3* gene having a fold change < 0.6 at two different time points, 6hr and 12hr, which are shown at Table 3.1 and Table 3.2.

Table 3.1. *APA candidates found by the comparison of control and 6 hours of 100 nM Estrogen treated MCF7 Cells by SAM.*

Gene Name	Probeset ID	PolyA Site ID	Fold Change (unlogged)
TCF3	213730_x_at	Hs.371282.1.10	0.5437103
EIF2B3	218488_at	Hs.533549.1.10	0.56412125

Table 3.2. *APA candidates found by the comparison of control, and 12 hours of 100 nM Estrogen treated MCF7 Cells by SAM.*

Gene Name	Probeset ID	PolyA Site ID	Fold Change (unlogged)
TCF3	213730_x_at	Hs.371282.1.10	0.5551587
UBE2D3	200669_s_at	Hs.518773.1.8	0.64215034
EIF2B3	218488_at	Hs.533549.1.10	0.5900842
EIF2B3	218488_at	Hs.533549.1.14	0.4826824
RB1	211540_s_at	Hs.408528.1.2	0.6509687
SCARB2	201647_s_at	Hs.349656.1.10	0.5344845
SCARB2	201647_s_at	Hs.349656.1.9	0.57973385
MKI67	212022_s_at	Hs.80976.1.16	0.7645466

Among these APA events, we focused on *EIF2B3* as a regulator of ribosome assembly, a process which may be regulated in response to E2 (Cochrane & Deeley, 1984).

EIF2B3 has five poly(A) sites. Three Poly(A) sites Hs.371282.1.2, Hs.371282.1.3, and Hs.371282.1.4 are found within the 3'UTR, remaining two poly(A) sites (Hs.371282.1.14 and Hs.371282.1.10) are located at the 9th exon, and at an intronic site between the 9th and 10th exons (PolyA_db2), respectively.

In our *in-silico* analysis, the Hs.371282.1.10 poly(A) site was of interest due to increased distal probe mean intensities compared to those of proximal ones in the presence of Estrogen.

Relative distributions of 11 Affymetrix probes are shown in Figure 3.1. Eight of the probes were mapped to the proximal part of the transcript marked by the intronic poly(A) site (Hs.533549.1.10), while 3 of them were mapped distal to the poly(A) site.

Figure 3.1. Gene schema for *EIF2B3* and the positions of the probes.

Exon intron boundaries, Affymetrix probe locations, and the five poly(A) sites found on *EIF2B3* transcript were shown in the figure. Intronic poly(A) site was indicated by a red circle in the figure. Eight of the Affymetrix probes are found at the upstream of the intronic poly(A) site, while three of them are found downstream. Also, exon numbers were indicated above the corresponding exons.

Ratios of the proximal and distal probe intensity means for both treated and control samples for *EIF2B3* are given in Table 3.3. The ratio of the means of proximal and distal probe intensities for the E2 treated samples (R_{E2}) normalized to that of control samples (R_c) gives the R_{E2}/ R_c ratio. *EIF2B3* was selected as an APA candidate as it

may undergo a lengthening event, indicated by a decreased R_{E2}/ $R_{c.}$ratio of Hs.371282.1.10 in E2 treated cells.

Table 3.3. *Ratios of the means of proximal and distal probe intensities for treated and control samples.*

Series Number	E2 Concentration	Treatment Duration	Proximal/ Distal -E2 (R_{E2})	Proximal/ Distal - Control (R_c)	Fold Change (R_{E2}/ R_c)
GSE11324	100 nM	6h	1.36	3.11	0.44
GSE11324	100 nM	12h	1.45	3.11	0.47

Next, gene structure and the transcript isoforms of this APA candidate was investigated.

3.2. *EIF2B3* Gene Structure and *EIF2B3* Transcript Isoforms

Eukaryotic Initiation Factor 2B-3 (*EIF2B3/ EIF2Bgamma/ EIF2Bγ*) gene is located on the minus strand of the first chromosome (1p34.1) in the human genome (GRCh38.p13). This gene is conserved among many eukaryotic organisms from *S. pombe* to chimpanzee (HomoloGene: 7005, NCBI)

EIF2B3 gene has three mRNA isoforms that are experimentally confirmed (Figure 3.2). The first transcript variant (NM_020365.5) is the longest isoform with a total of 11 coding exons and 136,201bp (including UTRs). This full-sized transcript encodes a 452 amino acid long peptide (NP_065098) which has a molecular weight of approximately 58kDa.

Figure 3.2. Transcript variants of *EIF2B3* gene.

Transcript variant 1 encodes the longest isoform with 11 coding exons. Transcript variant 2 is the intronic isoform having an exonized intron, which was shown as a grey box on the figure, at its terminal exon. Lastly, transcript variant 3 has one exon missing compared to the first isoform.

The second isoform (NM_001166588.2) is 106,528 bp long in the coding region. This isoform actually has 9 coding exons; however, among all 9 exons, one exon has a part which is formed by the inclusion of a part of the next intron. Intronic sequence encoded 11 new amino acids which are not found in the first isoform are shown in Figure 3.3. Lastly, the transcript variant 3 (NM_001261418.1) has 130,169bp long coding region size and it lacks one exon which is found in the first isoform. Intronic and full isoform differ in their C- termini and this could be realized in the protein structure predictions of the isoforms in Figure 3.4.

Figure 3.3. Alignment of the amino acids encoded by different isoforms of *EIF2B3* gene.

The wild type amino acids found in the full-length transcript were shown in a yellow rectangle. On the other hand, newly added 11 amino acids coming from the exonization of the intronic part are indicated with red color.

Figure 3.4. Protein structure predictions of two isoforms (SWISS-MODEL).

Structure predictions were made by using SWIS-MODEL algorithm. In the figure, both N-terminus and C-terminus were indicated by green and yellow arrows, respectively. The differences at C-termini of distinct isoforms could be seen in the figures.

Moreover, conserved domain search showed that GCD1 superfamily domain, is truncated in the IPA isoform. This domain is conserved between both *EIF2B3*(-gamma) and EIF2B5(-epsilon), which together form the catalytic subcomplex. Figure 3.5 shows the conserved domains on full-sized protein isoform. Between the amino acids of 4-447 are the part of this conserved GCD1, NDP-sugar pyrophosphorylase domain (NCBI, Conserved Domain Database). Since the intronic isoform has the initial 401 amino acids common with the full-sized isoform, part of this domain is lost in the intronic one.

Figure 3.5. Conserved domains of the first isoform.

Figure shows the conserved domains of the full isoform. eIF-2B_gamma_N (3-212amino acids) represents the N-terminal domain of EIF2B3. GCD1 superfamily (4-447) represents the NDP-sugar pyrophosphorylase domain which is conserved between both EIF2B-gamma and -epsilon subunits.

Considering the important function of *EIF2B3* in the translation initiation process and the structural alteration in this IPA isoform, we thought that this could be a promising IPA candidate. Next, we wanted to confirm existence of this isoform that uses an intronic poly(A) site.

3.3. Experimental Confirmation of *EIF2B3* in E2 Treated MCF7 Samples

To understand whether *EIF2B3* undergoes intronic polyadenylation with E2 stimulation, we performed E2 treatment on MCF7 cells, which are ER+ breast cancer cells. According to the APADetect results, 100 nM E2 treatment for 6 hr and 12 hr resulted with an SLR change for *EIF2B3*. Therefore, MCF7 cells were treated with 100 nM E2 for 12 hours after 72 hours of starvation. As a control, vehicle (EtOH) treatment was also performed at the same conditions. Then, the cell pellets were collected and used for RNA isolation. After DNase treatment, lack of DNA contamination was proven by PCR using GAPDH primers (see Appendix D). Then, these RNA samples were further used for cDNA production to validate the successful treatment. One of the most well-known target genes of E2, Trefoil Factor 1 (*TFF1*), was used to confirm the success of treatment (Bourdeau et al., 2008). Figure 3.6 shows *TFF1* upregulation at three different biological replicate treatments.

Figure 3.6. RT-qPCR Showing the TFF1 Expression on E2/EtOH Treated MCF7 cDNAs.

TFF1 expression was used to check the success of E2 treatment. Relative expression levels of *TFF1* was compared between E2 and EtOH treated MCF7 cDNAs in three different biological replicate treatments. All expressions were normalized to the reference gene, *RPLPO.* Quantifications were made by using ΔΔCq method *(Livak & Schmittgen, 2001)*

3.3.1. RT-qPCR Quantification

RT-qPCR standardizations were made by following MIQE guidelines (Bustin et al., 2009). Melt curve analyses were checked to test whether there is any non-specific product or primer dimer formation. It should be pointed out that specific primer pairs were designed for the intronic and long (full) transcripts to get the expressions of those transcripts separately. The reverse primer for the intronic transcript was designed from the exonized part of the intron so that it only maps to the intronic transcript. The primers for the full transcript were designed from the last two exons so that they will amplify only the long transcript.

In Figure 3.7, relative expression levels for both intronic and full transcripts are shown. The intronic isoform's expression was upregulated by E2 induction. The

upregulation was also detected at the full transcript, and it was slightly more than the intronic one.

Figure 3.7. Relative Quantification of *EIF2B3* intronic and long isoforms in E2 and EtOH Treated MCF7 Cells.

MCF7 cells were treated with 100nM E2 for 12 hours. Expression changes were normalized to RPLPO reference gene. Quantifications were made by using ΔΔCq method (Livak & Schmittgen, 2001). For the statistical analysis, unpaired student t-test was used resulting with p values of 0.024, and 0.011 for the intronic and long isoform, respectively.

3.3.2. 3' Rapid Amplification of cDNA Ends (3' RACE)

To further validate the usage of intronic poly(A) site, we performed 3'RACE. To do this, we synthesized 3'RACE-ready cDNAs from the E2, and EtOH treated samples by using oligo dT-anchor primers. For the first-round amplification, F1(forward) primers were used together with reverse anchor primers. For the second round, first round products were used as the template, and they were amplified by using F2 (forward) primers to perform nested 3'RACE PCR. Annealing of both F1&F2 primers and the predicted product sizes for each poly(A) site usage are shown in Figure 3.8.

Figure 3.8. Target sequence locations of F1&F2 primers and predicted product sizes.

For all nested 3'RACE PCR reactions, the same F1, F2, and Reverse Anchor primers were used. Predicted product sizes for each corresponding poly(A) sites are given in the figure.

3'RACE PCR result is given in Figure 3.9. On the same figure, predicted product sizes corresponding to usage of different poly(A) sites are also shown. The usage of intronic poly(A) site (Hs.533549.1.10) was not detected in starved condition; however, it was increased in E2 treated cDNAs.

PolyA Site ID	Expected Product Size
Hs.533549.1.2	593bp
Hs.533549.1.3	553bp
Hs.533549.1.4	491bp
Hs.533549.1.10	369bp
Hs.533549.1.14	188bp

Figure 3.9. First-round 3'RACE PCR products on an agarose gel.

Starvation, 12h EtOH and E2 treated RACE ready cDNAs were used as a template for the 3'RACE procedure. Predicted product sizes according to different poly(A)sites were given on the right side of the figure. 4 of 5 poly(A) site usages were detected in the gel photo.

Since there were many non-specific bands on the agarose gel figure of the first round 3'RACE, we performed nested 3'RACE PCR in order to increase the specificity of the bands we had seen. For the second round, the products of the first-round PCR were used, and amplified by F2 primer together with Race-anchor primer. In Figure 3.10, an agarose gel photo showing the products of the second round is presented. A PCR product that would suggest the usage of the intronic poly(A) site was detected at 361bp.

To further confirm that these PCR products truly represented the usage of those poly(A) sites, the bands were extracted from the gel, purified and inserted into pGEM-T vector as described in the Materials and Methods section. Then, these vectors were transformed into *E.coli* cells and the presence of pGEMT-3'UTR RACE products' insertion was validated by colony PCR method (Figure 3.11, Figure 3.12).

PolyA Site ID	Expected Product Size
Hs.533549.1.2	585bp
Hs.533549.1.3	545bp
Hs.533549.1.4	483bp
Hs.533549.1.10	361bp
Hs.533549.1.14	180bp

Figure 3.10. Second round 3'RACE PCR products on agarose gel validate the usage of intronic poly(A) site.

Diluted first-round RACE products were purified and used for the amplification process. 35 cycle of Touch-down PCR was performed as described in the Materials and Methods part in order to increase the specificity of the products. Predicted product sizes for different poly(A) site usages were given on the right side of the figure accordingly.

Figure 3.11. Colony PCR verifying the presence of pGEMT-3'UTR insertion.

The figure shows the agarose gel result of colony PCR to check whether the insertion is successful or not. 35 cycle colony PCR was performed by using F2&Race anchor primers. All seven colonies taken from the bacteria plate show the presence of the expected product at almost 483 bp.

Figure 3.12. Colony PCR verifying the presence of pGEMT- intronic 3'UTR insertion.

The figure shows the agarose gel result of colony PCR to check whether the intronic 3'UTR insertion is successful or not. 35 cycle colony PCR was performed by using intronic RT-qPCR primers. Six of seven colonies taken from the bacteria plate show the presence of the expected product at almost 143 bp.

Next, isolated plasmids from the positive colonies were sent for sequencing to confirm the PCR results. Sequencing was performed by using the pGEM-T specific universal primers; T7 and SP6. Sequencing results are shown in Figure 3.13, and Figure 3.14.

```
                F2 RACE primer
Query  1    GTCAGCCCAGATTGTCAGCAAACACCTGGTTGGAGTTGACAGCCTCATTGGGCCAGAGAC  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  431  GTCAGCCCAGATTGTCAGCAAACACCTGGTTGGAGTTGACAGCCTCATTGGGCCAGAGAC  372

Query  61   ACAGATTGGAGAGAAGTCATCCATTAAGCGCTCAGTCATTGGCTCATCCTGTCTCATAAA  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  371  ACAGATTGGAGAGAAGTCATCCATTAAGCGCTCAGTCATTGGCTCATCCTGTCTCATAAA  312

Query  121  AGATAGAGTGACTATTACCAATTGCCTTCTCATGAACTCAGTCACTGTGGAGGAAGGGTA  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  311  AGATAGAGTGACTATTACCAATTGCCTTCTCATGAACTCAGTCACTGTGGAGGAAGGGTA  252

Query  181  TGTTTCCCCCTGTACCCACTTGAGGCAAAGATAATGGAAAAATGCCTTCAAAATTCTGAA  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  251  TGTTTCCCCCTGTACCCACTTGAGGCAAAGATAATGGAAAAATGCCTTCAAAATTCTGAA  192

Query  241  GGAAGGTTAATTTTAACCTAAAATTGTATACCCAGGCAAAGTGCTAATCAAGAGGGCAAT  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  191  GGAAGGTTAATTTTAACCTAAAATTGTATACCCAGGCAAAGTGCTAATCAAGAGGGCAAT  132

                      Hs.533549.1.10 PolyA Site
                                                RACE anchor Reverse Primer
Query  301  AGAATAAAAACATTTTCAGATATCaaaaaaaaaaaaaaaaaaGTCGACATCGATACGCG  361
            |||||||||||||||||||||||| |    |||||||||||||||||||||||||||||||
Sbjct  131  AGAATAAAAACATTTTCAGATA-C----AAAAAAAAAAAAAAAGTCGACATCGATACGCG  75

Query  362  TGGTC  366
            |||||
Sbjct  74   TGGTC  70
```

*Figure 3.13.*Plasmid sequencing result confirming the usage of intronic poly(A) site.

The above figure shows the alignment of the expected product (query), and the sequencing result (subject). Alignments of F2 RACE primer and RACE anchor reverse primers were also shown in the figure. The sequencing result validates our intronic 3'UTR insertion into pGEM-T vector.

```
                F2 RACE primer
Query  1    GTCAGCCCAGATTGTCAGCAAACACCTGGTTGGAGTTGACAGCCTCATTGGGCCAGAGAC  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  69   GTCAGCCCAGATTGTCAGCAAACACCTGGTTGGAGTTGACAGCCTCATTGGGCCAGAGAC  128

Query  61   ACAGATTGGAGAGAAGTCATCCATTAAGCGCTCAGTCATTGGCTCATCCTGTCTCATAAA  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  129  ACAGATTGGAGAGAAGTCATCCATTAAGCGCTCAGTCATTGGCTCATCCTGTCTCATAAA  188

Query  121  AGATAGAGTGACTATTACCAATTGCCTTCTCATGAACTCAGTCACTGTGGAGGAAGGAAG  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  189  AGATAGAGTGACTATTACCAATTGCCTTCTCATGAACTCAGTCACTGTGGAGGAAGGAAG  248

Query  181  CAATATCCAAGGCAGTGTCATCTGCAACAATGCTGTGATCGAGAAGGGTGCAGACATCAA  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  249  CAATATCCAAGGCAGTGTCATCTGCAACAATGCTGTGATCGAGAAGGGTGCAGACATCAA  308

Query  241  GGACTGCTTGATTGGAAGTGGCCAGAGGATTGAAGCCAAAGCTAAACGAGTGAATGAGGT  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  309  GGACTGCTTGATTGGAAGTGGCCAGAGGATTGAAGCCAAAGCTAAACGAGTGAATGAGGT  368

Query  301  GATCGTGGGGAATGACCAGCTCATGGAGATCTGAGTTCTGAGCAAGTCAGACTCCTTCCT  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Sbjct  369  GATCGTGGGGAATGACCAGCTCATGGAGATCTGAGTTCTGAGCAAGTCAGACTCTTTCCT  428

Query  361  TTTGGCCTCCAAAGCCACAGATGTTGGCCGGCCCACCTGTTTAACTCTGTATTTATTTCC  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  429  TTTGGCCTCCAAAGCCACAGATGTTGGCCGGCCCACCTGTTTAACTCTGTATTTATTTCC  488

                        Hs.533549.1.4 PolyA Site
                                      RACE anchor Reverse Primer
Query  421  CAATAAAGAAGGGCTTCCAAAGGCaaaaaaaaaaaaaaaGTCGACATCGATACGCGTGG  480
            |||||||||||||||||||||||||  |||||||||||||||||||||||||||||||||
Sbjct  489  CAATAAAGAAGGGCTTCCAAAGGC--AAAAAAAAAAAAAAGTCGACATCGATACGCGTGG  546

Query  481  TC  482
            ||
Sbjct  547  TC  548
```

Figure 3.14. Plasmid sequencing result confirming the usage of Hs. 533549.1.4. poly(A) site.

The above figure shows the alignment of the expected product (query), and the sequencing result (subject). Alignments of F2 RACE primer and RACE anchor reverse primers were also shown in the figure. The sequencing result validates Hs. 533549.1.4-3'UTR insertion into pGEM-T vector. No mismatch or missing nucleotide was detected in the sequencing results.

Figure 3.13. shows the plasmid sequencing result indicating the presence of pGEM-T -intronic 3'UTR. Next, Figure 3.14. shows the sequencing result indicating the presence of pGEM-T-Hs.533549.1.4 3'UTR, which was normally present at the beginning of 3'UTR of the transcript isoform 1. Therefore, these results together confirmed the existence of both intronic poly(A) site, and Hs. 533549.1.4 poly(A) site.

After confirming that this intronic isoform existed and that C-terminus amino acid composition of *EIF2B3* can change, we began to delineate other potential differences in these isoforms. One of the important factors to change mRNA stability, and subcellular localization is the 3'UTR. In our case, the 3'UTRs of intronic and long isoform are completely different and this difference may be important. Hence, I cloned 3'UTR of the IPA and the long isoforms into pMIR-Report Luciferase vector.

3.4. Cloning of 3'UTR Isoforms into Reporter Plasmid

3' UTR isoforms were cloned into the multiple cloning site downstream of the firefly luciferase gene on pMIR-Report Luciferase vector. The detailed figure for the plasmid is given in Appendix (A). HindIII and MluI recognition sites were incorporated into newly designed pMIR cloning primers. Firstly, MCF7 cDNAs were amplified by using designed pMIR primers for each different poly(A) site, separately. Agarose gel results of the PCR products are shown in Figure 3.15. Then, these products were extracted from the agarose gel, purified, and ligated into pGEM-T vector. Later, pGEM-T vectors were transformed into *E.coli* cells, and colony PCR was performed to confirm the plasmid insertions. One of the positive colonies for each corresponding poly(A) site was chosen and plasmid isolation was performed. Resulting plasmids were double digested with MluI and HindIII enzymes together with pMIR empty vector. Next, digestion products were ligated into pMIR empty vector, and the resulting colonies from bacterial transformation were used in order to validate the presence of the inserts by colony PCR. Agarose gel photos of the colony PCRs are shown in Figure 3.16 and 3.17.

Figure 3.15. Cloning of Different 3'UTR Isoforms.

MCF7 cDNAs were amplified by pMIR cloning primers and loaded onto 2% agarose gel. For each 3'UTR isoform, there was a specific primer with different annealing temperatures. Therefore, the reactions were performed at the same PCR with a gradient program. The expected product for the second lane was 224 bp, for the third lane was 184 bp and for the fourth one 122 bp. On the right side of the figure, a separate gel photo representing the amplification of intronic 3'UTR isoform was shown. The expected product size was 126 bp.

Figure 3.16. Colony PCR for the validation of 3'UTR isoforms in pMIR vector.

35 cycle colony PCR was performed by using pMIR specific primers. The expected product size for Hs.533549.1.2-3'UTR in pMIR was 457bp, while it was 355bp for Hs.533549.1.4-3'UTR in pMIR vector. pMIR EV plasmid was used as a positive control, which gives the expected product of 233bp. All colonies have the expected inserts.

Figure 3.17. Colony PCR for the validation of 3'UTR isoforms in pMIR vector.

35 cycle colony PCR was performed by using pMIR specific primers. The expected product size for Hs.533549.1.3-3'UTR in pMIR was 417bp, while it was 359bp for Hs.533549.1.10-3'UTR in pMIR vector. pMIR EV plasmid was used as a positive control, which gives the expected product of 233bp. Most of the colonies gave the expected product size, except the ones with empty lanes.

One of the positive colonies were selected for each corresponding 3'UTR isoform, and plasmid isolation of those colonies were performed. Resulting plasmids were sent for sequencing to further confirm the PCR results. Sequencing results confirmed the successful cloning of all 3'UTR isoforms into pMIR vectors (see Appendix (E)).

3.5. Dual-Luciferase Assay

MCF7 cells were co-transfected with different pMIR-3'UTR isoform constructs and phRL-TK, which is the Renilla luciferase construct for the normalization. After 24 hours of co-transfection, both renilla and firefly luciferase activities were counted by Modulus Microplate Luminometer (Turner Biosystems). Then, relative luciferase activities were calculated by normalization of Firefly readings to Renilla luciferase activity. In Figure 3.18., the relative luciferase activities corresponding to different 3'UTR isoforms are shown. According to Dual-luciferase assay results, there was no significant difference between the luciferase activities of different constructs having different 3'UTR sequences.

Figure 3.18. Dual-luciferase reporter assay results.

Relative luciferase activities corresponding to different 3'UTR isoforms were shown in the figure. Firstly, firefly luciferase activities were normalized to Renilla luciferase activities. Then, Firefly/Renilla readings were normalized to those of pMIR EV. Relative luciferase activity of intronic 3'UTR is shown by red color.

3.6. Measuring the mRNA Stabilities

Since 3'UTR sequence did not alter translation efficiency, we turned into the possibility of altered mRNA stability. MCF7 cells were treated with 2 µg/mL Actinomycin D for 0, 30 minutes, 2 hours, and 12 hours to stop transcription and detect the half-lives of these isoforms.

RT-qPCR was performed by using cDNAs of treated samples as a template. RT-qPCR results showed no significant difference between intronic and full transcript stabilities (Figure 3.19).

Figure 3.19. Actinomycin D treatment showing mRNA Stability differences between intronic and long isoform.

2 µg/mL of Actinomycin D treatment of MCF7 cells was performed by İbrahim Özgül in our lab. Treatment durations were 0, 30 minutes, 2 hours and 12 hours. c-myc was used as a positive control since it is known to be an unstable transcript (Jones & Cole, 1987). No significant difference between the mRNA stability of intronic and full transcript was detected.

3.7. Detection of Protein Expressions

After seeing no difference between the mRNA stabilities of intronic and full isoforms, to test whether the intronic mRNA isoform could be translated into a protein product, we cloned the IPA and long isoform's CDSs into pcDNA 3.1 (-) empty vector. For easy detection, we incorporated HA-tags to N- and C- termini of coding sequences in different vector constructs. We generated four different pCDNA constructs namely; intronic HA-N terminus, intronic HA-C terminus, full HA-N terminus, and full HA-C terminus. Confirmation of these vector constructs by sequencing results are given in Appendix (F).

Next, we transfected HEK293T cells with these constructs for 24 hours and took the cell pellets for protein isolations. Then, Western Blotting was performed by using protein products. Antibody against HA-tag was used for the detection of protein levels of overexpression products. Both B-actin and GAPDH antibodies were used as

loading controls. When the protein expressions were examined, we observed that C-tagged proteins were not stable and possibly degraded. N-term tagged proteins gave 58 kDA and 54 KDa bands in agreement with expected sizes of full and intronic isoforms, respectively (Figure 3.20). This result showed us that the intronic transcript could be translated into a protein product. The size difference between the protein products of intronic and full-sized transcripts could be clearly seen in the figure.

Figure 3.20. Western Blotting for protein products of intronic and full transcripts.

Proteins were isolated from HEK293T cells after 24 hours of transfection with different HA- tagged pCDNA constructs. Anti HA- antibody was used to detect the presence of HA-tagged protein products of both transcripts. Untransfected HEK293T and pCDNA EV total proteins were used as negative controls. The expected products for the full HA N&C termini proteins were about 58kDa, while it was about 54 kDa for the intronic one. The size difference between full and intronic ones could be seen clearly. Both B-actin and GAPDH antibodies were used as loading controls.

3.8. Expression of Intronic *EIF2B3* in Breast Cell Lines

Given the existence of intronic transcript in Estrogen treated MCF7 cells and the structural significance of this isoform, we decided to further investigate this IPA isoform in different cell lines and tissues. First, we performed RT-qPCR in breast cancer cell lines to detect this isoform that uses an IPA. (Figure 3.21).

Figure 3.21. RT-qPCR showing the intronic transcript expression on different cell lines.

cDNAs from different cell lines were amplified by using intronic isoform specific primers. Expression levels were normalized to those of RPLP0 reference gene. Relative expression levels were found by normalization of intronic expression levels to that of normal breast cDNA. Quantifications were made by using ΔΔCq method *(Livak & Schmittgen, 2001)*

RT-qPCR suggested existence and expression of the intronic isoform in different cell lines. We used intron specific primers and we already confirmed lack of DNA contamination in our initial RNA samples. Therefore, it was curious why this isoform exists and how it may be functionally relevant to estrogen and to ribosomal function. We wanted to extend this analysis to different tissues and patient samples.

3.9. Expression of the Intronic Isoform in Different Tissues and Patient Samples

Next, we examined the presence of this intronic transcript in different tissues by using the publicly available datasets in order to support our results by showing the usage of intronic poly(A) site.

By using the data coming from GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2), we examined the expressions of both full-sized and intronic transcripts of *EIF2B3* in different tissues. From the graph below, it could be realized that the intronic transcript is present in all of the tissues (Figure 3.22). However, it should be also noted that the intronic transcript expression is generally very low in normal tissues.

Figure 3.22. Expression Profile for Both Full-Sized and Intronic Transcripts in Different Tissues.

In the figure, Transcripts Per Million (TPM) values of both transcripts are shown on the same bars. Intronic transcript expressions were demonstrated as striped bars, while the full transcript expressions were shown as plain bars.

After getting an idea about the expression of intronic transcript in different tissues, we have turned our attention to breast tissue since our initial APADetect analysis was based on E2 treated MCF7 samples. In normal breast tissues, there were almost a 6-

fold difference between the TPM values of intronic and full transcripts. To understand whether there is an increase in the usage of intronic poly(A) site in breast cancer cells, we analyzed a single-cell RNA sequencing data based on primary breast cancer cells. In this dataset (GEO Accession Number: GSE75688), there was a total of 515 samples from 11 breast cancer patients having different subtypes of the disease (Chung et al., 2017).

For the analysis of scRNA-seq data, we employed Seven Bridges Genomics tools (available at https://cgc.sbgenomics.com/). To do this, firstly, the SRA files were converted to FastQ Files by using the SRA Toolkit fastq-dump tool. Then, the output FastQ files were aligned to the reference genome (*GRCh37*) by using STAR 2.5.4b. Lastly, the aligned BAM files were visualized on the Integrative Genomics Viewer (IGV) 2.4.7. (available at https://igv.org/)

IGV Snapshots, which belong to two of the patients having ER+ subtype, were shown in Figure 3.23. One of the patients is coded as BC01, while the other is BC02. Interestingly, a robust increase of intronic poly(A) site usage could be realized when the ER+ cells were examined.

Figure 3.23. IGV Snapshots for ER+ Breast Cancer Cells Taken from the Patients.

BC01 and BC02 represent two distinct patients, and the numbers next to them mark the different single-cell RNA sequencing samples. The samples having the intronic transcript expression were indicated by blue color. The location of the intronic polyA site is shown in Poly(A)Db_2 track. And, the extension of the exon by the inclusion of the intronic sequence is indicated by a thinner line on RefSeq Genes track. Some of the ER+ samples have the intronic transcript expression excluding the pooled and bulk tumor samples.

In this single-cell RNA-sequencing data, not only the ER+ cells but also the other subtypes of breast cancer were analyzed. In Figure 3.24., samples belonging to a patient with double-positive (ER+ HER2+) subtype was shown as BC03. Moreover, the samples taken from the lymph node metastasis of the same patient were represented as BC03LN. When this IGV visualization is examined, it could be said that the intronic transcript expression is not common in this double positive subtype. Surprisingly, some of the samples taken from the lymph node metastatic part shows the expression of intronic transcript. Even though intronic transcripts were demonstrated to be highly expressed in these metastatic cases, the connection is currently not clear. Nevertheless, this discussion is beyond the scope of this thesis. Next, we analyzed the other patient samples having HER2+ subtype (Figure 3.25.). Some samples had the intronic transcript expression in this data too. In this case, some of the patient samples were under-represented due to our sample size, since many of the samples were filtered out because of lack of reads coming from *EIF2B3* gene. Lastly, the TNBC samples coming from five different patients were examined (Figure 3.26.). Many of the samples had the intronic expression in total.

Figure 3.24. IGV Snapshots for ER+ and HER2+ Breast Cancer Cells Taken from the Patients.

BC03 represents the RNA-sequencing samples from double-positive (ER+ and HER2+) subtype, while BC03LN shows the sequencing coming from the lymph node metastasis state of the same patient. As same with the previous one, the extension of the exon by the inclusion of the intronic sequence is indicated by a thinner line on RefSeq Genes track. And, the samples having the intronic expression could be followed by looking at the alignment of the intronic polyA site at Poly(A)Db_2 track. None of the BC03 samples have the intronic expression, while some of BC03LN samples have.

Figure 3.25. IGV Snapshots for HER2+ Breast Cancer Cells Taken from the Patients.

BC04, BC05 and BC06 represents the RNA-sequencing samples from three different patients having HER2+ subtype. Some of the samples, excluding the pooled ones, shows intronic expression.

Figure 3.26. IGV Snapshots for TNBC Cells Taken from the Patients.
BC07, BC08, BC09, and BC11 represent the RNA-sequencing samples from four different patients having a TNBC subtype. BC07LN indicates the samples taken from lymph node metastasis of BC07 patients. In total, many of the samples show intronic expression.

Overall, we confirmed expression of an IPA isoform for *EIF2B3*. This IPA isoform is expressed in breast cancer cell lines and patient samples. We confirmed the intronic poly(A) site usage using 3'RACE, cloning and sequencing. Moreover, we were able to express a protein from this IPA isoform. Next experiments will be to understand how this EIF2B3 IPA protein variant alters ribosome assembly dynamics and how this may be relevant to estrogen in breast cancer cells.

CHAPTER 4

CONCLUSION

Alternative polyadenylation is an important mechanism changing the 3'UTR diversity, which may alter both protein levels and functions. Intronic polyadenylation is an important APA type, in which the poly(A) site located on an intron is selected. When an intronic poly(A) site is chosen, some exons may be lost in the transcript resulting with a truncated protein, either functional or unfunctional.

To understand whether IPA occurs in the presence of E2, and its implications in breast cancers, first, we took an in-silico approach. By using APADetect analysis on publicly available microarray datasets consisting of 100 nM E2 treated MCF7 samples at different time points (0, 3, 6, and 12 hours) (GSE11324), we found several APA candidates whose probe levels between the proximal and distal poly(A) sites are altered in the presence of E2. Then, testing the statistically significant genes undergoing APA events by SAM revealed an important IPA candidate, *EIF2B3*, which may undergo a lengthening event, indicated by a decreased R_{E2}/ $R_{c.}$ratio of its intronic poly(A) site in E2 treated cells. *EIF2B3* encoding Eukaryotic Initiation Factor 2B-gamma is an important factor in ribosomal assembly and translation initiation. Therefore, we further investigated this intronic isoform and found that intronic poly(A) site selection causes loss of 51 amino-acids from the C-terminus of transcript. Moreover, 11 amino acids are added by the exonized part of the intron. Interestingly, conserved domain search showed loss of a GCD1 superfamily domain part which is common in both EIF2Bgamma and epsilon that forms the catalytic complex together. Given the structural changes and functional importance of this protein, we thought that this could be a promising IPA candidate. Then, we confirmed the existence of this IPA isoform by 3'RACE, cloning and sequencing. We also showed the increase in the expression of IPA isoform in the presence of E2, by RT-qPCR using 12 hours of

100nM E2 treated ER+ MCF7 cells. Moreover, by using the available scRNA-seqencing datasets consisting of patient samples (GSE75688), we showed that the intronic isoform expression is high in different subtypes of breast cancers.

After showing the usage of this intronic poly(A) site and its' expression increase in the presence of E2, we wondered if there is any functional change in 3'UTRs of different isoforms. Neither of 3'UTR isoforms showed a significant difference between the relative luciferase activities, meaning that 3'UTR sequences did not alter translational efficiency. As a next step, we tested whether mRNA stabilities are altered. For this purpose, we treated MCF7 cells with 2 µg/mL Actinomycin D for 0, 30 minutes, 2 hours, and 12 hours and detect the half-lives of these isoforms. No significant difference between mRNA stabilities of full and intronic isoform was observed. Next, we turned our attention to alterations at the protein levels. We managed to express a protein from the intronic isoform that has missing amino acids compared to long isoform as expected. Currently, the consequences of this truncation are not clear; however, many studies suggest that mutations in any of the five subunits of EIF2B, is linked with significant decrease of GEF activity up to 70%, resulting with leukodystrophies (Fogli et al., 2004). Hence, in future experiments, it will be important to unravel how this EIF2B3 IPA protein variant alters either GEF activity or ribosome assembly, and how this may be relevant to estrogen in breast cancer cells.

CHAPTER 5

REFERENCES

Akman, B. H., Can, T., & E. Erson-Bensan, A. (2012). Estrogen-induced upregulation and 3′-UTR shortening of CDC6. *Nucleic Acids Research*, *40*(21), 10679–10688. https://doi.org/10.1093/nar/gks855

Ali, M. U., Ur Rahman, M. S., Jia, Z., & Jiang, C. (2017). Eukaryotic translation initiation factors and cancer. *Tumor Biology*, *39*(6), 101042831770980. https://doi.org/10.1177/1010428317709805

Amrani, N., Miche`, M., Minet, M., Le Gouar, M., Lacroute, O., & Wyers, O. (1997). Yeast Pab1 Interacts with Rna15 and Participates in the Control of the Poly(A) Tail Length In Vitro. In *MOLECULAR AND CELLULAR BIOLOGY* (Vol. 17). Retrieved from http://mcb.asm.org/

An, J. J., Gharami, K., Liao, G. Y., Woo, N. H., Lau, A. G., Vanevski, F., … Xu, B. (2008). Distinct Role of Long 3′ UTR BDNF mRNA in Spine Morphology and Synaptic Plasticity in Hippocampal Neurons. *Cell*, *134*(1), 175–187. https://doi.org/10.1016/j.cell.2008.05.045

Asano, K. (2000). A multifactor complex of eukaryotic initiation factors, eIF1, eIF2, eIF3, eIF5, and initiator tRNAMet is an important translation initiation intermediate in vivo. *Genes & Development*, *14*(19), 2534–2546. https://doi.org/10.1101/gad.831800

Bartel, D. P. (2009, January 23). MicroRNAs: Target Recognition and Regulatory Functions. *Cell*, Vol. 136, pp. 215–233. https://doi.org/10.1016/j.cell.2009.01.002

Bava, F. A., Eliscovich, C., Ferreira, P. G., Miñana, B., Ben-Dov, C., Guigó, R., …

Méndez, R. (2013). CPEB1 coordinates alternative 3′-UTR formation with translational regulation. *Nature*, *495*(7439), 121–125. https://doi.org/10.1038/nature11901

Berg, M. G., Singh, L. N., Younis, I., Liu, Q., Pinto, A. M., Kaida, D., … Dreyfuss, G. (2012). U1 snRNP determines mRNA length and regulates isoform expression. *Cell*, *150*(1), 53–64. https://doi.org/10.1016/j.cell.2012.05.029

Bogorad, A. M., Lin, K. Y., & Marintchev, A. (2017). Novel mechanisms of eIF2B action and regulation by eIF2 phosphorylation. *Nucleic Acids Research*, *45*(20), 11962–11979. https://doi.org/10.1093/nar/gkx845

Bogorad, A. M., Lin, K. Y., & Marintchev, A. (2018). eIF2B Mechanisms of Action and Regulation: A Thermodynamic View. *Biochemistry*, *57*(9), 1426–1435. https://doi.org/10.1021/acs.biochem.7b00957

Bourdeau, V., Deschênes, J., Laperrière, D., Aid, M., White, J. H., & Mader, S. (2008). Mechanisms of primary and secondary estrogen target gene regulation in breast cancer cells. *Nucleic Acids Research*, *36*(1), 76–93. https://doi.org/10.1093/nar/gkm945

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., … Wittwer, C. T. (2009). The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry*, *55*(4), 611–622. https://doi.org/10.1373/clinchem.2008.112797

Bugiani, M., Boor, I., Powers, J. M., Scheper, G. C., & Knaap, M. S. Van Der. (2010). *Leukoencephalopathy With Vanishing White Matter : A Review*. *69*(10), 987–996.

Chan, S., Choi, E., & Shi, Y. (2011). Pre-mRNA 3′-end processing complex assembly and function. *Wiley Interdisciplinary Reviews: RNA*, *2*(3), 321–335. https://doi.org/10.1002/wrna.54

Chan, S. L., Huppertz, I., Yao, C., Weng, L., Moresco, J. J., Yates, J. R., … Shi, Y.

(2014). CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3′ processing. *Genes and Development*, *28*(21), 2370–2380. https://doi.org/10.1101/gad.250993.114

Chang, J. W., Zhang, W., Yeh, H. S., De Jong, E. P., Jun, S., Kim, K. H., … Yong, J. (2015). MRNA 3′-UTR shortening is a molecular signature of mTORC1 activation. *Nature Communications*, *6*. https://doi.org/10.1038/ncomms8218

Cho, P. F., Poulin, F., Cho-park, Y. A., Cho-park, I. B., Chicoine, J. D., Lasko, P., & Sonenberg, N. (2005). *A New Paradigm for Translational Control : Inhibition via 5 □ -3 □ mRNA Tethering by Bicoid and the eIF4E Cognate 4EHP*. *121*, 411–423. https://doi.org/10.1016/j.cell.2005.02.024

Chung, W., Eum, H. H., Lee, H. O., Lee, K. M., Lee, H. B., Kim, K. T., … Park, W. Y. (2017). Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nature Communications*, *8*(May), 1–12. https://doi.org/10.1038/ncomms15081

Cochrane, A. W., & Deeley, R. G. (1984). Estrogen-dependent Modification of Ribosomal Proteins. *The Journal of Biological Chemistry*, *259*(December 25), 15408–15413. https://doi.org/259(24):15408-13.

Colgan, D. F., & Manley, J. L. (1997). *Mechanism and regulation of mRNA polyadenylation*. (212), 2755–2766.

Derti, A., Garrett-Engele, P., MacIsaac, K. D., Stevens, R. C., Sriram, S., Chen, R., … Babak, T. (2012). A quantitative atlas of polyadenylation in five mammals. *Genome Research*, *22*(6), 1173–1183. https://doi.org/10.1101/gr.132563.111

Desterro, J., Bak-Gordon, P., & Carmo-Fonseca, M. (2019). Targeting mRNA processing as an anticancer strategy. *Nature Reviews Drug Discovery*. https://doi.org/10.1038/s41573-019-0042-3

Elkon, R., Ugalde, A. P., & Agami, R. (2013a). Alternative cleavage and polyadenylation: Extent, regulation and function. *Nature Reviews Genetics*,

14(7), 496–506. https://doi.org/10.1038/nrg3482

Elkon, R., Ugalde, A. P., & Agami, R. (2013b). Alternative cleavage and polyadenylation: Extent, regulation and function. *Nature Reviews Genetics*, *14*(7), 496–506. https://doi.org/10.1038/nrg3482

Erson-Bensan, A. E. (2016). Alternative polyadenylation and RNA-binding proteins. *Journal of Molecular Endocrinology*, *57*(2), F29–F34. https://doi.org/10.1530/JME-16-0070

Fogli, A., Schiffmann, R., Hugendubler, L., Combes, P., Bertini, E., Rodriguez, D., … Boespflug-Tanguy, O. (2004). Decreased guanine nucleotide exchange factor activity in elF2B-mutated patients. *European Journal of Human Genetics*, *12*(7), 561–566. https://doi.org/10.1038/sj.ejhg.5201189

Floor, S. N., & Doudna, J. A. (2016). Tunable protein synthesis by transcript isoforms in human cells. *ELife*, *5*, 1–25. https://doi.org/10.7554/eLife.10921

Gomez, E., Mohammad, S. S., & Pavitt, G. D. (2002). Characterization of the minimal catalytic domain within eIF2B: The guanine-nucleotide exchange factor for translation initiation. *EMBO Journal*, *21*(19), 5292–5301. https://doi.org/10.1093/emboj/cdf515

Gordiyenko, Y., Schmidt, C., Jennings, M. D., Matak-Vinkovic, D., Pavitt, G. D., & Robinson, C. V. (2014a). EIF2B is a decameric guanine nucleotide exchange factor with a γ2 ε. *Nature Communications*, *5*(May), 1–12. https://doi.org/10.1038/ncomms4902

Gordiyenko, Y., Schmidt, C., Jennings, M. D., Matak-Vinkovic, D., Pavitt, G. D., & Robinson, C. V. (2014b). EIF2B is a decameric guanine nucleotide exchange factor with a γ2 ε. *Nature Communications*, *5*(May). https://doi.org/10.1038/ncomms4902

Gruber, A. R., Martin, G., Keller, W., & Zavolan, M. (2012). Cleavage factor I m is a key regulator of 3’ UTR length. *RNA Biology*, *9*(12), 1405–1412.

https://doi.org/10.4161/rna.22570

Guo, C., Spinelli, M., Liu, M., Li, Q. Q., & Liang, C. (2016). A Genome-wide Study of "Non-3UTR" Polyadenylation Sites in Arabidopsis thaliana. *Scientific Reports*, *6*(1), 28060. https://doi.org/10.1038/srep28060

Hardy, J. G., & Norbury, C. J. (2016). Cleavage factor Im (CFIm) as a regulator of alternative polyadenylation. *Biochemical Society Transactions*, *44*(4), 1051–1057. https://doi.org/10.1042/BST20160078

Hilgers, V., Lemke, S. B., & Levine, M. (2012). ELAV mediates 3′ UTR extension in the Drosophila nervous system. *Genes and Development*, *26*(20), 2259–2264. https://doi.org/10.1101/gad.199653.112

Hinnebusch, A. G. (1996). 7 Translational Control of GCN4: Gene-specific Regulation by Phosphorylation of elF2. *Cold Spring Harbor Monograph Archive*, *30*(0), 199–244. https://doi.org/10.1101/0.199-244

Jones, T. R., & Cole, M. D. (1987). Rapid cytoplasmic turnover of c-myc mRNA: requirement of the 3' untranslated sequences. *Molecular and Cellular Biology*, *7*(12), 4513–4521. https://doi.org/10.1128/mcb.7.12.4513

Kamieniarz-Gdula, K., & Proudfoot, N. J. (2019). Transcriptional Control by Premature Termination: A Forgotten Mechanism. *Trends in Genetics*, *35*(8), 553–564. https://doi.org/10.1016/j.tig.2019.05.005

Kapasi, P., Chaudhuri, S., Vyas, K., Baus, D., Komar, A. A., Fox, P. L., … Mazumder, B. (2007). *Article L13a Blocks 48S Assembly : Role of a General Initiation Factor in mRNA-Specific Translational Control*. 113–126. https://doi.org/10.1016/j.molcel.2006.11.028

Kapp, L. D., & Lorsch, J. R. (2004). GTP-dependent Recognition of the Methionine Moiety on Initiator tRNA by Translation Factor eIF2. *Journal of Molecular Biology*, *335*(4), 923–936. https://doi.org/10.1016/j.jmb.2003.11.025

Kashiwagi, K., Takahashi, M., Nishimoto, M., Hiyama, T. B., Higo, T., Umehara, T., … Yokoyama, S. (2016). Crystal structure of eukaryotic translation initiation factor 2B. *Nature*, *531*(7592), 122–125. https://doi.org/10.1038/nature16991

Keller, R. W., Kühn, U., Aragón, M., Bornikova, L., Wahle, E., & Bear, D. G. (2000). The nuclear poly(A) binding protein, PABP2, forms an oligomeric particle covering the length of the poly(A) tail. *Journal of Molecular Biology*, *297*(3), 569–583. https://doi.org/10.1006/jmbi.2000.3572

Klann, E., & Dever, T. E. (2004). Biochemical mechanisms for translational regulation in synaptic plasticity. *Nature Reviews Neuroscience*, *5*(12), 931–942. https://doi.org/10.1038/nrn1557

Lee, S. H., Singh, I., Tisdale, S., Abdel-Wahab, O., Leslie, C. S., & Mayr, C. (2018). Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia. *Nature*, *561*(7721), 127–131. https://doi.org/10.1038/s41586-018-0465-8

Lemmon, M. A., & Schlessinger, J. (2010, June). Cell signaling by receptor tyrosine kinases. *Cell*, Vol. 141, pp. 1117–1134. https://doi.org/10.1016/j.cell.2010.06.011

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods*, *25*(4), 402–408. https://doi.org/10.1006/meth.2001.1262

Luo, W., Ji, Z., Pan, Z., You, B., Hoque, M., Li, W., … Tian, B. (2013). The Conserved Intronic Cleavage and Polyadenylation Site of CstF-77 Gene Imparts Control of 3′ End Processing Activity through Feedback Autoregulation and by U1 snRNP. *PLoS Genetics*, *9*(7). https://doi.org/10.1371/journal.pgen.1003613

Marzluff, W. F., Wagner, E. J., & Duronio, R. J. (2008). *Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail William*. *71*(2), 233–236. https://doi.org/10.1038/mp.2011.182.doi

Matoulkova, E., Sommerova, L., Pastorek, M., Vojtesek, B., & Hrstka, R. (2017). Regulation of AGR2 expression via 3'UTR shortening. *Experimental Cell Research*, *356*(1), 40–47. https://doi.org/10.1016/j.yexcr.2017.04.011

Mohammad-Qureshi, S. S., Jennings, M. D., & Pavitt, G. D. (2008). Clues to the mechanism of action of eIF2B, the guanine-nucleotide-exchange factor for translation initiation. *Biochemical Society Transactions*, *36*(4), 658–664. https://doi.org/10.1042/bst0360658

Pan, Z., Zhang, H., Hague, L. K., Lee, J. Y., Lutz, C. S., & Tian, B. (2006). An intronic polyadenylation site in human and mouse CstF-77 genes suggests an evolutionarily conserved regulatory mechanism. *Gene*, *366*(2), 325–334. https://doi.org/10.1016/j.gene.2005.09.024

Pavitt, G. D. (2005). eIF2B, a mediator of general and gene-specific translational control. *Biochemical Society Transactions*, *33*(6), 1487. https://doi.org/10.1042/bst20051487

Pinto, P. A. B., Henriques, T., Freitas, M. O., Martins, T., Domingues, R. G., Wyrzykowska, P. S., … Moreira, A. (2011). RNA polymerase II kinetics in polo polyadenylation signal selection. *EMBO Journal*, *30*(12), 2431–2444. https://doi.org/10.1038/emboj.2011.156

Prall, O. W. ., Rogan, E. M., & Sutherland, R. L. (1998). Estrogen regulation of cell cycle progression in breast cancer cells. *The Journal of Steroid Biochemistry and Molecular Biology*, *65*(1–6), 169–174. https://doi.org/10.1016/S0960-0760(98)00021-1

Proudfoot, N. J. (2011). *Ending the message : poly (A) signals then and now*. 1770–1782. https://doi.org/10.1101/gad.17268411.

Proudfoot, N. J., Furger, A., & Dye, M. J. (2016). Integrating mRNA Processing with Transcription. *Journal of Lightwave Technology*, *34*(13), 3148–3154. https://doi.org/10.1109/JLT.2016.2552719

Richter, J. D., & Sonenberg, N. (2005). Regulation of cap-dependent translation by eIF4E inhibitory proteins. *Nature*, *433*(7025), 477–480. https://doi.org/10.1038/nature03205

Singh, I., Lee, S. H., Sperling, A. S., Samur, M. K., Tai, Y. T., Fulciniti, M., … Leslie, C. S. (2018). Widespread intronic polyadenylation diversifies immune cell transcriptomes. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-04112-z

Szostak, E., & Gebauer, F. (2013). Translational control by 3'-UTR-binding proteins. *Briefings in Functional Genomics*, *12*(1), 58–65. https://doi.org/10.1093/bfgp/els056

Tian, B.; Manley, J. L. (2017). *Alternative polyadenylation of mRNA precursors*. *25*(5), 1032–1057. https://doi.org/10.1111/mec.13536.Application

Tian, B., & Manley, J. L. (2016a). Alternative polyadenylation of mRNA precursors. *Nature Reviews Molecular Cell Biology*, *18*(1), 18–30. https://doi.org/10.1038/nrm.2016.116

Tian, B., & Manley, J. L. (2016b). Alternative polyadenylation of mRNA precursors. *Nature Reviews Molecular Cell Biology*, *18*(1), 18–30. https://doi.org/10.1038/nrm.2016.116

Tian, B., Pan, Z., & Ju, Y. L. (2007). Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. *Genome Research*, *17*(2), 156–165. https://doi.org/10.1101/gr.5532707

Vorlová, S., Rocco, G., LeFave, C. V., Jodelka, F. M., Hess, K., Hastings, M. L., … Cartegni, L. (2011). Induction of Antagonistic Soluble Decoy Receptor Tyrosine Kinases by Intronic PolyA Activation. *Molecular Cell*, *43*(6), 927–939. https://doi.org/10.1016/j.molcel.2011.08.009

West, S., & Proudfoot, N. J. (2008). Human Pcf11 enhances degradation of RNA polymerase II-associated nascent RNA and transcriptional termination. *Nucleic*

Acids Research, *36*(3), 905–914. https://doi.org/10.1093/nar/gkm1112

Wiestner, A., Tehrani, M., Chiorazzi, M., Wright, G., Gibellini, F., Nakayama, K., … Staudt, L. M. (2007). Point mutations and genomic deletions in CCND1 create stable truncated cyclin D1 mRNAs that are associated with increased proliferation rate and shorter survival. *Blood*, *109*(11), 4599–4606. https://doi.org/10.1182/blood-2006-08-039859

Xia, Z., Donehower, L. A., Cooper, T. A., Neilson, J. R., Wheeler, D. A., Wagner, E. J., & Li, W. (2014). Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'2-UTR landscape across seven tumour types. *Nature Communications*, *5*. https://doi.org/10.1038/ncomms6274

Yang, W., & Hinnebusch, A. G. (1996). Identification of a regulatory subcomplex in the guanine nucleotide exchange factor eIF2B that mediates inhibition by phosphorylated eIF2. *Molecular and cellular biology*, *16*(11), 6603-6616.

Zhao, J., & Hyman, L. (1999). *Formation of mRNA 3 J Ends in Eukaryotes : Mechanism , Regulation , and Interrelationships with Other Steps in mRNA Synthesis*. *63*(2), 405–445.

APPENDICES

A. Vector Constructs Used in the Experiments

Invitrogen pCDNA 3.1(-) vectors were used for the overexpression purposes. The map of the vector was given in Figure 0.1.

Figure 0.1. Map of pCDNA 3.1(-) vector

XbaI and HindIII cut sites used for the cloning purposes are shown in the figure.

pMIR-REPORT™ Luciferase vectors were used for the Dual- Luciferase assay. The map of the vector was given in Figure 0.2.

Figure 0.2. Map of pMIR-REPORT™ Luciferase vector

MluI and HindIII cut sites used for the cloning purposes are shown in the figure.

Promega pGEM®-T Easy Vectors were used for all the cloning purposes as a first-hand vector. Before cloning into either pCDNA or pMIR vectors, desired gene parts were first cloned into pGEM-T vectors. The map of the figure is shown in Figure 0.3.

Figure 0.3. Map of pGEM-T Easy Vector

Single -T nucleotides on the middle part of the lacZ gene was used for the cloning purposes.

B. Markers Used in the Experiments

Figure 0.4. PageRuler Prestained Protein Ladder

Thermo Scientific Page Ruler Prestained Protein Ladder covering the range of 10-250 kDa was used for the Western Blotting experiments.

Figure 0.5. GeneRuler™ 100bp Plus DNA Ladder

Thermo Scientific GeneRuler™ 100bp Plus DNA Ladder covering the range of 100-3000bp was used as a marker in the agarose gel electrophoresis procedures.

C. Buffers Recipes for the Experiments

Tris Buffered Saline (10X) Buffer

- 12.1 grams of Tris (20mM) and 40 grams of NaCl2 (137mM) are dissolved in 200Ml of dH_2O, then the volume is completed to 500mL by dH2O.

- pH should be adjusted to 7.6 by adding HCl

- The buffer should be stored at room temperature.

1X TBST Buffer

- 50 mL of 10X TBS is mixed with 450 mL of dH20.

- 500 uL of Tween 20 is added onto the mixture.

- The buffer should be stored at +4°C.

6X Laemmli Buffer

- 5.25 mL of Tris-HCl (at pH 6.8& 1M concentration), 1.54 grams of SDS, 5.4 mL of Glycerol, 750 mL of B-mercaptoethanol and 2 mg Bromophenol blue are dissolved in 15 mL of dH2O.

- Mixture is heated a bit and shaked for easy dissolving.

- The buffer should be stored at -20°C.

10% SDS solution

- 10 grams of SDS is dissolved in 100 mL dH2O.

- Solution should be stored at room temperature.

1M Tris-HCl

- 12 grams of Tris Base is dissolved in 60 mL dH2O, then the pH is adjusted to 6.8 and the volume is completed to 100 mL.

- Solution should be stored at +4°C.

1.5M Tris-HCl

- 18 grams of Tris Base is dissolved in 60 mL dH2O, then the pH is adjusted to 8.8, and the volume is completed to 100 mL.

- Solution should be stored at +4°C.

10 % APS

- 1gram of Ammonium Persulfate is dissolve in 10 mL dH2O.

- The solution is filtered by using 0.2 um filter and stored at -20°C.

Running Buffer (10X)

- 30 grams of Tris (250mM), 180 grams of Glycine (2500mM), and 10 grams of SDS (1%) are dissolved in dH2O, and the volume is completed to 1 liter with dH2O.

- Solution should be stored at room temperature, and it should be diluted to 1X with dH2O upon using it.

Transfer Buffer (10X)

- 30.3 grams of Tris, and 144 grams of Glycine are dissolved in 500 mL of dH2O, and the volume is completed to 1 liter with dH2O.

- Solution should be stored at room temperature.

Transfer Buffer (1X)

- 100 mL of 10X Transfer Buffer, 700 mL of dH2O, and 200 mL of Methanol are mixed.

- Solution should be prepared fresh, and should be stored at +4°C.

Mild Stripping Buffer

- 15 grams of Glycine, 15 grams of SDS, and 10 mL of Tween 20 are added onto 500mL of dH2O, then the volume is completed to 1liter with dH2O.

- pH of the solution should be adjusted to pH of 2.2.

- 50 mL of aliquots are taken and they are heated up to 65°C prior to use.

Blocking Buffer

- 2 grams of either Skim Milk or BSA is dissolved in 4 mL TBST or PBST (components might change according to antibody that is used).

- It should be freshly prepared upon usage.

- Antibody is also dissolved in blocking buffer

D. Validation of lack of DNA contamination on RNA samples

Figure 0.6. PCR with GAPDH primers to check DNA contamination.

35 cycles of PCR with GAPDH primers was performed to check if there is any DNA contamination. MCF7 cDNA was used as a positive control, on the other hand, no template control was the negative control. None of the RNA samples have DNA contamination in the figure.

E. Sequencing Results for the pMIR Clonings

Figure 0.7 shows the alignment of the expected product(query) and the sequencing result (subject) for the pMIR-Hs.533549.1.2.-3'UTR vector.

```
                MluI
                cut
                site        Forward Primer
Query  1        CGCGTGTTCTGAGCAAGTCAGACTCCTTCCTTTTGGCCTCCAAAGCCACAGATGTTGGCC   60
                ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  322      CGCGTGTTCT-AGCAAGTCAGACTCCTTCCTTTTGGCCTCCAAAGCCACAGATGTTGGCC   264

Query  61       GGCCCACCTGTTTAACTCTGTATTTATTTCCCAATAAAGAAGGGCTTCCAAAGGCATGCT   120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  263      GGCCCACCTGTTTAACTCTGTATTTATTTCCCAATAAAGAAGGGCTTCCAAAGGCATGCT   204

Query  121      GGAGACTTGTGGAGCAGTCCAAAGCTCCATGTCAGGTGGGCTCCAGGTGTACACAGTGTA   180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  203      GGAGACTTGTGGAGCAGTCCAAAGCTCCATGTCAGGTGGGCTCCAGGTGTACACAGTGTA   144
                                Reverse Primer          HindIII
                                                        cut site
Query  181      TGTTCATGTGTCATGTGGTAAAGATCATCTGGAGCAAAAGCT   222                PolyA
                ||||||||||||||||||||||||||||||||||||||||                       Site
Sbjct  143      TGTTCATGTGTCATGTGGTAAAGATCATCTGGAGCAAAAGCT   102
```

Figure 0.7. Sequencing result for pMIR-Hs.533549.1.2.-3'UTR vector.

Alignment of the sequencing result and the expected 3'UTR sequence is shown above. MluI cut site incorporated into forward primer, and HindIII cut site incorporated into reverse primer are highlighted by green color. Rest of the primers were highlighted by yellow color. Poly(A) site is highlighted by red color.

Figure 0.8 shows the alignment of the expected product(query) and the sequencing result (subject) for the pMIR-Hs.533549.1.4.-3'UTR vector. Also, Figure 0.9 shows the alignment of the expected product(query) and the sequencing result (subject) for the pMIR-Hs.533549.1.10.-3'UTR vector.

```
                MluI
                cut        Forward Primer
                site      ------------------>
Query   1       CGCGTGTTCTGAGCAAGTCAGACTCCTTCCTTTTGGCCTCCAAAGCCACAGATGTTGGCC   60
                ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   218     CGCGTGT-CTGAGCAAGTCAGACTCCTTCCTTTTGGCCTCCAAAGCCACAGATGTTGGCC   160
                                                        Reverse Primer   HindIII
                                                        <------------------ cut site
Query   61      GGCCCACCTGTTTAACTCTGTATTTATTTCCCAATAAAGAAGGGCTTCCAAAGGCAAGCT   120
                ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   159     GGCCCGCCTGTTTAACTCTGTATTTATTTCCCAATAAAGAAGGGCTTCCAAAGGCAAGCT   100
```

Figure 0.8. Sequencing result for pMIR-Hs.533549.1.4.-3'UTR vector.

Alignment of the sequencing result and the expected 3'UTR sequence is shown above. MluI cut site incorporated into forward primer, and HindIII cut site incorporated into reverse primer are highlighted by green color. Rest of the primers were highlighted by yellow color. Poly(A) site is highlighted by red color.

```
                MluI
                cut        Forward Primer
                site      ------------------>
Query   1       CGCGTTGGAAAAATGCCTTCAAAATTCTGAAGGAAGGTTAATTTTAACCTAAAATTGTAT   60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   232     CGCGTTGGAAAAATGCCTTCAAAATTCTGAAGGAAGGTTAATTTTAACCTAAAATTGTAT   173
                                                        Reverse Primer
                                                        <------------------
Query   61      ACCCAGGCAAAGTGCTAATCAAGAGGGCAATAGAATAAAAACATTTTCAGATATCAAAAA   120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   172     ACCCAGGCAAAGTGCTAATCAAGAGGGCAATAGAATAAAAACATTTTCAGATATCAAAAA   113
                HindIII
                cut site
Query   121     AGCT   124
                ||||
Sbjct   112     AGCT   109
```

Figure 0.9. Sequencing result for pMIR-Hs.533549.1.4.-3'UTR vector.

Alignment of the sequencing result and the expected 3'UTR sequence is shown above. MluI cut site incorporated into forward primer, and HindIII cut site incorporated into reverse primer are highlighted by green color. Rest of the primers were highlighted by yellow color. Poly(A) site is highlighted by red color.

F. Sequencing Results for the pCDNA Clonings

For the sequencing of overexpression vectors having coding sequences of both full and intronic isoform, T7 and BgH reverse primers were used and the sequencing results were aligned according to both reads. Figure 0.10. shows the sequencing result for the pCDNA-Full-HA C terminus vector, while Figure 0.11 shows the sequencing result for pCDNA-Full-HA N terminus vector. Since the sequencing results are too long, first and last parts of the alignment were given below.

```
                XbaI
                cut
                site        Forward Primer
Query  6        TCTAGAAAAATGGAATTTCAAGCAGTAGTGATGGCAGTAGGTGGAGGATCTCGGATGACA  65
                ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  24       TCTAGAAAA-TGGAATTTCAAGCAGTAGTGATGGCAGTAGGTGGAGGATCTCGGATGACA  82

Query  66       GACCTAACTTCCAGCATTCCCAAACCTCTGCTTCCAGTTGGGAACAAACCTTTAATTTGG  125
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  83       GACCTAACTTCCAGCATTCCCAAACCTCTGCTTCCAGTTGGGAACAAACCTTTAATTTGG  142

                                                                Reverse Primer
Query  1312     AAGCCAAAGCTAAACGAGTGAATGAGGTGATCGTGGGGAATGACCAGCTCATGGAGATCT  1371
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  102      AAGCCAAAGCTAAACGAGTGAATGAGGTGATCGTGGGGAATGACCAGCTCATGGAGATCT  43

                                  -HA tag          HindIII
                                                   cut site
Query  1372     ACCCATACGATGTTCCAGATTACGCTTGAAAGCTTA  1407
                |||||||||||||||||||||||||||||||||||
Sbjct  42       ACCCATACGATGTTCCAGATTACGCTTGAAAGCTTA
```

Figure 0.10. Sequencing result for pCDNA-Full-HA C terminus vector.

First and the last part of the sequencing result is shown above. Sequencing was done by using both T7 and BgH Reverse primers, therefore the numbers above are not quite representative for the lengths of the sequences. Both primers are highlighted with yellow color in the figure, and a red arrow. -HA tag is also indicated by purple color, and the cut sites are indicated by green color.

```
                XbaI
                cut
                site                        -HA tag        Forward Primer
                                                          ------------------------------------------------------------>
Query  6        TCTAGAAAAATGTACCCATACGATGTTCCAGATTACGCTGAATTTCAAGCAGTAGTGATG  65
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  24       TCTAGAAAAATGTACCCATACGATGTTCCAGATTACGCTGAATTTCAAGCAGTAGTGATG  83

Query  66       GCAGTAGGTGGAGGATCTCGGATGACAGACCTAACTTCCAGCATTCCCAAACCTCTGCTT  125
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  84       GCAGTAGGTGGAGGATCTCGGATGACAGACCTAACTTCCAGCATTCCCAAACCTCTGCTT  143

Query  1319     TGGAAGTGGCCAGAGGATTGAAGCCAAAGCTAAACGAGTGAATGAGGTGATCGTGGGGAA  1378
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  731      TGGAAGTGGCCAGAGGATTGAAGCCAAAGCTAAACGAGTGAATGAGGTGATCGTGGGGAA  790

                        Reverse Primer      HindIII
                                            cut site
                <------------------------------------
Query  1379     TGACCAGCTCATGGAGATCTGAAAGCTTA  1407
                |||||||||||||||||||||||||||||
Sbjct  791      TGACCAGCTCATGGAGATCTGAAAGCTTA  819
```

Figure 0.11. Sequencing result for pCDNA-Full-HA N terminus vector.

First and the last part of the sequencing result is shown above. Sequencing was done by using both T7 and BgH Reverse primers, therefore the numbers above are not quite representative for the lengths of the sequences. Both primers are highlighted with yellow color in the figure, and a red arrow. -HA tag is also indicated by purple color, and the cut sites are indicated by green color.

Fort he cloning of intronic coding sequence into pCDNA vector, exonized part of the intron was also added to coding sequence. Figure 0.12. shows the sequencing result for the pCDNA-intronic-HA C terminus vector, while Figure 0.13 shows the sequencing result for pCDNA-intronic-HA N terminus vector. Since the sequencing results are too long, first and last parts of the alignment were given below.

```
                XbaI
                cut
                site
                          Forward Primer
Query  6        TCTAGAAAAATGGAATTTCAAGCAGTAGTGATGGCAGTAGGTGGAGGATCTCGGATGACA  65
                ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  32       TCTAGAAAA-TGGAATTTCAAGCAGTAGTGATGGCAGTAGGTGGAGGATCTCGGATGACA  90

Query  66       GACCTAACTTCCAGCATTCCCAAACCTCTGCTTCCAGTTGGGAACAAACCTTTAATTTGG  125
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  91       GACCTAACTTCCAGCATTCCCAAACCTCTGCTTCCAGTTGGGAACAAACCTTTAATTTGG  150

                                                                  Reverse
                                                                  Primer

Query  1181     TTGCCTTCTCATGAACTCAGTCACTGTGGAGGAAGGGTATGTTTCCCCCTGTACCCACTT  1240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  111      TTGCCTTCTCATGAACTCAGTCACTGTGGAGGAAGGGTATGTTTCCCCCTGTACCCACTT  52
                                                          HindIII
                                  -HA tag                 cut site
Query  1241     GAGGCAAAGATACCCATACGATGTTCCAGATTACGCTTAAAAGCTTA  1285
                ||||||||||||||||||||||||||||||||||||||| ||||||||
Sbjct  51       GAGGCAAAGATACCCATACGATGTTCCAGATTACGCTT-AAAGCTTA  10
```

Figure 0.12. Sequencing result for pCDNA-intronic-HA C terminus vector.

First and the last part of the sequencing result is shown above. Sequencing was done by using both T7 and BgH Reverse primers, therefore the numbers above are not quite representative for the lengths of the sequences. Both primers are highlighted with yellow color in the figure, and a red arrow. -HA tag is also indicated by purple color, and the cut sites are indicated by green color.

```
                XbaI cut
                  site                    -HA tag                          Forward Primer
                                                                          ------------------->
Query  6        TCTAGAAAAATGTACCCATACGATGTTCCAGATTACGCTGAATTTCAAGCAGTAGTGATG  65
                ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  23       TCTAGAAAA-TGTACCCATACGATGTTCCAGATTACGCTGAATTTCAAGCAGTAGTGATG  81

Query  66       GCAGTAGGTGGAGGATCTCGGATGACAGACCTAACTTCCAGCATTCCCAAACCTCTGCTT  125
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  82       GCAGTAGGTGGAGGATCTCGGATGACAGACCTAACTTCCAGCATTCCCAAACCTCTGCTT  141
                                                                  Reverse Primer
                                                                  <-------------------
Query  1218     ATGAACTCAGTCACTGTGGAGGAAGGGTATGTTTCCCCCTGTACCCACTTGAGGCAAAGA  1277
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  79       ATGAACTCAGTCACTGTGGAGGAAGGGTATGTTTCCCCCTGTACCCACTTGAGGCAAAGA  20
                   HindIII
                  cut site
Query  1278     TAAAAGCTTA  1285
                ||||||||||
Sbjct  19       TAAAAGCTTA  12
```

Figure 0.13. Sequencing result for pCDNA-intronic-HA N terminus vector.

First and the last part of the sequencing result is shown above. Sequencing was done by using both T7 and BgH Reverse primers, therefore the numbers above are not quite representative for the lengths of the sequences. Both primers are highlighted with yellow color, and a red arrow in the figure. -HA tag is also indicated by purple color, and the cut sites are indicated by green color.