IDENTIFICATION AND CHARACTERIZATION OF AN INTRONIC TRANSCRIPT FOR *IGFBP4*

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

BY

UTKU CEM YILMAZ

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

JANUARY 2023

Approval of the thesis:

IDENTIFICATION AND CHARACTERIZATION OF AN INTRONIC TRANSCRIPT FOR *IGFBP4*

submitted by **Utku Cem Yılmaz** in partial fulfillment of the requirements for the degree of **Master of Science** i**n Molecular Biology and Genetics, Middle East Technical University** by,

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

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

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

Examining Committee Members:

Prof. Dr. Mesut Muyan
Biology, METU _____

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

Prof. Dr. Yusuf Çetin Kocaefe
Medical Biology, Hacettepe University _____

Date: 20.01.2023

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 Last name : Utku Cem Yılmaz

Signature :

ABSTRACT

IDENTIFICATION AND CHARACTERIZATION OF AN INTRONIC TRANSCRIPT FOR *IGFBP4*

Yılmaz, Utku Cem
Master of Science, Molecular Biology and Genetics
Supervisor : Prof. Dr. Ayşe Elif Erson Bensan

January 2023, 66 pages

Alternative polyadenylation is a widespread mRNA processing event affecting the length of the 3'UTR of mRNAs. Alternative use of polyadenylation sites sometimes occurs in terminal exons, proximal exons, or proximal introns, affecting the coding 3'UTR length and even the coding sequence of the mRNA. To identify estrogen (E2) related polyadenylation changes by a 3'-seq experiment, we detected an early intronic polyadenylation site for *IGFBP4* (insulin-like growth factor binding protein 4). In this thesis, I present data to confirm the existence of an intronic transcript that is robustly upregulated in response to E2 exposure in estrogen receptor-positive breast cancer cell lines. The expression pattern for this isoform follows an opposite pattern to that of the full-length isoform. Hence, we focused on the initial characterization of this isoform. I confirmed the 3'-end of the transcript with an intronic novel poly(A) signal. I performed reporter assays to suggest the functionality of the candidate poly(A) signal compared to the poly(A) signal found at the 3'UTR of the canonical isoform. I provide additional insight into the characterization of the intronicly polyadenylated isoform including mRNA stability and coding potential. While the functional role of this isoform is still not clear, these results highlight the

need to detect and investigate different isoforms that may originate from a gene locus.

Keywords: Intronic Polyadenylation, IGFBP4, Alternative Polyadenylation, Breast Cancer

ÖZ

IGFBP4 GENİNE AİT İNTRONİK BİR TRANSKRİPTİN TANIMLANMASI VE KARAKTERİZE EDİLMESİ

Yılmaz, Utku Cem
Yüksek Lisans, Moleküler Biyoloji ve Genetik
Tez Yöneticisi: Prof. Dr. Ayşe Elif Erson Bensan

Ocak 2023, 66 sayfa

Alternatif poliadenilasyon, mRNA'ların 3'UTR'sinin uzunluğunu etkileyen bir mRNA işleme olayıdır. Poliadenilasyon bölgelerinin alternatif kullanımı bazen terminal ekzonlarda, proksimal ekzonlarda veya proksimal intronlarda meydana gelir. Bunun sonucu olarak da mRNA'ların 3'UTR uzunluğu ve kodlama dizisini etkiler. 3'-seq deneyi ile östrojen (E2) ile ilişkili poliadenilasyon değişikliklerini belirlemek için, *IGFBP4* için erken bir intronik poliadenilasyon bölgesi saptadık. Bu tezde, östrojen reseptörü pozitif meme kanseri hücre hatlarında E2 maruziyetine yanıt olarak güçlü bir şekilde ifadesi pozitif yönde etkilenen intronik bir transkriptin varlığını doğrulamak için veriler sunuyorum. Bu izoformun ifade paterni, tam uzunluktaki izoformunkine zıt bir paterni takip ettiğini gözlemledik. Bu nedenle, ilk olarak bu izoformun karakterizasyonuna odaklandık. Transkriptin 3'-ucunu bir intronik yeni poli(A) sinyaliyle onayladım. Kanonik izoformun 3'UTR'sinde bulunan poli(A) sinyaline kıyasla aday poli(A) sinyalinin işlevselliğini test etmek için raportör tahlilleri yaptım. mRNA stabilitesi ve kodlama potansiyeli dahil olmak üzere intronik olarak poliadenilatlanmış izoformun karakterizasyonuna ilişkin ek bilgiler sağlıyorum. Bu izoformun fonksiyonel rolü hala net olmasa da, bu sonuçlar

bir gen lokusundan kaynaklanabilecek farklı izoformların saptanması ve araştırılması ihtiyacını vurgulamaktadır.

Anahtar Kelimeler: İntronik Poliadenilasyon, IGFBP4, Alternatif Poliadenilasyon, Meme Kanseri

To my beloved family

ACKNOWLEDGMENTS

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. Mesut Muyan for his precious guidance. Also, I would like to thank my thesis committee member, Prof. Dr. Yusuf Çetin Kocaefe for his valuable comments.

I am grateful to all my fellow lab mates; Elanur Almeriç, İrem Eroğlu, İbrahim Özgül, Murat Erdem, Didem Naz Döken, Irmak Gürcüoğlu and Deniz Karagözoğlu for all stimulating discussions, feedbacks, and for all the fun we have had in the last three years. Special thanks to Ayça Çırçır, who is the best lab partner ever, besides for being the person that thought me everything, for her precious friendship and invaluable support all the time.

Lastly, I wish to express my deepest gratitude to my mother, Gülüşen Yılmaz, and my father, Atilla Cengiz Yılmaz whose support, love and guidance were with me all the time.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

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

IPA: Intronic Polyadenylation

PAS: Polyadenylation Signal

RBP: RNA Binding Protein

CTD: Carboxy Terminal Domain

SDM: Site Directed Mutagenesis

LIST OF SYMBOLS

SYMBOLS

IGFBP4: Insulin-like Growth Factor Binding Protein 4

TFF1: Trefoil Factor 1

GAPDH: Glyceraldehyde-3-Phosphate Dehydrogenase

RPLP0: Ribosomal Protein Lateral Stalk Subunit P0

CHAPTER 1

INTRODUCTION

1.1 Polyadenylation

There are three major steps in eukaryotic mRNA maturation, namely, 5' m^7G capping, splicing, and polyadenylation. Eukaryotic mRNAs undergo polyadenylation to get a stretch of adenine nucleotides at their 3'UTR, which is known to determine the mRNA's fate in several ways (Wilton et al., 2021). This process is carried out by the cleavage and polyadenylation (CP) machinery.

Polyadenylation occurs after recognizing a hexameric poly(A) signal, flanking upstream (U-rich elements and UGUA motif) and downstream elements by the CP machinery. Consequently, the poly(A) tail is added to the 3'end of the transcript by poly(A) polymerase (Tian & Manley, 2016).

Polyadenylation usually adds around 250 bases of adenine nucleotides to the 3'-end of transcripts. As this process creates a barrier that protects the mRNAs from exosome nuclease and extends its half-life, it also plays a crucial role in localization within the cell, translation, and stability (Kühn et al., 2009).

Several cis-elements and complex machinery composed of RBPs coordinate the process of polyadenylation. Cis-elements that contribute to this process are hexameric poly(A) signal, USEs (U- and UGUA rich elements), and DSEs (U- and GU- rich elements). The major factors involved in the cleavage and polyadenylation machinery are; cleavage stimulatory factor (CSTF), cleavage, polyadenylation specificity factor (CPSF), and cleavage factors CFIm, CFIIm and poly(A) polymerase (PAP) (Tian & Manley, 2016). Besides these elements, 3'-end processing machinery is quite complex, comprising around 85 proteins (Shi et al., 2009).

After transcription initiation, the carboxyl-terminal domain of RNA Pol II gets differentially phosphorylated on serine residues. This phosphorylation is known to induce the recruitment of cleavage and polyadenylation machinery to the mRNA 3'-ends. RNA Pol II transcribing through a functional polyadenylation site (PAS) is known to have a detractive effect on elongation (Wilton et al., 2021). Furthermore, CPSF and CstF factors induce a lagging effect on elongation, which after that, causes the gradual dissociation of Pol II from the DNA template (Zhang, 2015). Conserved sequence elements also have a high impact on the positioning of these RBPs. Upstream Sequence Elements (USEs), U-rich elements, Downstream Sequence Elements (DSEs), UG-rich motifs, and PAS in between these two motifs are directly recognized by RBPs.

General consensus canonical sequences and their positions are shown in Figure 1.2. In addition to the USEs, DSEs and PAS, there is a cleavage site of 15-30 nucleotides downstream of PAS (Laishram, 2014).

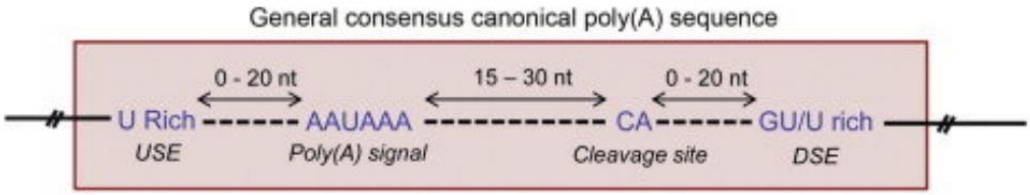


Figure 1.1. Cis-elements recognized by CP machinery and their estimated positions on RNA (Laishram, 2014).

Other structure-based properties of chromatin, such as heterochromatin, may also induce transcription termination. These structure-based blockages on Pol II movement through DNA template may cause recognition of a cryptic poly(A) signal due to slowed elongation (Wilton et al., 2021). Following the recognition of PAS by Pol II and reduced elongation rate, RNA is cleaved 10-30 nucleotides downstream of recognized PAS. Simultaneously, poly(A) polymerase (PAP) adds a short stretch of adenine nucleotides to the 3'-end of nascent RNA. Nuclear poly(A) binding protein (PABPN) can bind to this 11-14 nucleotide stretch of adenines and increases

the efficiency of PAP, expanding the synthesis of poly(A) tail up to 200-250 nucleotides (Nicholson & Pasquinelli, 2018).

1.2 Alternative Polyadenilation

Functional poly(A) signals defined in the previous section can be located at different locations on the gene. Seventy-four percent of all human genes contain more than one PAS (Wilton et al., 2021). This multiplicity of poly(A) signals on the genome, which is mainly on the 3'UTR of the genes, creates an environment for the alternation of the transcription termination sites. This mechanism of alternation among different poly(A) signals is called "alternative polyadenylation (APA)". As a result of APA, transcripts with different 3'UTR lengths (3'UTR-APA) and proteins with various functions after a change in their coding sequence (CDS) can be produced if the poly(A) sites reside in proximal introns or exons (Derti et al. 2012). Cis-elements and their cognate RBP are shown in Figure 1.2.

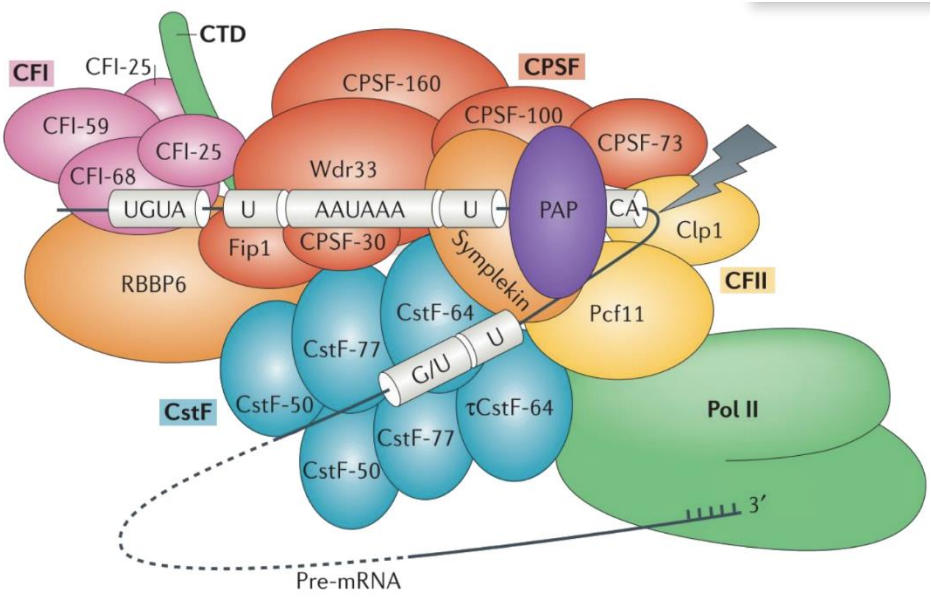


Figure 1.2. Elements involved in the cleavage and polyadenylation machinery (Tian & Manley, 2016).

1.2.1 Types of APA

The presence of functional PASs on precursor mRNA can lead to the formation of distinct isoforms. Hence, APA is a process that results with different transcripts from the same gene that differ at their 3'ends. This process adds another layer to gene regulation. APA can be examined under two main categories according to the position of the PAS on the gene, namely, 3'UTR-APAs, which creates a diversity on isoform's 3'UTR lengths and upstream region APA (UR-APAs) which occurs at the upstream of the exon. UR-APA events may alter the C-terminus of the protein and hence protein function (Turner et al., 2018).

3'UTR-APA is a widespread event in cancer where isoforms are shortened by this process (Park et al., 2018). Variation of the 3'UTR length of transcripts affects the number of possible miRNA binding sites and RBPs, which consequently has an impact on mRNA stability, nuclear export, intracellular localization, and translation efficiency (Zhang et al., 2021). The change in 3'UTR length induced variation of miRNA binding cis-elements and secondary structures on RNA created by RBPs can change mRNA's fate, affecting gene regulation by influencing its recognition by post-transcriptional factors (Lembo et al., 2012).

UR-APA can be divided into three subclasses: splicing APA, intronic APA and internal exon APA, as seen in Figure 1.3.

Intronic-APA among UR-APA types can also function in the repression of gene expression by creating a feedback loop to minimize the full-length transcript levels in the cell (Tian, Manley, 2016).

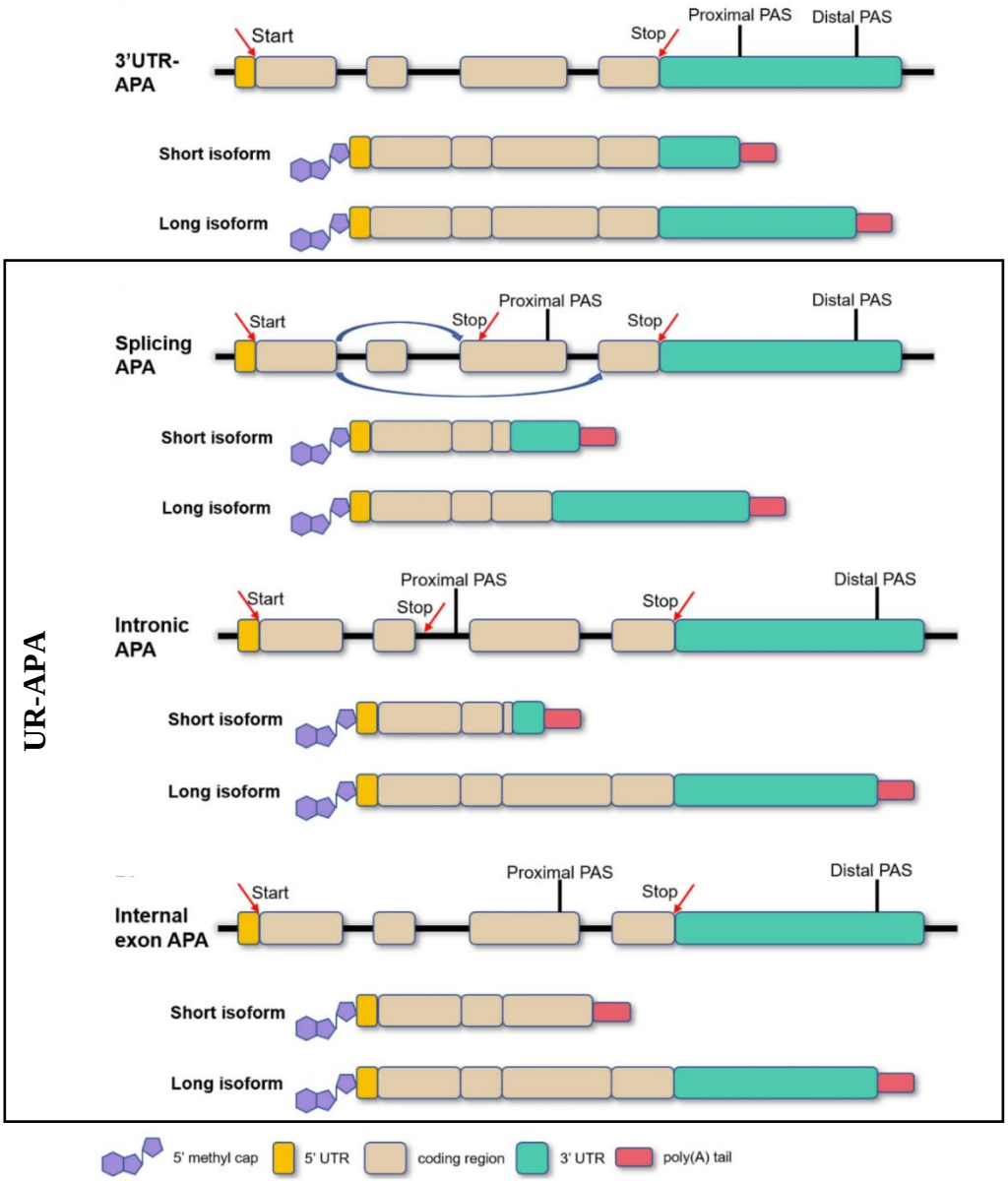


Figure 1.3. Categorization of alternative polyadenylation types (Zhang et al., 2021).

1.2.2 APA in Cancer

APA is a widespread regulation of gene expression that results in diverse consequences from transcripts stability and localization to the protein to be translated from that transcript. Recent analyses of Pan-cancer regulation of APA in cancer revealed that APA events are not random; there is a conserved pattern (Venkat et al., 2020).

With next-generation sequencing technologies, the regulation of APA patterns in cancer compared to healthy tissues is being examined in more detail. It was discovered that cancer genes related to cell proliferation undergo 3'UTR shortening, which eliminates the target regions of several miRNAs and extends the mRNA's half-life. The result of global analysis on examining APAs suggested that 3'UTR events on proto-oncogenes extend these mRNA's half-life and have a positive contribution to cancer progression, and CDS-APA events on tumor suppressors can repress the expression of these genes and promote cancer cell progression even more (Yuan et al., 2019).

In this thesis, I focused on an intronic polyadenylation event in Insulin-like growth factor binding protein 4 (IGFBP4) in breast cancer cell line models.

1.3 Insulin-like Growth Factor Binding Proteins

Insulin-like growth factor binding proteins (IGFBPs) are a group of proteins that bind to insulin-like growth factors (IGFs) and regulate their activity by inhibiting their interaction with their corresponding receptor upon bound. IGFs are hormones involved in a wide range of physiological processes, including cell proliferation, differentiation, and apoptosis (Jones and Clemmons, 1995). The binding of IGFBPs to these hormones can interfere with their activity. There are six known IGFBPs (Allard & Duan, 2018). Considering their ability to inhibit the binding of a growth factor which consequently disables the following processes such as cell proliferation, some IGFBPs are reported as tumor suppressors. Therefore, besides the IGF

concentration in bloodstream and cell surface receptor density for IGFs, IGFBPs add another layer of modulation to the signaling activity of IGFs (Hjortebjerg, 2018). It is also reported that IGFBP binding to IGFs prolong the half-life of the ligand, creating a constantly available ligand pool (Allard and Duan, 2018). Among IGFBP family, especially *IGFBP4* is a highly studied gene in this context. In a recent study, silencing of the tumor suppressor resulted in a significant upregulation of EZH2- a histone methyltransferase- and poor survivals of hepatocellular carcinoma patients (Lee et al., 2018).

1.3.1 Insulin-like Growth Factor Binding Protein 4

IGFBP4 is the smallest member of the IGFBP family. It is a dimeric glycoprotein consisting of N-terminal and C-terminal connected by a linker region. Protein can be found in both glycosylated and non-glycosylated forms, which does not affect the binding affinity to IGFs (Konev et al., 2015). In all IGFBPs there are conserved Cys residues. In IGFBP4 there are two extra conserved Cys residues (Rajaram, 1997). It was shown that the Cys residues are important for IGF binding properties of IGFBP4 and the Cys residues at the hydrophobic binding motif of N-terminal are more crucial for efficient binding to IGF (Byun et al., 2001).

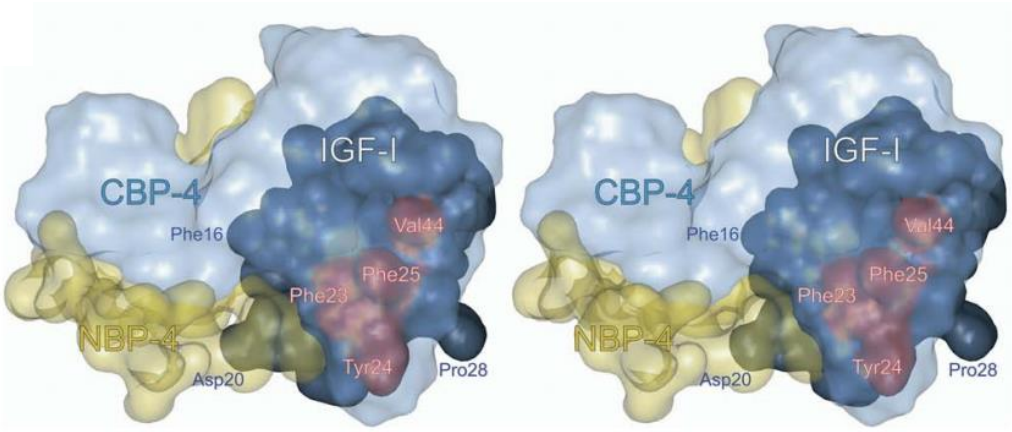


Figure 1.4. Structure of IGFBP4 interaction with IGF (Siwanowicz et al., 2005).

Figure illustrates the IGFBP4 interaction with IGF. N-terminal (NBP-4) and C-terminal (CBP-4) of IGFBP4 were indicated separately.

IGFBP4 is found in all biological fluids and expressed in a number of organs (Hjortebjerg, 2018; Zhou et al., 2003). Among all IGFBPs, IGFBP4 is the only one that has no report on positive mitogenic functions, it only has inhibitory functions (Wetterau et al., 1999). Supportingly, overexpression of the protein in mice resulted with several growth defects on organs. (Schneider et al., 2000; Zhou et al., 2004; Ning et al., 2008).

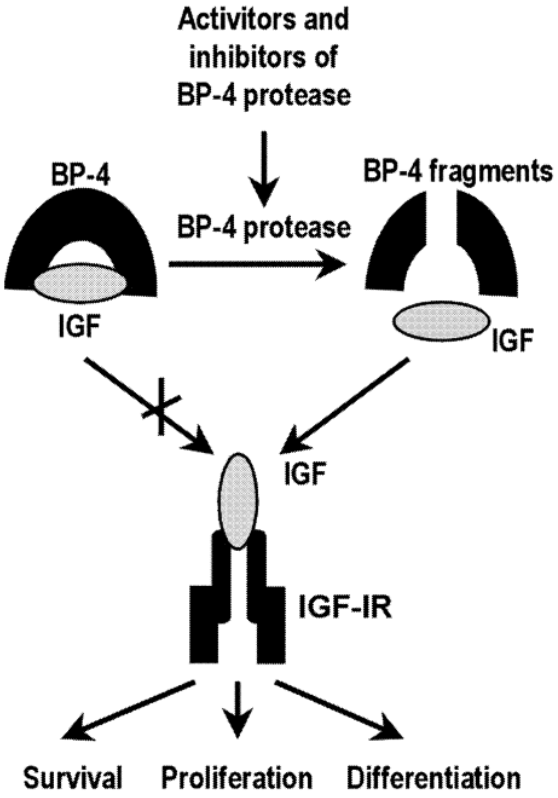


Figure 1.5.Inhibiton of the mitogenic activity of IGF-IR upon IGFBP4 binding (Zhou et al., 2003).

1.4 Aim of Study

A 3’-end specific RNA-sequencing experiment revealed differential usage of polyadenylation sites upon E2 treatment in a time-course experiment in MCF7 cells. The sequencing data showed a slow but dynamic upregulation of the full-length *IGFBP4* transcript implicated by increased usage of a 3’UTR polyA site, whereas a novel intronic polyadenylation site was more robustly upregulated in response to E2. In this thesis, I aimed to characterize this novel intronic polyadenylation event for *IGFBP4*.

CHAPTER 2

MATERIAL AND METHODS

2.1 Databases

PolyA_DB_2 database was used to examine the poly(A) sites on mRNAs. https://exon.apps.wistar.org/polya_db/

UCSC Genome Browser and Human Feb. 2009 (GRCh37/hg19) Assembly was used to show introns and exons genomic location and to image Chip-seq Data retrieved from Cistrome. https://genome.ucsc.edu/

2.2 Gene expression and CHIP Datasets

Gene Expression Omnibus (GEO) database was used to examine gene expression profiles of different mRNAs under different treatments by using different RNA sequencing data. https://www.ncbi.nlm.nih.gov/geo/

Cisrome Data Browser was used to check ChIP-seq experiments under different E2 treatment time points.

GSE94023 is a H3K4me3 ChIP-seq dataset performed after 40 minutes 10 nm E2 treatment to MCF7 cells. Not-treated and 40-minute E2-treated samples were included to the analysis.

GSE57436 is a H3K4me3 ChIP-seq dataset performed after 30 minutes of 100 nm E2 treatment to MCF7 cells.

GSE120756 is a H3K4me3 ChIP-seq dataset performed after 45 minutes of 10 nm E2 treatment to MCF7 cells.

GSE57436 is a H3K27ac ChIP-seq dataset performed after 30 minutes of 100 nm E2 treatment to MCF7 cells. Time 0 and 30-minute treated samples were included to the analysis.

GSE78913 is a H3K27ac ChIP-seq dataset performed after 45 minutes of 100 nm E2 treatment to MCF7 cells.

2.3 Cell Lines and Cell Culture

MCF7 and T47D cells were maintained in high glucose containing Dulbecco's Modified Eagle's Medium (4500 mg/L Glucose). Additionally, sodium pyruvate (final concentration: 1mM), L-glutamine (final concentration: 2mM), penicillin-streptomycin (final volume: 1%), and 10% Fetal Bovine Serum (FBS) was supplemented into the medium. To protect cells against mycoplasma contamination, cells were also treated with Plasmocin. Cells were cultured as monolayers in T25 flasks and initially later to be removed into T75 flasks, which were incubated in CO_2 (5%) incubators at 37°C. To obtain stocks from the cell line, cell pellets were taken after 70-80% confluency was observed on flasks under a microscope, and the pellet was resuspended in the previously described medium, which contained 5% DMSO (dimethyl sulfoxide). Cells were transferred into cryovials in liquid nitrogen for long-term preservation.

2.4 Estradiol Treatment

Estradiol (E2) treatment was applied to MCF7 and T47D cells with 60-70% confluency in T75 flasks. Before treatment, cells were washed with PBS. Then cells were hormone deprived for 72 hours. The hormone deprivation medium contained phenol-red free DMEM, 10% dextran-coated charcoal-stripped FBS, 1% P/S, and 1% L-glutamine. MCF7 cells were then treated with 100 nM 17 β-Estradiol (Sigma-Aldrich, CAT#E2257) and 100 nM Ethanol as vehicle control. Treatment was performed for 45 minutes, 3 hours, and 12 hours. At each time point, pellets were

obtained from corresponding E2 or ethanol treated cells. The E2 treatment was tested by examining *TFF1* expression. *TFF1* is a known E2-responsive gene (Carroll and Brown, 2006).

2.5 Actinomycin D Treatment

MCF7 cells in T75 flasks with 50-60% confluency were treated with Actinomycin D (Tocris Bioscience, CAT#1229). Before Actinomycin D treatment, cells were treated with 100 nM E2 for 45 minutes to capture the intronic transcript. Following E2 treatment, cells were treated with 2 μg/mL Actinomycin D or DMSO as vehicle control for 0, 30 minutes, and 12 hours.

2.6 RNA Isolation and DNase Treatment

Total RNA was isolated using a High Pure RNA Isolation Kit (Roche, CAT#11828665001). RNA samples were always checked for DNA contamination by PCR using *GAPDH* primers (the sequence is given in Table 2.1). Any DNA contamination was eliminated by treating RNA samples with DNase Recombinant Enzyme (Thermo Fisher Scientific, CAT#EN0521) by using the steps described in the manufacturer's protocol. (100 ul reaction containing 10% DNase Recombinant Enzyme by volume).

Sample concentrations and purity levels were calculated using Nanodrop (Maestrogen).

Table 2.1. List of primers

Primer Name	Primer Sequence (5’ to 3’)	Experiments that are used
GAPDH_F	GGGAGCCAAAAGGGTCATCA	PCR
GAPDH_R	TTTCTAGACGGCAGGTCAGGT	PCR
TFF1_F	TTGTGGTTTTCCTGGTGTCA	RT-qPCR
TFF1_R	CCGAGCTCTGGGACTAATCA	RT-qPCR
RPLP0_F	GGAGAAACTGCTGCCTCATA	RT-qPCR
RPLP0_R	GGAAAAAGGAGGTCTTCTCG	RT-qPCR
IGFBP4_UTR_F	GTCTGAGCCCTGGTGTGTTT	RT-qPCR
IGFBP4_UTR_R	CCCACAGGCTTGAACTCTCC	RT-qPCR
IGFBP4_Intron_F	TGGGTAGGGAGAGATGGGTC	RT-qPCR
IGFBP4_Intron_R	TCTGCCCCAAGATTAGCGAC	RT-qPCR
IGFBP4_Intron_3’RACE_F_1	GTCGCTAATCTTGGGGCAGA	3’RACE
IGFBP4_Intron_3’RACE_F2	CCTCCTCCTCTCTCAGCACT	3’RACE
IGFBP4_ Poly(A)_cloning Hs.462998.1.9_ΔpApMIR_F	CGCATGAGCTCTTAGGAACCTACCAGTTGGC	pMIR Cloning
IGFBP4_ Poly(A)_cloning Hs.462998.1.9_ΔpApMIR_R	CGCATACGCGTGCATCCCTGTGTCTAACTGAGAA	pMIR Cloning
IGFBP4_intronic Poly(A)_cloning_nested_product_F	GGAGGGCATGGCATGAGAAT	pMIR Cloning

IGFBP4_intronic Poly(A)_cloning_nested_product_R	ACCTAAAGCCACCCCATTCC	pMIR Cloning
IGFBP4_intronic Poly(A)_cloning_ΔpApMIR_F	CGCATGAGCTCGGTGTGGAGGTTCATGCTTG	pMIR Cloning
IGFBP4_intronic Poly(A)_cloning_ΔpApMIR_R	CGCATACGCGTCATGAGCCACTGCAGTTGCC	pMIR Cloning
IGFBP4_intronic Poly(A)_site_directed_mutation_ΔpApMIR_F	GGGTCTGGCTTTATTGCTCAGGCTGGTCTC	pMIR Site Directed Mutation
IGFBP4_intronic Poly(A)_site_directed_mutation_ΔpApMIR_R	GAGACCAGCCTGAGCAATAAAGCCAGACCC	pMIR Site Directed Mutation
3'RACE_OligodT	GACCACGCGATCGATTGACTTTTTTTTTTTTTTV	3'RACE
Anchor_Primer_R	GACCACGCGTATCGATGTCGAC	3'RACE
T7	TAATACGACTCACTATAGGG	Sequencing
BGH/pcDNA3.1_R	TAGAAGGCACAGTCGAGG	Sequencing
pMIR_Sequencing_F	AGGCGATTAAGTTGGGTA	Sequencing
pMIR_Sequencing_R	GGAAAGTCCAAATTGCTC	Sequencing

2.7 cDNA Synthesis

cDNA synthesis was performed using RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, CAT# EP0441) as shown in Table 2.2. Random Hexamer primer was used for the validation of transcripts (see Appendix B).

Table 2.2. Reaction Conditions for cDNA synthesis

Components	Amounts
RNA	1µg
Oligo(dT)primer (100µM) Or Random Hexamer Primer (100µM)	1µL
Nuclease-free Water	Up to 12 µL
*Incubation at 70°C for 5 minutes. *Spin-down *Incubation on ice for 1 minute*	
5X Reaction Buffer. (250mM Tris-HCl (pH: 8.3), 250mM KCl, 20mM MgCl2, 50mM DTT)	4µL
dNTP Mix (10mM)	2µL
RiboLock RNase Inhibitor (20U/µL)	1µL
RevertAid Reverse Transcriptase (200U/µL)	1µL
*Incubation at 42°C for 60 minutes. *Reaction inhibition by incubating at 70°C for 5 minutes.	

The quality of the obtained cDNAs were validated by PCR using *GAPDH* primers.

2.8 RT-qPCR

Real-Time Quantitative PCR (RT-qPCR) was performed using the QIAGEN-Rotor-GeneQ detection system. For 10 µL reactions, BioRAD SsoAdvanced Universal SYBR® Green Supermix (CAT#1725270), which contains both reverse transcriptase and hot-start DNA Polymerase was used. Primers shown in Table 2.1 were used normalization for each sample was done by using the Ct value of the

housekeeping gene *RPLP0* for the corresponding samples. ΔΔCq was calculated to determine fold changes in expression levels. Melt peaks were also examined to check the specificity of the reaction. Every process during RT-qPCR was done by following MIQE Guidelines (Bustin et al., 2009).

IGFBP4 full transcripts were amplified with IGFBP4_UTR_F, IGFBP4_UTR_R (products size: 199 bp, annealing temperature: 55°C) and IGFBP4 intronic transcripts were amplified with IGFBP4_Intron_F, IGFBP4_Intron_R (product size: 113bp, annealing temperature: 61°C).

The expression of an estrogen-responsive gene, *TFF1*, was used to validate estradiol treatment success. TFF1 was amplified by using TFF1_F, TFF1_R (products size: 209bp, annealing temperature: 56°C).

RPLP0 was amplified by using RPLP0_F, RPLP0_R (product size: 191bp, annealing temperature: 60°C).

2.9 3’ Rapid Amplification of cDNA Ends

cDNAs for 3’RACE that contain an anchor sequence at the 3’-end were synthesized by using RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, CAT# EP0441) and oligo dT-anchor primer with 5µg total RNA of 45 minutes E2 treated MCF7 cells. For nested PCR, a reverse primer specific to the anchor sequence at the 3’-end was used for both rounds. Two rounds of nested PCR were performed. For the first round, IGFBP4_Intron_3’RACE_F_1 was used as forward primer together with anchor primer in the following conditions: 95°C for 3 minutes, 34 cycles of 95°C for 30 seconds, 56°C for 30 seconds, 72°C for 30 seconds and lastly 72°C for 10 minutes. Expected product sizes for the first round of 3’RACE is given in Table 2.3.

Table 2.3. Expected Product Sizes For The First Round of 3'RACE

PolyA Site ID	Expected Product Size
Not-Identified (1st peak)	498bp
Not-Identified (2nd peak)	640bp
Not-Identified (3rd peak)	971bp

For the second round, IGFBP4_Intron_3'RACE_F_2 was used as forward primer with the anchor reverse primer by using the 1/5 diluted PCR products of the first 3'RACE. Conditions of the second round PCR were as follows: 95°C for 3 minutes, 34 cycles of 95°C for 30 seconds, 56°C for 30 seconds, 72°C for 30 seconds, and 72°C for 10 minutes. Expected product sizes after the second round of 3'RACE are given in Table 2.4.

Table 2.4. Expected Product Sizes For The Second Round of 3'RACE

PolyA Site ID	Expected Product Size
novel (1st peak)	322bp
novel (2nd peak)	464bp
novel (3rd peak)	795bp

After serial PCR reactions, second round 3'RACE products were loaded into 1% Agarose gel. The desired band around 322bp was extracted by using Zymoclean™ Gel DNA Recovery Kit (CAT#D4008). The purity and concentration of the gel extracts were measured by Nanodrop (Maestrogen).

2.10 Cloning of 3'RACE products into pGEM-T Vectors

The expected band from 3'RACE around 322 bp was cut and extracted from agarose by using the DNA recovery kit mentioned above. The sample was cloned into 50 ng

pGEM®-T Easy Vector (Promega, CAT#A1360). Insert and linear vector were ligated to each other by using T4 DNA ligase (Promega, CAT#A1360) at 4°C for 16 hours. For the ligation of the vector and insert with known concentrations and sizes, the following ratio was used:

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

$$Volume\ of\ insert\ (\mu l) = \frac{\text{Insert (ng)}}{\text{Measured concentration of extraction product } \left(\frac{\text{ng}}{\mu\text{l}}\right)}$$

Following ligation, products were transformed into *E. coli* cells. Observed colonies on Ampicillin-containing agar plates were tested with colony PCR using IGFBP4_Intron_3'RACE_F_2 and reverse anchor primer. Plasmids from positive colonies were isolated and sent for sequencing.

2.11 Cloning of Intronic and UTR Poly(A) Signals into ΔpA-pMIR-Report Luciferase Vectors

Reported poly(A) signal for the canonical isoform of *IGFBP4* and identified potential poly(A) signal of the intronic isoform of *IGFBP4* were cloned into ΔpA-pMIR-Report Luciferase Vectors. ΔpA-pMIR-Report Luciferase Vector is a construct designed in our laboratory by Elanur Almeriç, which lacks the SV40 poly(A) signal at downstream of the Luciferase Reporter gene. *MluI* (Forward) and *SacI* (Reverse) cut sites were added to the 5'-end of gene-specific primers (Figure 2.1). PCR reactions with 50 µl total volume were prepared. For intronic poly(A) signal cloning 45-minute E2-treated MCF7 cDNA was used. For canonical isoform poly(A) signal cloning, 12-hour E2-treated MCF7 cDNA was used as a template. The intronic site aimed to be cloned was located in a short interspersed nuclear elements (SINE). SINE, a class of transposable elements are found at high copy numbers in human genome (Deininger, 2011; Weiner, 2002). Therefore, to eliminate the non-specific annealing to these elements on the genome with high copy number,

the intronic poly(A) signal along with estimated flanking sequences (~50bp upstream and downstream of poly(A) signal) PCR product was produced in two rounds with nested PCR using the primers given in Table 2.1. Following PCR, products were loaded into agarose gel, and desired bands were extracted from the gel. Extracted products and ΔpA-pMIR-Report Luciferase Vector were then digested by using *MluI* and *SacI* enzymes at 37°C for 90 minutes. After running the digestion products on agarose gel, bands were extracted from the gel by using a gel recovery kit. Products were then ligated to each other using the T4 DNA Ligase enzyme by using the given equation above. Ligation products were transformed into *E. coli* cells and seeded on a medium containing Ampicillin. Colonies were used for plasmid isolation and sent to sequencing with pMIR-specific sequencing primers (pMIR_Sequencing_F, pMIR_Sequencing_R).

2.12 Site-Directed Mutagenesis

Intronic potential poly(A) signal sequence (AATATA) was converted into canonical poly(A) signal (AATAAA) with a single site-directed mutagenesis reaction.

For this purpose, QuickChange Primer Design-Agilent (https://www.agilent.com/store/primerDesignProgram.jsp) was used to design the primers shown in Table 2.1.

300 ng IGFBP4 intronic Poly(A) containing ΔpApMIR, 5µl of each primer (10µM) (IGFBP4_intronic_Poly(A)_site_directed_mutation_ΔpApMIR_F and IGFBP4_intronic_Poly(A)_site_directed_mutation_ΔpApMIR_R), 5 µl of 2 mM dNTPs, and 0.5 µl Phusion DNA Polymerase (2 U/µl) were mixed with 5X HF buffer and reaction volume was completed to 50µl by adding nuclease-free H_2O. PCR conditions for the reaction were as follows: 98°C for 3 minutes as initial denaturation, followed by 15 cycles of 98°C 30 seconds, 65°C 30 seconds, and 72°C for 3.5 minutes. To check whether the newly synthesized plasmid is still intact, PCR product was run on 0.8% agarose gel at 100 volts for 1.5-2 hours.

Then the PCR product was treated with *DpnI* restriction enzyme (FD1704, Thermoscientific). The reaction was prepared by mixing 10 µL PCR product, 4 µL of 10X FastDigest buffer, 23 µL of nuclease-free H_2O, and 3 µL *DpnI* enzyme. The reaction was prepared in a 0.2 mL tube and incubated overnight at 16°C. *DpnI*-treated samples were transformed into *E.coli* cells. SDM was confirmed by sequencing.

2.13 Dual-Luciferase Assay

MCF7 cells were co-transfected with phRL-TK (Renilla Luciferase) (25ng) and pMIR-Report Luciferase (Firefly) (225ng) 48-well cell culture plate using Turbofect Transfection Reagent. Following 24-hour incubation after transfection, cells were lysed by following the instructions described in manufacturer's instructions for Dual-Luciferase® Reporter Assay System (Promega, CAT# E1910). Luminesance of the samples were measured by using Modulus Microplate Luminometer (Turner Biosystems). Obtained luminesecence reads of Firefly were then normalized to those of Renilla.

2.14 Transfections

Transfections were performed using Turbofect Transfection Reagent (Thermo Fisher Scientific, CAT#053) by following the described steps given by the manufacturer. For the transfection with pcDNA3.1(-), MCF7 cells were grown up to a confluency of 50-60% in 6-well cell culture plates. 2 µg/mL plasmids were mixed with 200 µl serum-free DMEM and 4 µl transfection reagent. The transfection mixture was incubated for 30 minutes and slowly added to each well. Next, cells were incubated at 37°C for 24 hours in a CO_2 incubator.

2.15 Cloning of Coding Sequences into pcDNA 3.1 (-) Vectors

For the cloning of coding sequences of *IGFBP4* intronic and full isoforms, specific primers that flank the desired region with proper restriction sites for the directionality were designed by excluding the start codon (ATG) of the open reading frame. The *NotI* enzyme cut site was incorporated into the forward primer, while the *EcoRI* enzyme cut site was incorporated into the reverse primer. Primers were designed to have 3'UTRs of both *IGFBP4* intronic and the canonical isoform. The random sequence 'CGTCTA' was added at the 5'-end of all primers to increase the endonuclease activity of the selected restriction enzymes. All cloning steps for pcDNA3.1(-) were the same with poly(A) signal cloning into ΔpA-pMIR vectors except the initial PCR performed to amplify the insert, instead of Taq Polymerase (Thermo Scientific, CAT# EPEP0401) Phusion Polymerase Enzyme (Thermo Scientific, CAT# F530S) was used. 45-minute E2 treated and 12-hour E2 treated MCF7 cell cDNAs were used as the template for insert amplification of *IGFBP4* intronic and full isoform, respectively. Plasmids and inserts were double digested with NotI and EcoRI at 37°C for 1.5 hours then digestion products were run on 1% agarose gel. After confirming the successful digestion of samples, corresponding double-digestion bands were cut and extracted from the gel. Then digested intronic and full isoforms were ligated into digested pcDNA vectors by using T4 DNA ligase enzyme at 16°C overnight. After transforming ligation products into *E.coli*, colony PCR for positive colonies and sequencing was performed using vector-specific T7 and BGH primers (Table 2.1) as confirmation.

CHAPTER 3

RESULTS AND DISCUSSION

3.1 A Novel Isoform of *IGFBP4* detected by 3'-end Sequencing

Estradiol (E2) is a well-known cell proliferating hormone in estrogen receptor (ER) positive breast cancer cells. Our lab has performed a 3'-end RNA-sequencing analysis of transcripts in MCF7 cells treated by E2 for 45 minutes, 3-hour and 12-hour times points.

In this experiment, 3'-end reads of the transcripts were aligned to the GRC37/hg19 human genome. Accumulation of reads on certain regions around the genome were then examined through Integrative Genomics Viewer (IGV).

Based on RNA-seq data, I focused on *IGFBP4* to investigate the potential isoform, which has an intronicly terminated 3'end. the 3'-end sequencing data for *IGFBP4* are shown in Figure 3.1

Figure 3.1. 3'end RNA-sequencing result for *IGFBP4*.

BedGraphs were visualized on IGV (Integrated Genome Viewer). 3'-end sequencing result of IGFBP4 on IGV aligned to GRC37/hg19. Observed intronic peak locations

on GRC37/hg19 for three peaks are ch17:38602241-38602430, chr17:38602434-38602572 and chr17:38602785-38602903, respectively and UTR peak locations is chr17: :38613845-38613981.

IGFBP4 has two reported poly(A) sites in PolyA_DB. These two reported poly(A) signals, Hs.462998.1.9 and Hs.462998.1.11, are located at the 3'UTR of the gene (Figure 3.2.).

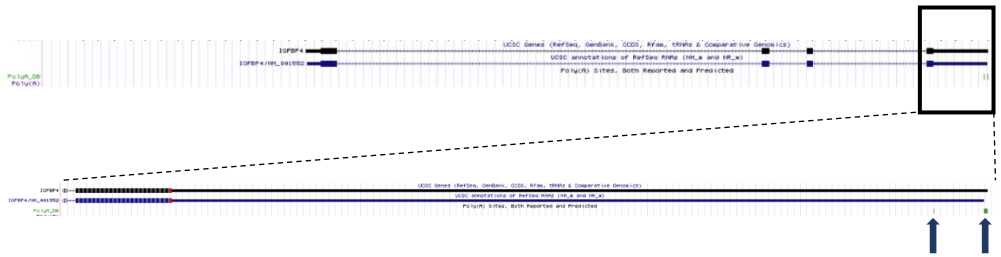


Figure 3.2. Reported poly(A) signals of *IGFBP4*.

The figure shows the reported transcript variants, exon intron boundaries, reported and predicted poly(A) signal locations of *IGFBP4* on the gene schema. Blue arrows indicate two signals, Hs.462998.1.9 and Hs.462998.1.11. No other poly(A) sites are reported for *IGFBP4*.

According to 3'seq data, the 3'UTR polyA site usage increased gradually upon E2 treatment. However, there were more robust reads coming from a promoter proximal and intronic region of *IGFBP4*. Hence, we aimed to characterize these peaks originating from the first intron of *IGFBP4*.

3.2 Experimental Confirmation of Intronic/Full Transcript Isoforms of *IGFBP4* in E2 Treated MCF7 and T47D Samples

3'-seq reads at polyA sites indicate ends of transcripts, however 3'seq reads are short and do not cover the whole length of transcripts. Hence, we wanted to experimentally test the existence of transcripts that have the indicated 3'ends. To this end, I treated

MCF7 cells with E2 for 45 minutes, 3 hours, and 12 hours (n=3). E2 treatment was confirmed by looking at the expression of a well-known estrogen-responsive gene, Trefoil Factor 1 (*TFF1*) (Bourdeau et al., 2007). Figure 3.3. shows the expected upregulation of *TFF1*. The same treatment was repeated on T47D ER+ breast cancer line (n=1) since it has the same receptor profile with MCF7, we wanted to observe the isoform expression in another cancer cell line.

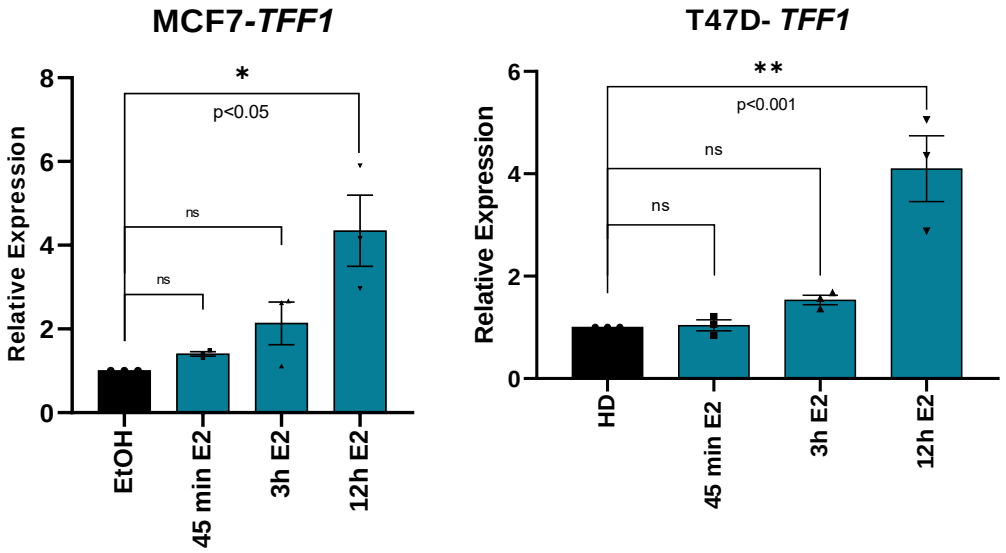


Figure 3.3. RT-qPCR result of TFF1 expression on E2/EtOH treated MCF7 and T47D cDNAs. E2 treatment in each biological replicate on MCF7 cells (n=3) and T47D cells (n=1) was verified by *TFF1* upregulation in RT-qPCR. *TFF1* relative expression was normalized for each replicate to the corresponding sample's *RPLP0* expression value. Quantification was done by using the ΔΔCq method (Livak & Schmittgen, 2001) For the statistical analysis, one-way ANOVA was applied (HD: Hormone deprived, *p<0.05, n=3).

E2 treatment success was observed for all E2 treatments.

3.2.1 *IGFBP4* expression

Specific primers were designed for the intronic and full isoforms of *IGFBP4*. Illustrations of the primer locations relative to 3'seq reads are shown in Figure 3.4.

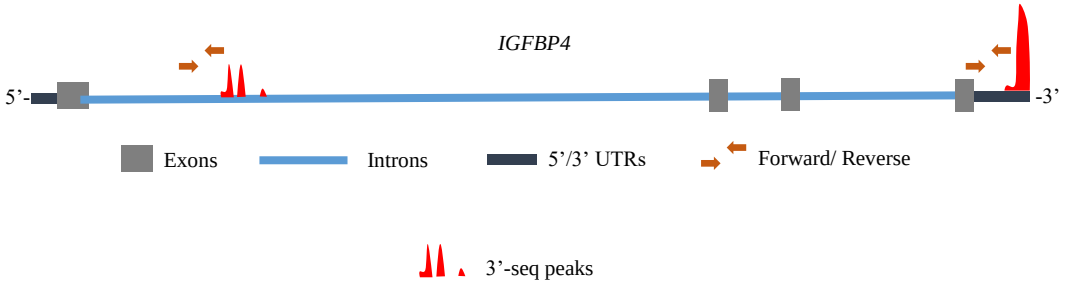


Figure 3.4. 3'-seq polyA peaks and RT-qPCR primer locations on *IGFBP4* gene structure.

The figure shows the polyA peaks detected by 3'-end sequencing and primer locations.

RT-qPCR was performed using E2-treated MCF7 cDNAs and E2 treated T47D cDNAs. Relative expressions of intronic and full isoforms of *IGFBP4* are shown in Figure 3.5. The expression patterns for the intronic/full isoforms of *IGFBP4* observed in 3'-end sequencing results was confirmed by RT-qPCR. Robust upregulation of intronic *IGFBP4* seems to decrease gradually in a time course manner, whereas the full-length isoform is upregulated in time.

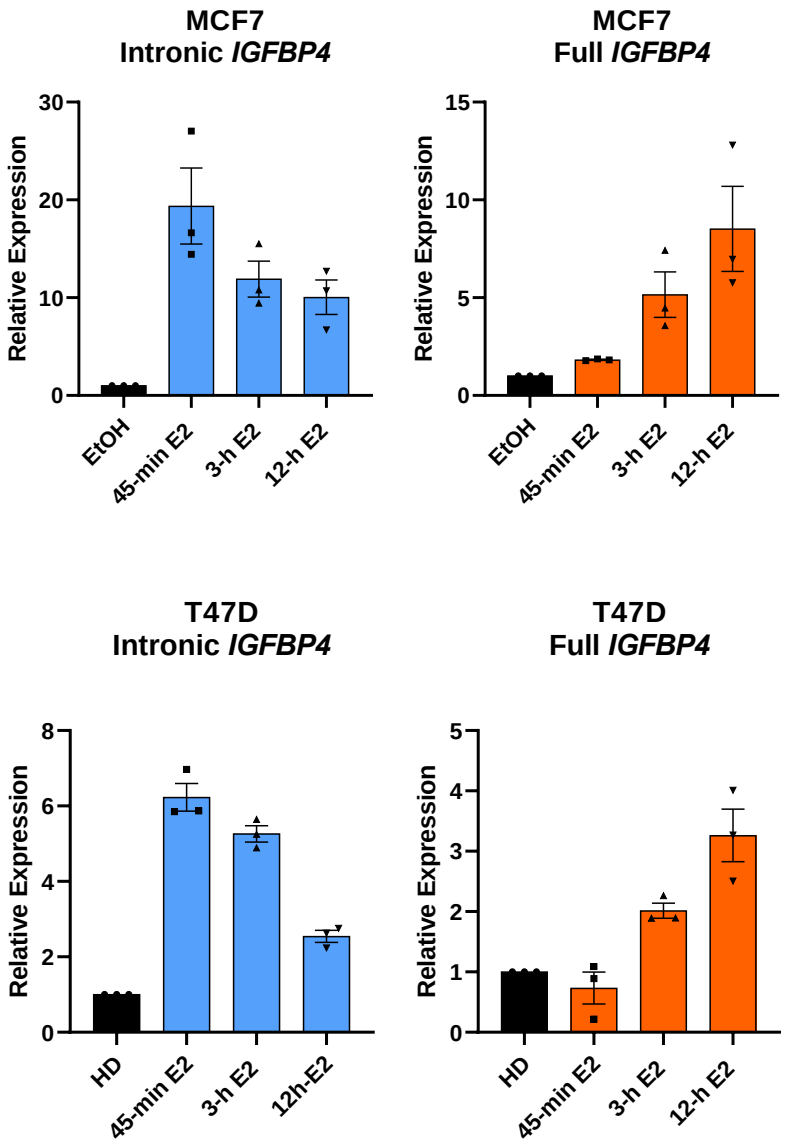


Figure 3.5. RT-qPCR results of IGFBP4 intronic and full transcripts in E2/EtOH treated MCF7 and T47D cDNAs. The expression of each sample was normalized to the corresponding sample's reference gene expression, *RPLP0*. Quantification was done by using the ΔΔCq method (Livak & Schmittgen, 2001). For the statistical analysis, one-way ANOVA was applied (**p<0.001,***p<0.0001, ns: not significant, HD: hormone deprived, MCF7 E2 treatment n=3, T47D E2 treatment

significant, HD: hormone deprived, MCF7 E2 treatment n=3, T47D E2 treatment n=1).

3.2.2 3’ Rapid Amplification of cDNA Ends (3’RACE)

I conducted 3’RACE to test whether there is indeed an isoform that has the intronic polyA site. For 3’RACE, 3’end anchored cDNAs were used as a template for the first round. To increase the final product’s specificity, two-round nested PCR was performed with two different forward primers nesting each other, one for each round. F1 and F2 forward 3’RACE primers were used, respectively, as the same anchor reverse primer was used for both rounds. For both rounds, illustration of expected sizes alongside the location of primers and the experiment results on agarose are given in Figure 3.6., 3.7., 3.8. and 3.9.

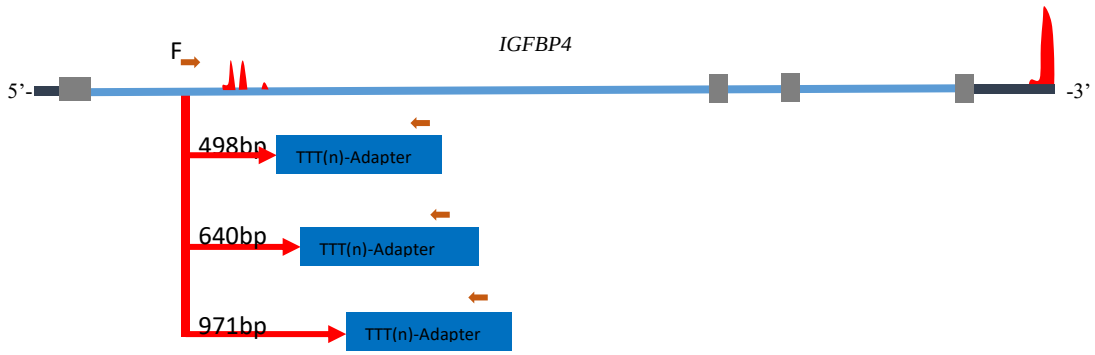


Figure 3.6. Illustration of expected product sizes of IGFBP4 1st-round 3’RACE with respect to 3’seq reads and primer locations.

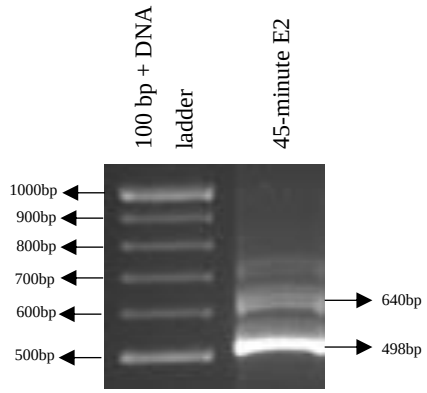


Figure 3.7. Agarose gel image of *IGFBP4* 1st-round 3'RACE.

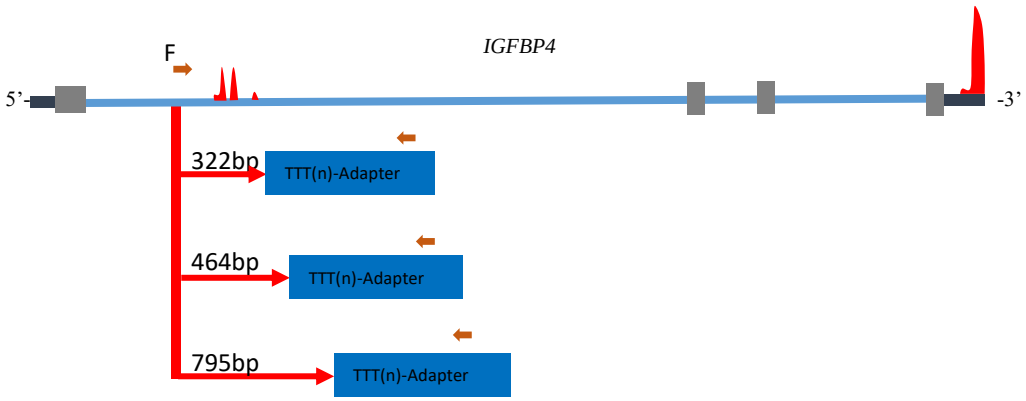


Figure 3.8. Illustration of expected product sizes of *IGFBP4* 2nd-round 3'RACE with respect to 3'seq reads and primer locations.

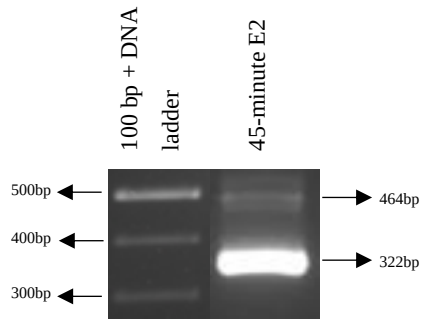


Figure 3.9. Agarose gel image of IGFBP4 2nd-round 3'RACE.

In each round, PCR products were amplified to support the locations of 3'seq peaks for *IGFBP4*. I did not see the third PCR product that would match with the third 3'seq peak, probably because of the low levels, which can be foreseen by looking at the smaller peak at this region visualized through IGV. In both rounds, non-specificity was low, increasing the possibility of obtaining a specific product at the end of the second round.

To confirm that these products are genuinely representing the 3'-end of the intronic isoform of *IGFBP4*, the 322 bp band of 2nd-round 3'RACE was extracted from the gel, purified, and cloned into pGEM-T vector as described in the Materials and Methods section. Sequencing results are shown in Figure 3.10.

Figure 3.10. Sequencing result confirms the 3'-end of an intronicly transcribed *IGFBP4* isoform.

3'RACE confirmed the 3'-end of the transcript as was detected by 3'-seq reads.

Interestingly when we analyzed the sequence at around the 3'seq reads, we detected a polyA strech on the DNA, (Figure 3.11). A polyA stretch on the DNA and on the pre-mRNA was alarming for us because 3'RACE results or the peaks at the 3'-seq data could be due to internal priming of oligo-dT primers.

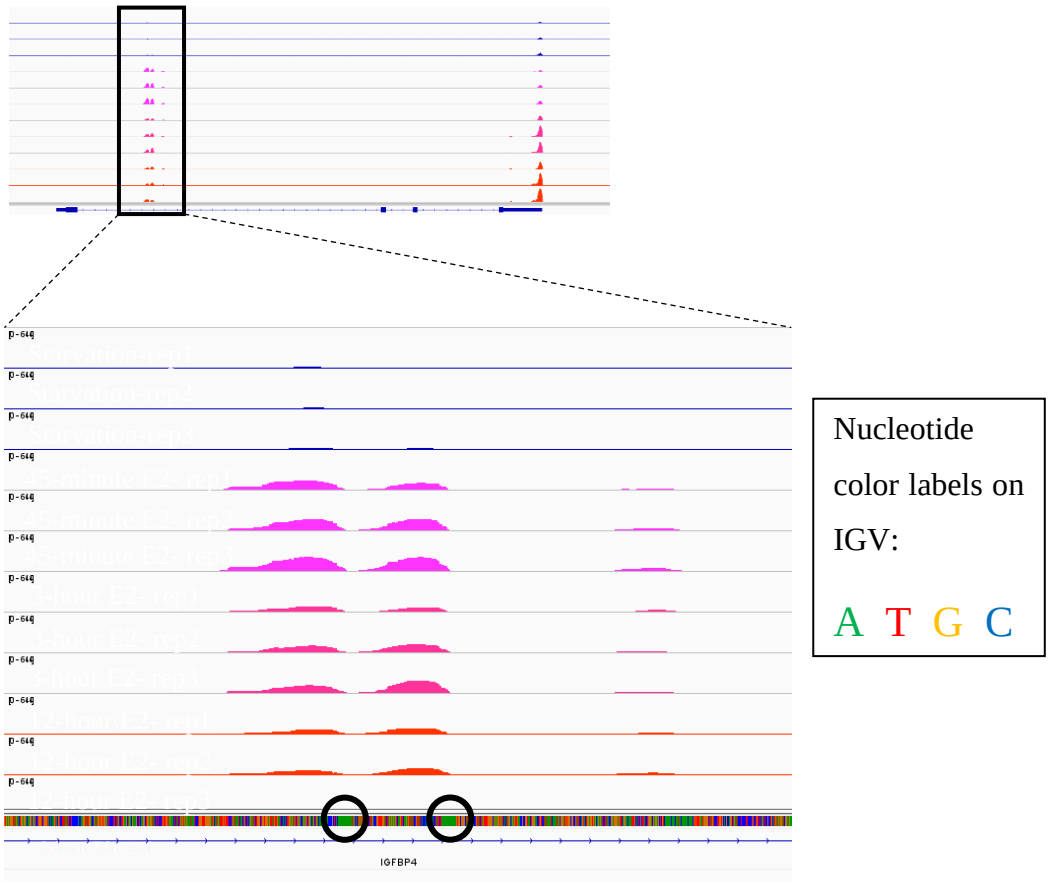


Figure 3.11. Adenine nucleotide stretches at the intronic peak downstream.

Figure shows the A-stretches downstream of intronic peaks of *IGFBP4*. Adenine nucleotides are indicated with green color on IGV. Adenines mapped to peak ends are marked with black circles.

To test these possibilities, I first looked into DNA contamination in my RNA samples. We always check for DNA contamination in our RNA samples with PCR and only proceed to synthesize cDNA when we don't detect any DNA based amplication in a PCR reaction where we use RNA as a template. DNA contamination validation results are presented in the Appendix (A).

Next, to eliminate the possibility of an internal priming, I synthesized cDNAs using random hexamers. RNAs used for this task was validated on genomic DNA contamination by using *IGFBP4* intronic primers.

cDNAs from E2 treated MCF7 cells were synthesized using random hexamers were used to perform PCR and RT-qPCR using *IGFBP4* intronic primers, (Figure 3.12 and Figure 3.13).

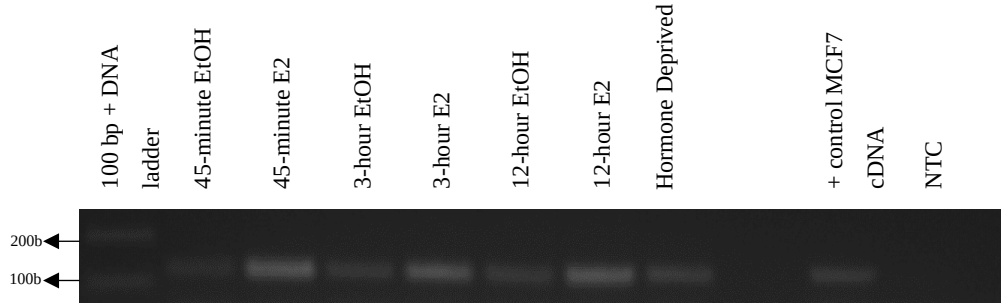


Figure 3.12.Agarose gel image of the PCR product of the *IGFBP4* intronic transcript for E2 treated samples. cDNAs were optained by using random hexamer synthesized cDNAs.

Using random hexamer synthesized cDNAs, RT-qPCR results confirmed the previous results (Figure 3.13). These results eliminated the possibility of an internal priming of the oligo-dT primers.

Random Hexamer (n=1)

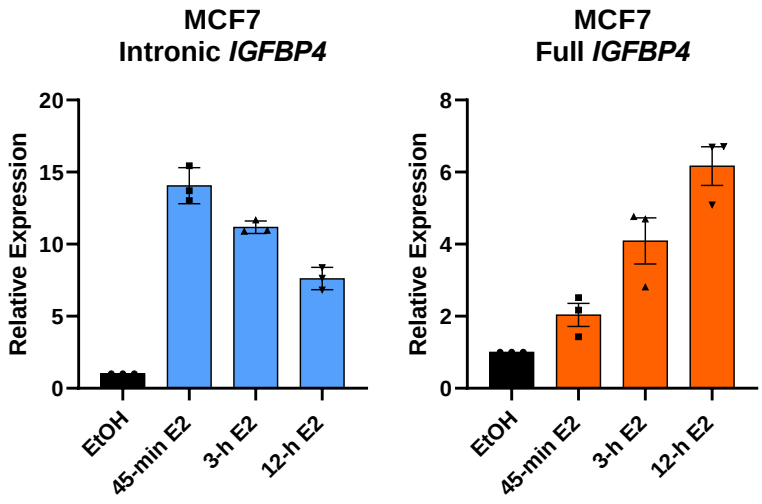


Figure 3.13. Expression of intronic and full isoforms of *IGFBP4* in E2 treated MCF7 cells. cDNAs of E2 treated samples were synthesized using random hexamers. Quantification was done by using the ΔΔCq method (Livak & Schmittgen, 2001). For the statistical analysis, one-way ANOVA was applied (***p<0.0001, n=3 technical replicates).

Next, I checked whether these regions are conserved throughout evolution to understand the source of the A stretch on DNA. Interestingly, the A stretches were part of a SINE element that appeared in the primate lineage after the bushbaby monkeys.

Conserved sequence analysis and evolutionary conservation of the SINE element for *IGFBP4* are shown in Figure 3.14.

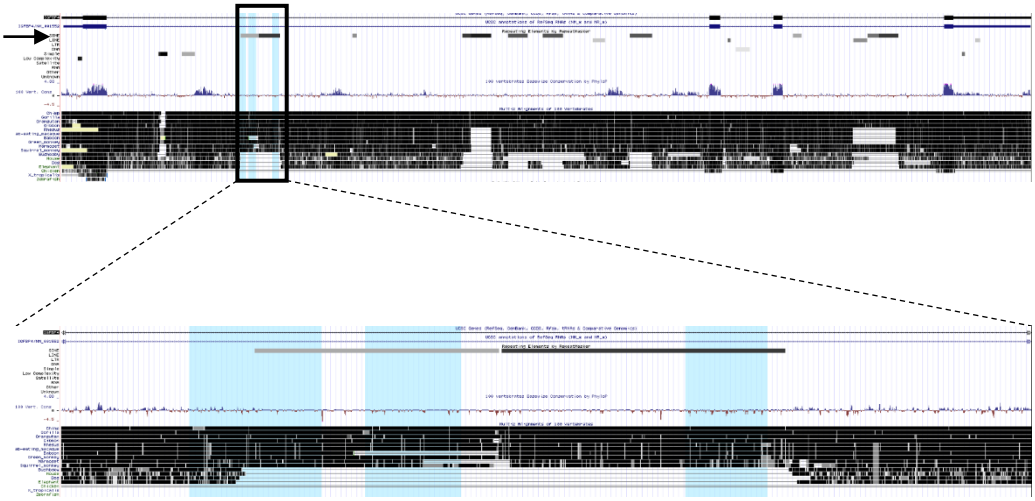


Figure 3.14. Conserved sequence analysis and evolutionary conservation of *IGFBP4*.

The SINE element is indicated with a black arrow. Three intronic peaks of *IGFBP4* are indicated with light blue highlights. SINE element is integrated into primate lineage after bushbaby monkeys.

A class of retrotransposable elements, short interspersed nuclear element (SINE), are found in mammalian genomes with high copy numbers (Lee et al., 2008; Doucet et al., 2015). Interestingly, retrotransposon-originated polyadenylation sites in host genes are implicated in transcription termination (Lee et al. 2008).

Therefore, in our case, when we analyzed the identified SINE element at the first intron of *IGFBP4*, I identified a potential poly(A) site. Hence to test the functionality of the predicted poly(A) site I found before this A stretch; I used a reporter system.

3.3 Poly(A) Signal Functionality for the Intronic Isoform of *IGFBP4*

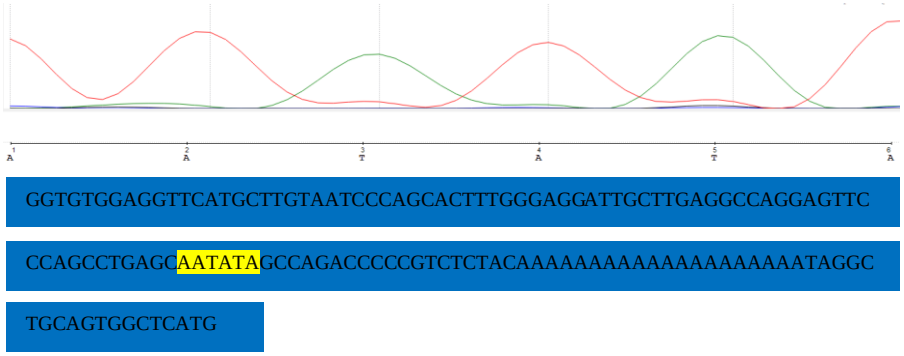
I amplified a 143 bp region upstream to the polyA stretch and the SINE element for the intronic isoform *IGFBP4*. I also amplified a 128 bp region that encompasses the 3'UTR polyA site. The intronic region contains AAUAUA as a putative signal and the 3'UTR region contains AACAAA signal at the upstream of the polyA site (Hs.462998.1.9)

Then, the intronic poly(A) signal region and canonical 3'UTR poly(A) signal region of *IGFBP4* were cloned into pMIR- Report Δpoly(A) vector which was modified in our lab, see Appendix (C). This vector lacks the SV40 polyA signal.

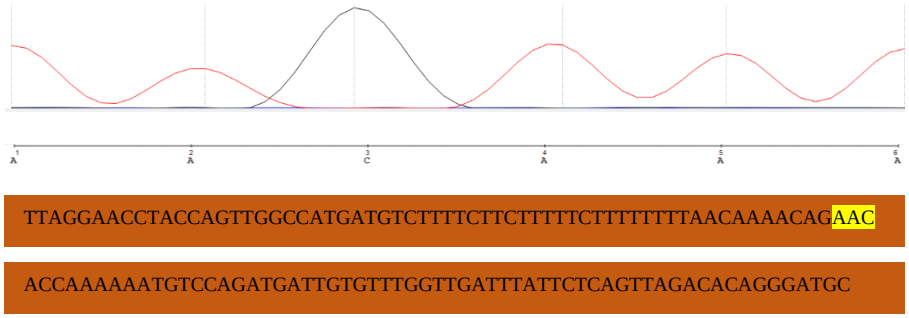
Products for this experiment were amplified using MCF7 cDNAs. Obtained products were extracted from agarose gel, purified and ligated into the pMIR-Report Δpoly(A) vector. They were then transformed into *E.coli* cells. Ligation was then confirmed by colony PCR.

I also generated an SDM version of the putative intronic pA signal to a more canonical signal to test the strength of the signal. For this aim, site-directed mutagenesis was applied to the intronic poly(A) signal containing pMIR- Report Δpoly(A) vector for changing the fifth thymine (T) nucleotide in the intronic poly(A) signal (AAUAUA) into adenine (A) nucleotide, therefore, mutated it into the most used poly(A) signal is eukaryotic mRNAs, AAUAAA (Sun et al., 2017). After performing the steps for site-directed mutagenesis as described in the Materials and Methods section, constructs were transformed into *E.coli* cells, then all plasmid constructs were isolated and then sent for sequencing. Sequencing results alongside the PAS and cleavage site which is indicated with two nucleotides, CA, are shown in Figure 3.15 (Laishram, 2014).

A) Intronic poly(A) signal and flanking elements of *IGFBP4*



B) Canonical 3'UTR poly(A) signal and flanking elements of



C) Site-dircted mutated intronic poly(A) signal and flanking elements of

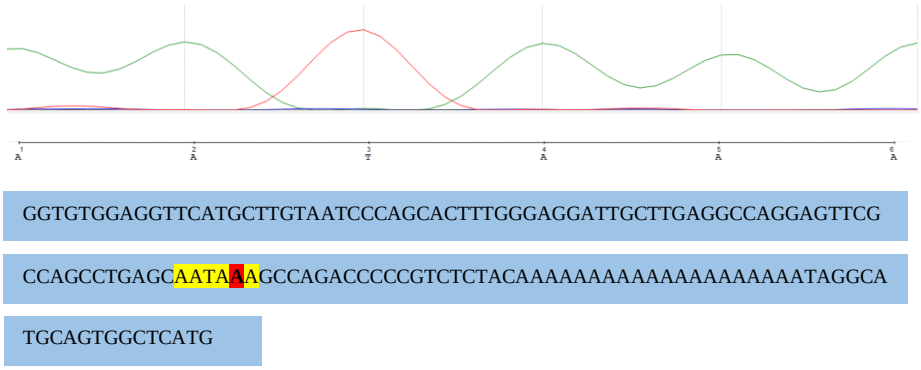


Figure 3.15. Sequencing result chromotograms of cloned intronic (A), canonical (B) and mutated intronic poly(A) signals (C) of *IGFBP4* with vector specific forward primer.

3.3.1 Dual- Luciferase Assay

MCF7 cells were co-transfected with the three vectors containing poly(A) signals and phRL-TK (Renilla). After 24-hour transfection, both renilla and firefly luciferase activities were measured for each sample using Modulus Microplate Luminometer (Turner Biosystems). After measurement, each Firefly luminescence read was normalized to the Renilla luminescence. Normalized luciferase activities of each sample are given in Figure 3.16.

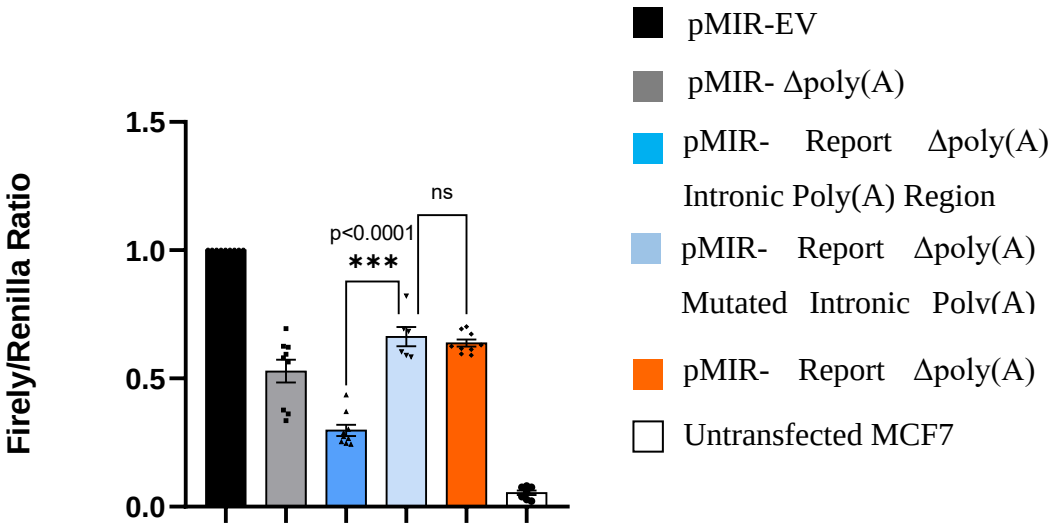


Figure 3.16. The figure shows the relative luciferase activities of different poly(A) signals. Initially, the firefly luciferase activities of samples were normalized to renilla luciferase activity. Then, in each biological replicate, Firefly/Renilla reads were normalized to pMIR-EV which has the SV40 polyA signal. For the statistical analysis, one-way ANOVA was applied (ns: not significant, ***p<0.0001, n=3).

According to the dual-Luciferase Assay results, the *IGFBP4* intronic poly(A) signal is 55% weaker than the 3'UTR poly(A) region. When the intronic signal was mutated into AAUAAA canonical poly(A) signal, the luciferase activity was increased. This experiment suggested the following conclusions.

First, the reporter system confirmed the SINE originated polyA site was functional. Second, the SDM of the signal indicated that the intronic signal is not as strong as the 3'UTR signal. And finally, the SDM vector decreased the possibility of an unintended trans factor (e.g., a microRNA) binding to the 143bp region cloned downstream of the luciferase coding sequence.

Up to this point, my data suggested that the intronic polyA peak detected by 3'seq was valid and there was indeed a regulated transcript expressed from the first intron of *IGFBP4*.

Next, I wanted to test whether this transcript has a potential to be translated. For this purpose, I examined the coding potential of the isoform.

3.4 Chromatin Architecture of *IGFBP4*

To start understanding the chromatin architecture at around the intronic polyA site, I examined several active chromatin markers on publicly available ChIP-seq data in the Cistome Database. We looked into ChIP-seq data of E2-treated MCF7 cells (Figure 3.17., 3.18). I used *TFF1* as a well-characterized control for its E2 responsiveness, see Appendix (D).

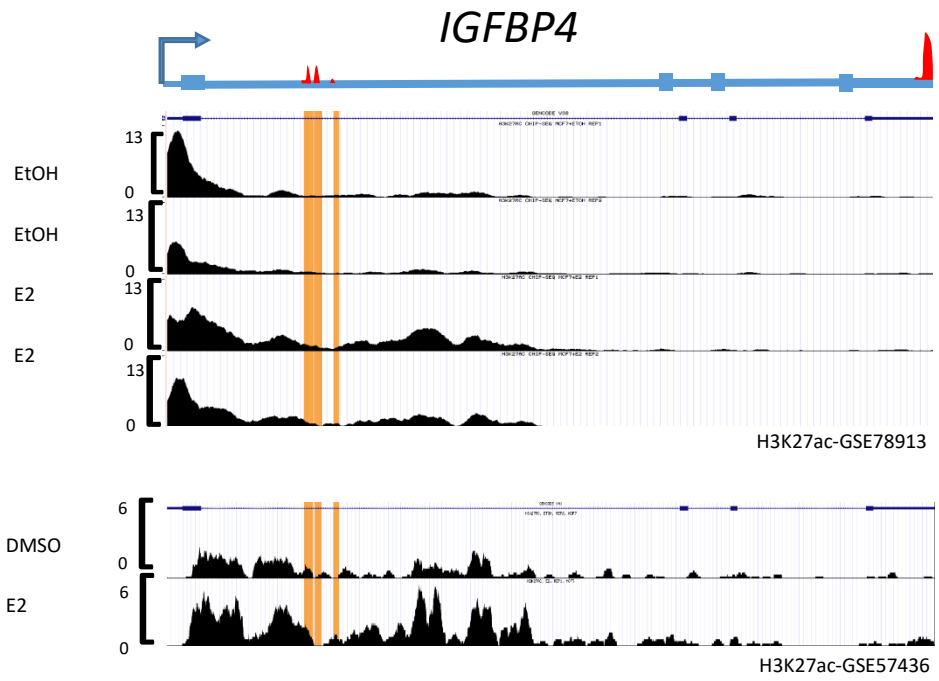


Figure 3.17. H3K27ac ChIP-seq results for IGFBP4. GSE78913 is a H3K27ac ChIP-seq dataset performed after 45 minutes of 100 nm E2 treatment to MCF7 cells. GSE57436 is a H3K27ac ChIP-seq dataset performed after 30 minutes of 100 nm E2 treatment to MCF7 cells. Time 0 and 30-minute treated samples were included to the analysis.

H3K27ac histone modification leads to neutralization of the positive charges on histone tails which consequently opens chromatin. It is mostly accumulated on enhancers and is accepted as an active chromatin marker (Zhang et al., 2020; Beacon, et al., 2020). Figure 3.17. shows that the 3'seq data generated peaks that mark the end of the intronic isoform maps to a highly acetylated H3 composition.

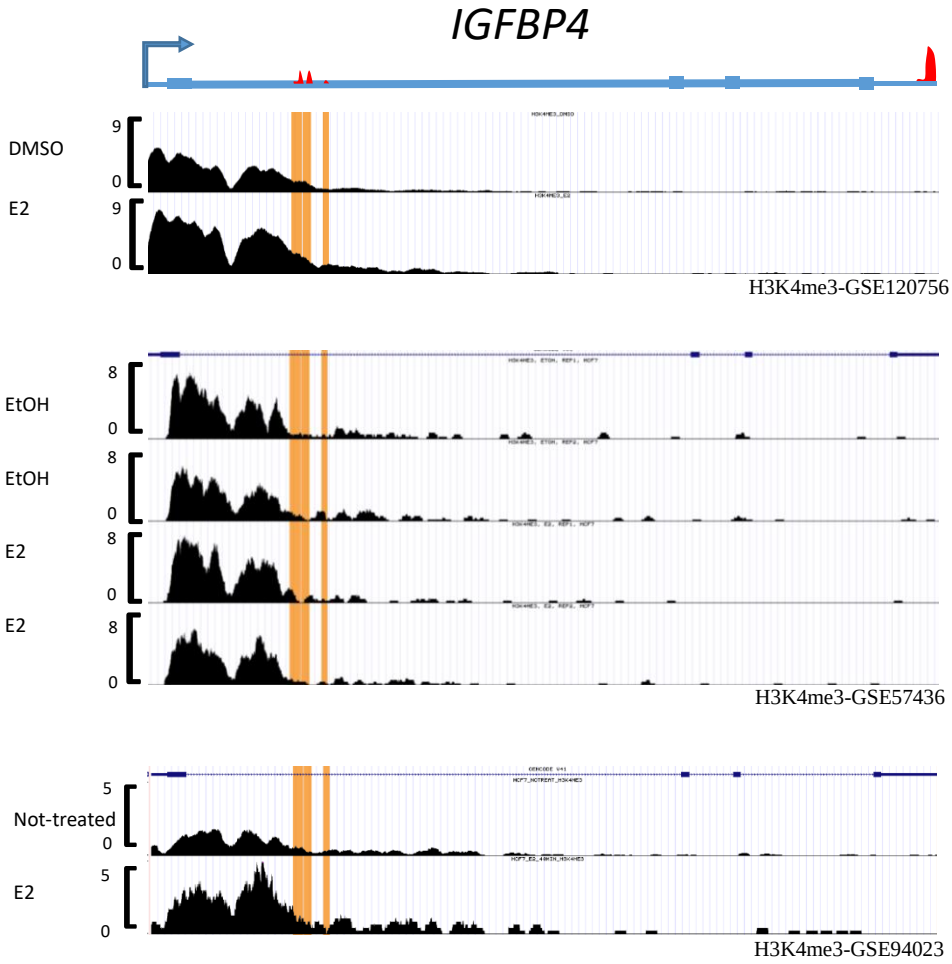


Figure 3.18. H3K4me3 ChIP-seq results for IGFBP4. GSE120756 is a H3K4me3 ChIP-seq dataset performed after 45 minutes of 10 nm E2 treatment to MCF7 cells. GSE57436 is a H3K4me3 ChIP-seq dataset performed after 30 minutes of 100 nm E2 treatment to MCF7 cells. GSE94023 is a H3K27ac ChIP-seq dataset performed after 40 minutes of 10 nm E2 treatment to MCF7 cells. Time 0 and 30-minute treated samples were included to the analysis.

H3K4me3 histone modification is accepted as transcriptionally active chromatin marker. This epigenetic marker along chromatin is mostly accumulated at the transcription start site and 5' end of the gene bodies (Beacon et al., 2021). Similar to H3K27ac, the peak ends were also a boundary for H3K4me3.

Highlighted intronic peak locations on GRC37/hg19 for three observed peaks are ch17:38602241-38602430, chr17:38602434-38602572 and chr17:38602785-38602903, respectively. Increased RNA PolII occupation on IGFBP4 after E2 treatment is also suggesting that the gene is E2 responsive.

Overall, the intronic isoform length coincides with active chromatin marks. These modification patterns suggested that either the intronic isoform is produced as a result of the chromatin architecture or chromatin has an open structure at that region because of the continuous transcription of the intronic isoform.

3.5 Coding Potential and mRNA Stability Measurement of Intronic Isoform of *IGFBP4*

After obtaining supportive evidence Coding Potential Calculator 2 (CPC2) tool was used to determine the coding potential of the intronic isoform. CPC2 uses machine learning algorithms to merge codon usage, open reading frame and amino acid composition information to predict the likelihood that a given DNA sequence will produce a protein (Kang et al., 2017). The intronic isoform sequence of *IGFBP4* was uploaded along with full isoform sequence of *IGFBP4* as a positive control since it is known to be translated. I also included *HOTAIR* long non-coding RNA sequence as a negative control. The coding potentials of the uploaded sequences are given at Figure 3.19.

ID	Label	Coding probability	Peptide length(aa)	Isoelectric point	ORF integrity
HOTAIR	noncoding	0.184882	48	11.539855957	incomplete
IGFBP4_FULL	coding	0.999808	259	6.80853271484	complete
IGFBP4_INTRONIC	coding	0.824951	151	5.89373779297	complete

Figure 3.19. CPC2 tool predictions on coding potential of IGFBP4 intronic isoform.

Interestingly, the intronic isoform exhibited a considerably high possibility to be translated by a score of 0.82. The coding potential of the full isoform was 0.99.

Next, I used in silico ORF finder tools to determine a putative coding sequence and looked into the protein primary structure. Amino acid sequence alignments of predicted intronic and known full isoforms, along with domain analysis are given in Figure 3.20

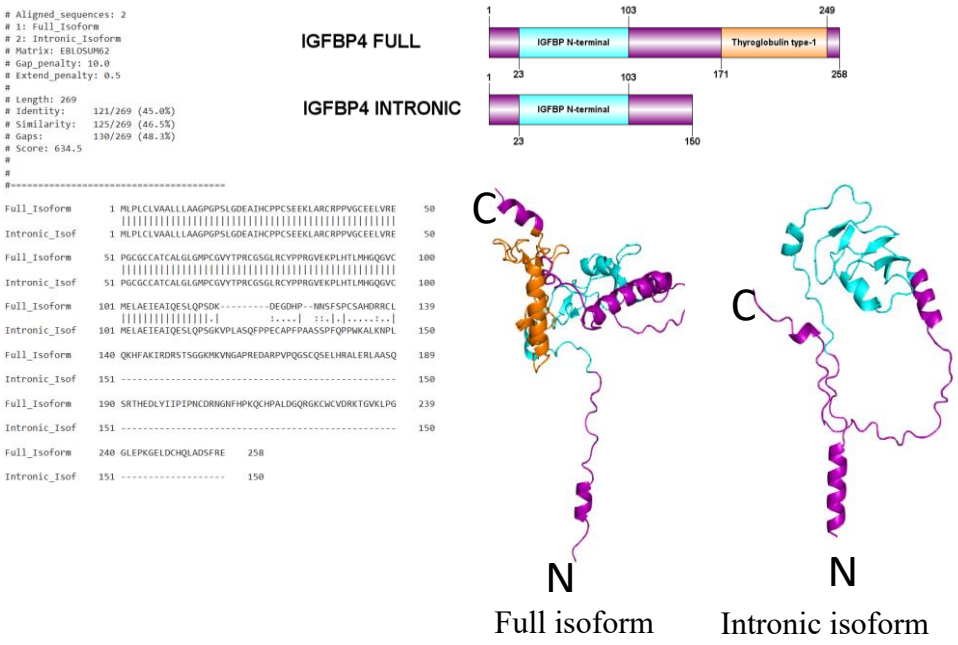


Figure 3.20. Amino acid sequences of intronic and full isoform of IGFBP4 were aligned using EMBOSS_Needle. Domain analysis and 3D structures of both isoforms were given.

The region 23-103 represents the domain IGFBP4 N-terminal, colored blue, and according to the scheme it exists in both isoforms. The region 171-249 represents the domain Thyroglobulin type-1, colored orange, according to the scheme it only exists in the full isoform. Full isoforms known protein model was retrieved from AlphaFold Protein Structure Database and illustrated by using PyMOL. Intronic isoforms predicted structure properties were retrieved from ColabFold database and illustrated by using PyMOL.

From Figure 3.18. it can be seen that the N-terminal of the truncated IGFBP4 is present in this isoform with 150 amino acids compared to the full isoform, whereas the amino acids between 150-258 are missing which correspond to C-terminal. In previous studies, via mutagenesis based approach to understand the binding of IGFBPs to IGFs, it was shown that the high affinity binding of IGFBPs mostly rely on the residues at the N-terminal of the protein (Zeslawski et al., 2001). Therefore, if translated, truncated IGFBP4 may still have the ability to bind and inhibit IGFs.

After observing a high possibility of being translated into a protein, we assessed mRNA stability for both *IGFBP4* intronic and full isoform. In order to measure the mRNA decay rate of two isoforms, actinomycin D (2 μg/mL) treatment was performed. All samples were treated with E2 for 45 minutes before Actinomycin D treatment. After E2 treatment, 12- hour Actinomycin D treatment was applied to samples with intermediate time points. RNA was collected, and cDNA was synthesized for RT-qPCR to determine the remaining transcript levels (Figure 3.21).

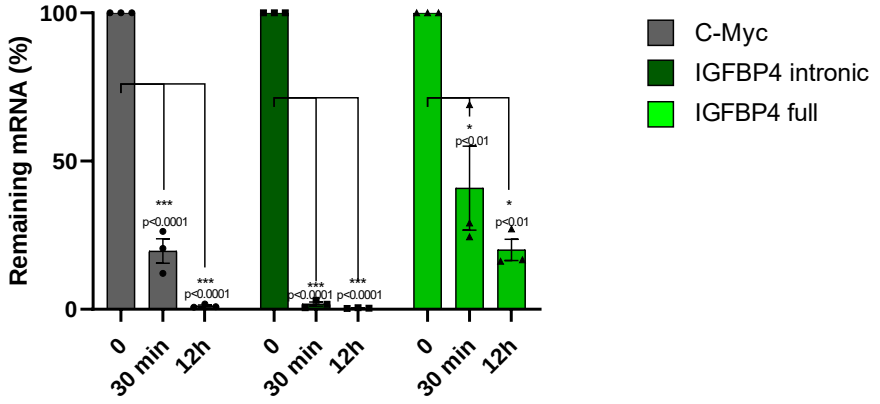


Figure 3.21. Actinomycin D treatment in MCF7 cells for 12 hours. MCF7 cells were first kept under 72-hour starvation and then treated with 100 nM E2 for 45 minutes. After the E2 treatment, 100 nM actionomycin-D treatment was applied. Cell pellets were collected at given time points, 0, 30 minutes, and 12 hours. The graph shows the remaining mRNA amount at 0, 45 minutes, and 12 hours of the genes given above after actinomycin D treatment. Normalization was done with respect to RPLP0, time 0, and the control DMSO samples of actinomycin D samples of each time point. *c-Myc* is used as the actinomycin-D treatment control. For the statistical analysis, one-way ANOVA was applied (*p<0.05, ***p<0.0001, n=3).

c-*Myc* was used as a control for validating the actinomycin D treatment. It is known that c-*Myc* has a short mRNA half-life (Sharova et al., 2009). As can be seen in the figure, c-*Myc* mRNA levels decreased rapidly where it becomes undetactable at the end of 12 hour actinomycin D treatment. When the intronic and full isoforms of *IGFBP4* were compared, the full isoform was significantly more stable than the intronic isoform of *IGFBP4*. By the end of 12-hour incubation, the remaining full isoform mRNA was 47-fold higher than the intronic mRNA.

Given the high instability of this transcript, translation of this isoform did not seem very possible or significant.

3.6 Intronic and Full Isoforms of *IGFBP4* in Different Breast Cancer Cell Lines

We were curious whether there was any correlation between the intronic and full length isoform levels. To test this, I analyzed publicly available RNA expression data for full *IGFBP4* expression in 61 different cell lines. Among examined cell lines, six cell line were selected to study isoform expressions. Three were in high *IGFBP4* expressing group, other three were in low *IGFBP4* expressing group. Comprehensive cell line analysis of *IGFBP4* expression is given in Figure 3.22.

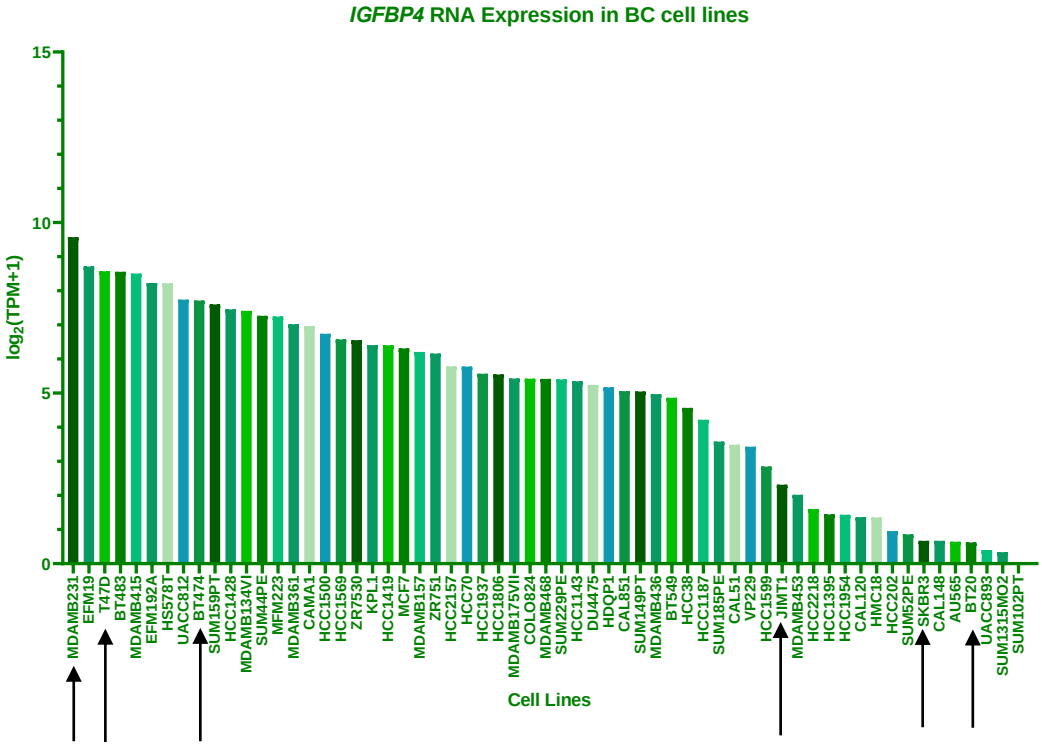


Figure 3.22.IGFBP4 expression if 61 different breast cancer cell line.

Selected cell lines for further experiments are marked with black arrows. T47D (ER +, PR +, HER2 -), MDA-MB231 (triple-negative) and BT474 (ER +, PR +, HER2 +) ,cell lines picked as high *IGFBP4* expressing cell lines. JIMT1 (ER -, PR -, HER2

-), SKBR3 (ER +/-, PR -, HER2 +), and BT20 (triple-negative) cell lines were picked as low *IGFBP4* expressing cell lines.

RT-qPCR was carried out for the selected cell lines in order to determine the relative expression of *IGFBP4* isoforms. RT-qPCR results are given in Figure 3.23.

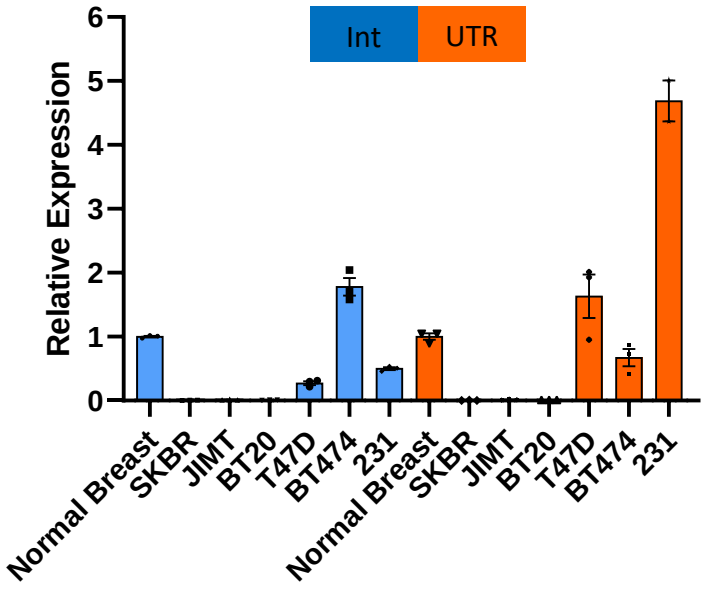


Figure 3.23. RT-qPCR IGFBP4 intron/UTR relative expression result in different cell lines. All samples were normalized to normal breast.

Figure shows the intronic and full-length isoform levels of *IGFBP4* in six selected cell lines. Consistent with the data obtained from publicly available resources, full length isoform expression SKBR3, JIMT1 and BT20 showed low *IGFBP4* full isoform expression; meanwhile, MDA-MB-231, T47D and BT474 showed high *IGFBP4* full isoform expression. Intronic isoform was present in cell lines with the full isoform expression, although the isoform ratios differ. In high *IGFBP4* expressing cell lines intronic isoform expression was observed at similar levels, meanwhile, in low *IGFBP4*-expressing cell lines, intronic isoform expression was also low or undetectable.

Because the intronic and full-length isoforms seemed to co-exist, we were curious to see whether the intronic isoform had a role in facilitating the expression of the full length by opening up the chromatin structure-based properties of chromatin which may also induce transcription dynamics and selection of cryptic polyA sites.

CHAPTER 4

CONCLUSION

Polyadenylation is a cotranscriptional process at the 3'untranslated regions (3'UTRs), but also sometimes in introns and internal exons of pre-mRNAs (Millevoi and Vagner, 2009; Turner et al., 2018). Polyadenylation signal and other elements are important elements on the mRNA sequence in terms of recognition by the polyadenylation machinery. U-rich Upstream Sequence Elements (USEs), UG-rich Downstream Sequence Elements (DSE) and hexameric poly(A) signal are the major cis-elements in the cleavage and polyadenylation process (Caggiano et al., 2022). Binding of RNA Binding Proteins (RBPs) to these motifs after recognition defines the 3'end of transcripts. 70% of the genes in humans contain more than one poly(A) signal (Linder et al., 2022). Depending on the selection of these poly(A) signals on nascent RNA can produce transcripts with varying 3'UTR length and coding sequences (Tian et al., 2005). This process is called alternative polyadenylation (APA). Selection of intronic polyadenylation sites in cancer causes genes with distorted coding sequences (Zhao et al., 2021). Especially, partial function loss after intronic polyadenylation on tumor suppressor genes is a process that may contribute to tumorigenesis (Lee at al., 2018).

To understand the E2 induced regulation of polyadenylation patterns in breast cancer, a 3'-end sequencing experiment was conducted in our laboratory. *IGFBP4* gene locus emerged as a candidate for multiple polyadenylation.

Insulin-like growth factor binding protein 4 (*IGFBP4*) is an important member of the IGF signaling system. It is a dimeric glycoprotein that binds to IGF with high affinity and blocks its binding to target receptors, IGF-IR and IGF-IIR (Chelius et al., 2001; Cheung et al., 1991; Ceda et al., 1991). This inhibitory effect of the protein on IGF

binding hamper the IGF signaling cascade and repress the outcomes of this signaling such as differentiation and proliferation. Due to this function of *IGFBP4*, it is also referred as a tumor suppressor (Lee et al., 2018).

Therefore, we further investigated the presence and characteristics of the intronic isoform of *IGFBP4*. We showed the APA pattern switch in E2 treated MCF7 cells and demonstrate the existence of the isoform via 3'RACE. From the sequencing result of 3'RACE we noticed A-stretches at the intronic poly(A) site which made us suspect the internal annealing of oligo-dT during cDNA synthesis. This possibility was tested via RT-qPCR by using E2 treated sample cDNAs synthesized by using random hexamers. In course of functional poly(A) signal motif search, we identified an unknown poly(A) signal at the intronic transcription termination site. Through conserved sequence analysis, we discovered that this newly identified potential poly(A) signal is located in a SINE. SINEs are a class of transposable elements that are known to be integrated into genome at high copy numbers to affect the biological processes such as gene expression and splicing events (Lee et al., 2008; Doucet et al., 2015). Hence, we investigated the intronic poly(A) signal functionality and observed a modest activity which was enhanced upon site-directed mutagenesis approach that changed the signal to a more canonical variant.

However, we identified this isoform to be rapidly degraded. The fact that the transcript does not reside in cells for a long time, we considered the protein product to have low significance. To this end, we also consider the SINE element and previous retrotransposition event to allow a cryptic polyadenylation signal to be activated. SINE elements and their roles in facilitating polyadenylation for host genes is an interesting suggestion (Lee et al., 2008).

Then, we investigated the chromatin architecture of the first intron of the gene. Surprisingly, open chromatin marks were enriched up until to the point of the intronic isoform length. I can suggest that either open chromatin allows activation of this SINE derived polyA site or transcription of this intronic isoform facilitates an open chromatin architecture which may modulate the full-length isoform expression.

Further experiments to eliminate the intronic polyA site or the intronic isoform may help answer these questions.

REFERENCES

Alharbi, A. B., Schmitz, U., Bailey, C. G., & Rasko, J. E. J. (2021). CTCF as a regulator of alternative splicing: new tricks for an old player. Nucleic Acids Research, 49(14), 7825–7838. https://doi.org/10.1093/nar/gkab520

Allard, J. B., & Duan, C. (2018). IGF-Binding Proteins: Why Do They Exist and Why Are There So Many? Frontiers in Endocrinology, 9.

Ağuş, H. H., & Erson Bensan, A. E. (2016). Mechanisms of mRNA polyadenylation. TURKISH JOURNAL OF BIOLOGY, 40, 529–538.

Beacon, T. H., Delcuve, G. P., López, C., Nardocci, G., Kovalchuk, I., van Wijnen, A. J., & Davie, J. R. (2021). The dynamic broad epigenetic (H3K4me3, H3K27ac) domain as a mark of essential genes. Clinical Epigenetics, 13(1). doi:10.1186/s13148-021-01126-1

Beaudoing, E. (2000). Patterns of Variant Polyadenylation Signal Usage in Human Genes. Genome Research, 10(7), 1001–1010. doi:10.1101/gr.10.7.1001

Bourdeau, V., Deschenes, J., Laperriere, D., Aid, M., White, J. H., & Mader, S. (2007). Mechanisms of primary and secondary estrogen target gene regulation in breast cancer cells. Nucleic Acids Research, 36(1), 76–93.

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. doi:10.1373/clinchem.2008.112797

Byun, D. (2001). Localization of the IGF binding domain and evaluation of the role of cysteine residues in IGF binding in IGF binding protein-4. Journal of Endocrinology, 169(1), 135–143. doi:10.1677/joe.0.1690135

Caggiano, C., Pieraccioli, M., Pitolli, C., Babini, G., Zheng, D., Tian, B., Bielli, P., & Sette, C. (2022). The androgen receptor couples promoter recruitment of RNA processing factors to regulation of alternative polyadenylation at the 3' end of transcripts. Nucleic Acids Research, 50(17), 9780–9796. https://doi.org/10.1093/nar/gkac737

Carroll, J. S., & Brown, M. (2006). Estrogen Receptor Target Gene: An Evolving Concept. Molecular Endocrinology, 20(8), 1707–1714. doi:10.1210/me.2005-0334

Ceda GP, Fielder PJ, Henzel WJ, Louie A, Donovan SM, Hoffman AR and Rosenfeld RG: Differential effects of insulin-like growth factor (IGF)-I and IGF-II on the expression of IGF binding proteins (IGFBPs) in a rat

neuroblastoma cell line: isolation and characterization of two forms of IGFBP-4. Endocrinology 128: 2815-2824, 1991.

Chelius D, Baldwin MA, Lu X and Spencer EM: Expression, purification and characterization of the structure and disulfide linkages of insulin-like growth factor binding protein-4. J Endocrinol 168: 283-296, 2001.

Cheung PT, Smith EP, Shimasaki S, Ling N and Chernausek SD: Characterization of an insulin-like growth factor binding protein (IGFBP-4) produced by the B104 rat neuronal cell line: chemical and biological properties and differential synthesis by sublines. Endocrinology 129: 1006-1015, 1991.

Deininger, P. (2011). Alu elements: know the SINEs. Genome Biology, 12(12), 236. doi:10.1186/gb-2011-12-12-236

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.

Doucet, A. J., Wilusz, J. E., Miyoshi, T., Liu, Y., & Moran, J. V. (2015). A 3′ Poly(A) Tract Is Required for LINE-1 Retrotransposition. Molecular Cell, 60(5), 728–741. doi:10.1016/j.molcel.2015.10.012

Fiorito, E., Sharma, Y., Gilfillan, S., Wang, S., Singh, S. K., Satheesh, S. V., … Hurtado, A. (2016). CTCF modulates Estrogen Receptor function through specific chromatin and nuclear matrix interactions. Nucleic Acids Research, 44(22), 10588–10602. doi:10.1093/nar/gkw785

Hjortebjerg, R. (2018). IGFBP-4 and PAPP-A in normal physiology and disease. Growth Hormone & IGF Research, 41, 7–22. doi:10.1016/j.ghir.2018.05.002

Huch, S., & Nissan, T. (2014). Interrelations between translation and general mRNA degradation in yeast. Wiley Interdisciplinary Reviews: RNA, 5(6), 747–763. doi:10.1002/wrna.1244

Jones, J. I., & Clemmons, D. R. (1995). Insulin-Like Growth Factors and Their Binding Proteins: Biological Actions*. Endocrine Reviews, 16(1), 3–34. doi:10.1210/edrv-16-1-3

Kamieniarz-Gdula, K., & Proudfoot, N. J. (2019). Transcriptional Control by Premature Termination: A Forgotten Mechanism. Trends in Genetics, 35(8), 553–564.

Kang, Y.-J., Yang, D.-C., Kong, L., Hou, M., Meng, Y.-Q., Wei, L., & Gao, G. (2017). CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. Nucleic Acids Research, 45(W1), W12–W16. doi:10.1093/nar/gkx428

Konev, A. A., Smolyanova, T. I., Kharitonov, A. V., Serebryanaya, D. V., Kozlovsky, S. V., Kara, A. N., … Postnikov, A. B. (2015). Characterization

of endogenously circulating IGFBP-4 fragments—Novel biomarkers for cardiac risk assessment. Clinical Biochemistry, 48(12), 774–780. doi:10.1016/j.clinbiochem.2015.05

Kühn, U., Gündel, M., Knoth, A., Kerwitz, Y., Rüdel, S., & Wahle, E. (2009). Poly(A) Tail Length Is Controlled by the Nuclear Poly(A)-binding Protein Regulating the Interaction between Poly(A) Polymerase and the Cleavage and Polyadenylation Specificity Factor. Journal of Biological Chemistry, 284(34), 22803–22814.

Laishram, R. S. (2014). Poly(A) polymerase (PAP) diversity in gene expression - Star-PAP vs canonical PAP. FEBS Letters, 588(14), 2185–2197. doi:10.1016/j.febslet.2014.05.029

Lawrence, J. B., Oxvig, C., Overgaard, M. T., Sottrup-Jensen, L., Gleich, G. J., Hays, L. G., … Conover, C. A. (1999). The insulin-like growth factor (IGF)-dependent IGF binding protein-4 protease secreted by human fibroblasts is pregnancy-associated plasma protein-A. Proceedings of the National Academy of Sciences, 96(6), 3149–3153. doi:10.1073/pnas.96.6.3149

Lee, J. Y., Ji, Z., & Tian, B. (2008). Phylogenetic analysis of mRNA polyadenylation sites reveals a role of transposable elements in evolution of the 3′-end of genes. Nucleic Acids Research, 36(17), 5581–5590. doi:10.1093/nar/gkn540

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. doi:10.1038/s41586-018-0465-8

Lee, Y.-Y., Mok, M. T. S., Kang, W., Yang, W., Tang, W., Wu, F., … Cheng, A. S. L. (2018). Loss of tumor suppressor IGFBP4 drives epigenetic reprogramming in hepatic carcinogenesis. Nucleic Acids Research.

Lembo, A., Di Cunto, F., & Provero, P. (2012). Shortening of 3′UTRs Correlates with Poor Prognosis in Breast and Lung Cancer. PLoS ONE, 7(2), e31129.

Linder, J., Koplik, S. E., Kundaje, A., & Seelig, G. (2022). Deciphering the impact of genetic variation on human polyadenylation using APARENT2. Genome Biology, 23(1). https://doi.org/10.1186/s13059-022-02799-4

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.

Mandel CR, Bai Y, Tong L. Protein factors in pre-mRNA 3′-end processing. Cell Mol Life Sci. 2008;65:1099–122.

Mauger, D. M., Cabral, B. J., Presnyak, V., Su, S. V., Reid, D. W., Goodman, B., … McFadyen, I. J. (2019). mRNA structure regulates protein expression through changes in functional half-life. Proceedings of the National Academy of Sciences, 201908052. doi:10.1073/pnas.1908052116

Millevoi S, Vagner S. Molecular mechanisms of eukaryotic pre-mRNA 3′ end processing regulation. Nucleic Acids Res. 2009;38:2757–74.

Mishra, R. R., Belder, N., Ansari, S. A., Kayhan, M., Bal, H., Raza, U., … Şahin, Ö. (2018). Reactivation of cAMP Pathway by PDE4D Inhibition Represents a Novel Druggable Axis for Overcoming Tamoxifen Resistance in ER-positive Breast Cancer. Clinical Cancer Research, 24(8), 1987–2001. doi:10.1158/1078-0432.ccr-17-2776

Nicholson, A. L., & Pasquinelli, A. E. (2018). Tales of Detailed Poly(A) Tails. Trends in Cell Biology.

Ning, Y., Schuller, A. G. P., Conover, C. A., & Pintar, J. E. (2008). Insulin-Like Growth Factor (IGF) Binding Protein-4 Is Both a Positive and Negative Regulator of IGF Activityin Vivo. Molecular Endocrinology, 22(5), 1213–1225. doi:10.1210/me.2007-0536

Park, H. J., Ji, P., Kim, S., Xia, Z., Rodriguez, B., Li, L., … Li, W. (2018). 3′ UTR shortening represses tumor-suppressor genes in trans by disrupting ceRNA crosstalk. Nature Genetics, 50(6), 783–789.

Rajaram, S. (1997). Insulin-Like Growth Factor-Binding Proteins in Serum and Other Biological Fluids: Regulation and Functions. Endocrine Reviews, 18(6), 801–831. https://doi.org/10.1210/er.18.6.801

Roy, B., & Jacobson, A. (2013). The intimate relationships of mRNA decay and translation. Trends in Genetics, 29(12), 691–699. doi:10.1016/j.tig.2013.09.002

Schneider, M. R., Lahm, H., Wu, M., Hoeflich, A., & Wolf, E. (2000). Transgenic mouse models for studying the functions of insulin-like growth factor-binding proteins. The FASEB Journal, 14(5), 629–640. doi:10.1096/fasebj.14.5.629

Siwanowicz, I., Popowicz, G. M., Wisniewska, M., Huber, R., Kuenkele, K.-P., Lang, K., … Holak, T. A. (2005). Structural Basis for the Regulation of Insulin-like Growth Factors by IGF Binding Proteins. Structure, 13(1), 155–167. doi:10.1016/j.str.2004.11.009

Sharova, L. V., Sharov, A. A., Nedorezov, T., Piao, Y., Shaik, N., & Ko, M. S. H. (2009). Database for mRNA Half-Life of 19 977 Genes Obtained by DNA Microarray Analysis of Pluripotent and Differentiating Mouse Embryonic Stem Cells. DNA Research, 16(1), 45–58. doi:10.1093/dnares/dsn030

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., … Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature, 479(7371), 74–79. doi:10.1038/nature10442

Sun, Y., Zhang, Y., Hamilton, K., Manley, J. L., Shi, Y., Walz, T., & Tong, L. (2017). Molecular basis for the recognition of the human AAUAAA polyadenylation signal. Proceedings of the National Academy of Sciences, 115(7), E1419–E1428. doi:10.1073/pnas.1718723115

Tian B, Hu J, Zhang H, Lutz CS. A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucleic Acids Res. 2005;33: 201–12.

Tian, B., & Manley, J. L. (2016). Alternative polyadenylation of mRNA precursors. Nature Reviews Molecular Cell Biology, 18(1), 18–30.

Turner, R. E., Pattison, A. D., & Beilharz, T. H. (2018). Alternative polyadenylation in the regulation and dysregulation of gene expression. Seminars in Cell & Developmental Biology, 75, 61–69.

Venkat, S., Tisdale, A. A., Schwarz, J. R., Alahmari, A. A., Maurer, H. C., Olive, K. P., … Feigin, M. E. (2020). Alternative polyadenylation drives oncogenic gene expression in pancreatic ductal adenocarcinoma. Genome Research, gr.257550.119.

Weiner, A. M. (2002). SINEs and LINEs: the art of biting the hand that feeds you. Current Opinion in Cell Biology, 14(3), 343–350. doi:10.1016/s0955-0674(02)00338-1

Wilton, J., Tellier, M., Nojima, T., Costa, A. M., Oliveira, M. J., & Moreira, A. (2021). Simultaneous studies of gene expression and alternative polyadenylation in primary human immune cells. mRNA 3’ End Processing and Metabolism, 349–399.

Wetterau, L. A., Moore, M. G., Lee, K.-W., Shim, M. L., & Cohen, P. (1999). Novel Aspects of the Insulin-like Growth Factor Binding Proteins. Molecular Genetics and Metabolism, 68(2), 161–181. doi:10.1006/mgme.1999.2920

Yuan, F., Hankey, W., Wagner, E. J., Li, W., & Wang, Q. (2019). Alternative polyadenylation of mRNA and its role in cancer. Genes & Diseases.

Zeslawski, W. (2001). The interaction of insulin-like growth factor-I with the N-terminal domain of IGFBP-5. The EMBO Journal, 20(14), 3638–3644. doi:10.1093/emboj/20.14.3638

Zhang, H., Rigo, F., & Martinson, H. G. (2015). Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site. Molecular Cell, 59(3), 437–448.

Zhang, T., Zhang, Z., Dong, Q., Xiong, J., & Zhu, B. (2020). Histone H3K27 acetylation is dispensable for enhancer activity in mouse embryonic stem cells. Genome Biology, 21(1). doi:10.1186/s13059-020-01957-w

Zhang, Y., Liu, L., Qiu, Q., Zhou, Q., Ding, J., Lu, Y., & Liu, P. (2021). Alternative polyadenylation: methods, mechanism, function, and role in cancer. Journal of Experimental & Clinical Cancer Research, 40(1). doi:10.1186/s13046-021-01852-7

Zhao, Z., Xu, Q., Wei, R., Wang, W., Ding, D., Yang, Y., Yao, J., Zhang, L., Hu, Y. Q., Wei, G., & Ni, T. (2021). Cancer-associated dynamics and potential regulators of intronic polyadenylation revealed by IPAFinder using standard RNA-seq data. Genome Research, 31(11), 2095–2106. https://doi.org/10.1101/gr.271627.120

Zhou, R. (2003). IGF-binding protein-4: biochemical characteristics and functional consequences. Journal of Endocrinology, 178(2), 177–193. doi:10.1677/joe.0.1780177

Zhou, R. (2004). Insulin-like growth factor-binding protein-4 inhibits growth of the thymus in transgenic mice. Journal of Molecular Endocrinology, 32(2), 349–364. doi:10.1677/jme.0.0320349

APPENDICES

A. Confirmation of lack of DNA contamination on RNA samples

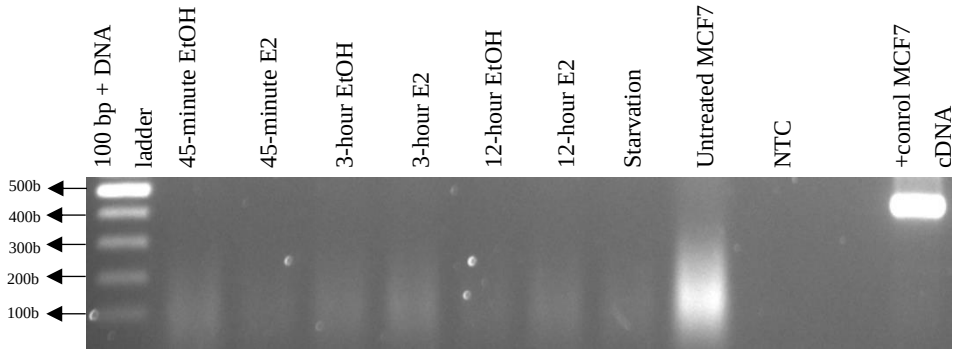


Figure 0.1. Agarose gel image of PCR products obtained by using *GAPDH* primers with RNA extracts

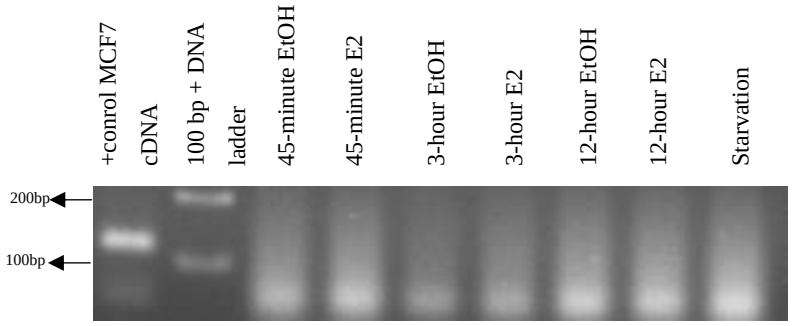


Figure 0.2. Agarose gel image of PCR products obtained by using *IGFBP4* intron primers with RNA extracts

DNA contamination in RNA samples were validated by performing 40 cycle PCR with *GAPDH* primers and *IGFBP4* intron primers. MCF7 cDNA was used as a positive control which had an expected size around 464 bp for *GAPDH* and 113bp for *IGFBP4* intron primers.

B. Confirmation of APA pattern with E2 treated MCF7 cDNAs synthesized by using Random Hexamer

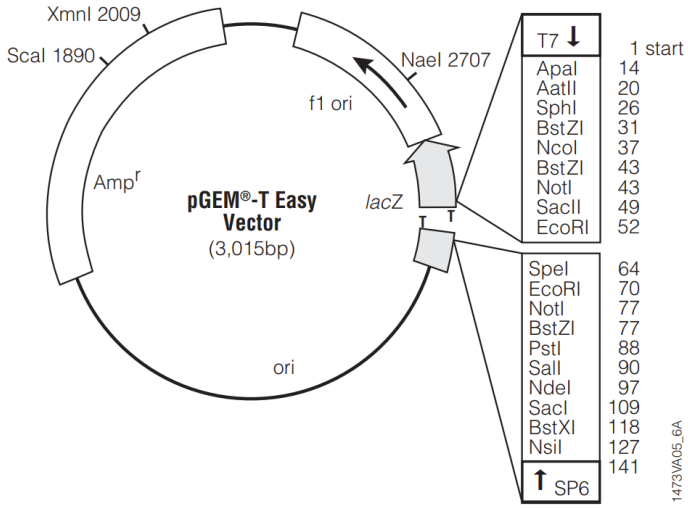


Figure 3.10. Illustration of pGEM-T Easy Vector map

C. Vector Constructs Used in the Experiments

pMIR- Report Δpoly(A) Luciferase vectors were used fort he Dual- Luciferase assay. Map of the modified vector was given in Figure 0.4.

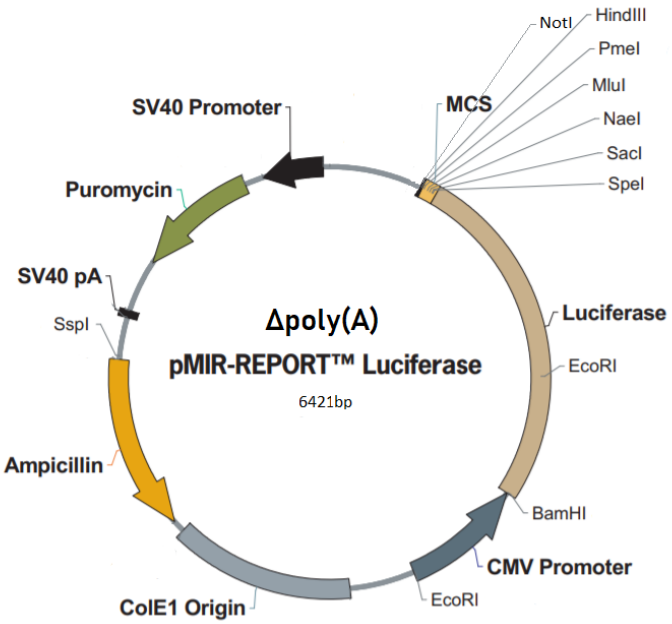


Figure 0.4. Map of pMIR- Report Δpoly(A) Luciferase modified vector

SV40 poly(A) signal from the original pMIR- Report™ Luciferase vector was deleated for poly(A) signal functionality assays.

D. Markers Used in the Experiments

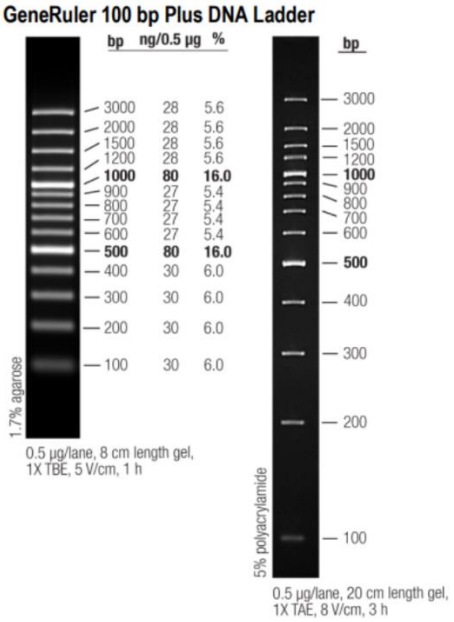


Figure 0.5. GeneRuler™ 100bp Plus DNA Ladder

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

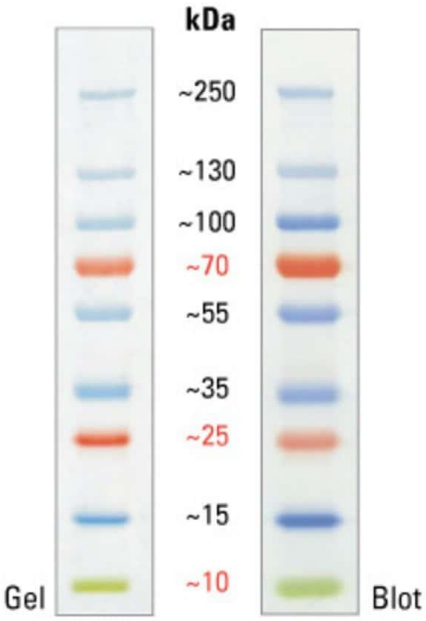


Figure 0.6. PageRuler Prestained Protein Ladder

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

E. TFF1 ChIP-seq results in studied datasets

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

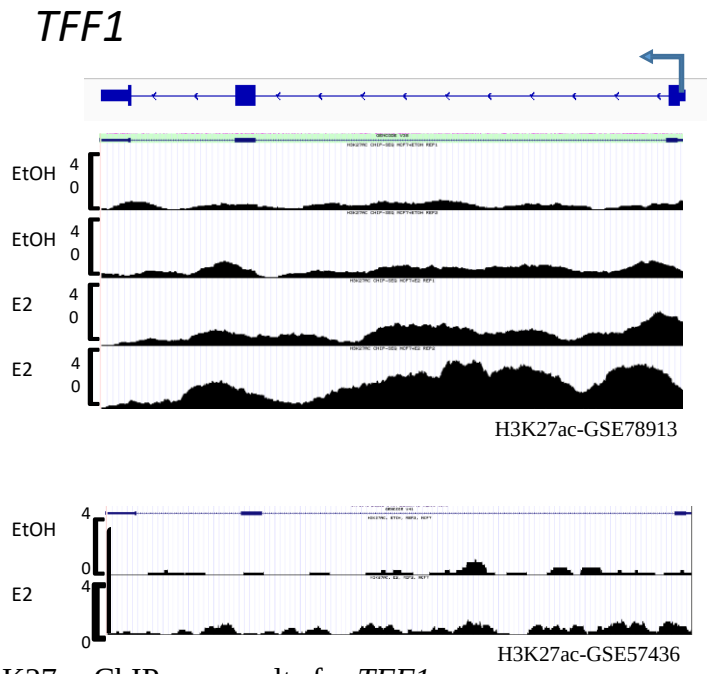


Figure 0.7. H3K27ac ChIP-seq results for *TFF1*.

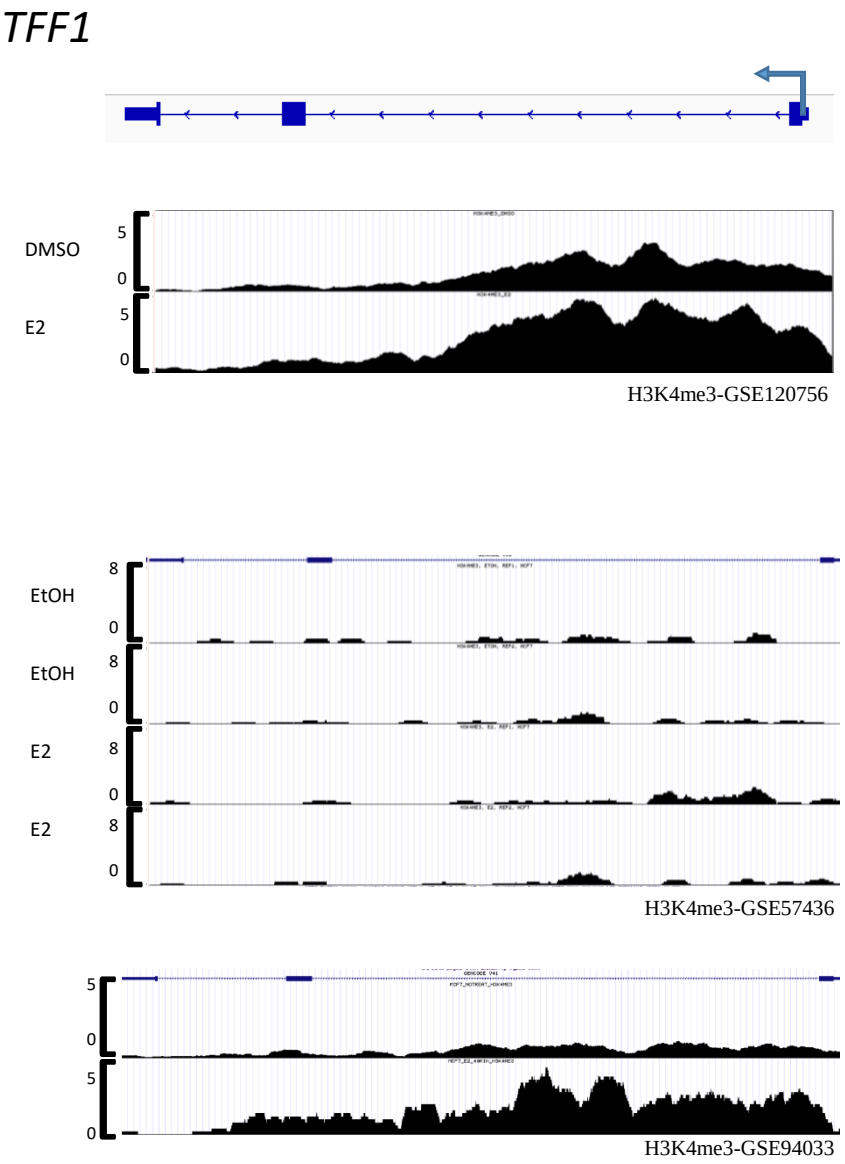


Figure 0.8. H3K4me3 ChIP-seq results for *TFF1*.