

AN EFFORT TO IDENTIFY NOVEL PROTEINS  
IN  
ALTERNATIVE POLYADENYLATION

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

BY  
MUSTAFA ÇİÇEK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR  
THE DEGREE OF DOCTOR OF PHILOSOPHY  
IN  
BIOTECHNOLOGY

JANUARY 2020



Approval of the thesis:

**AN EFFORT TO IDENTIFY NOVEL PROTEINS  
IN  
ALTERNATIVE POLYADENYLATION**

submitted by **MUSTAFA ÇİÇEK** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Biotechnology, Middle East Technical University** by,

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

Assoc. Prof. Dr. Can Özen  
Head of the Department, **Biotechnology** \_\_\_\_\_

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

Assoc. Prof. Dr. Nurcan Tunçbağ  
Co-Supervisor, **Informatics Institute, METU** \_\_\_\_\_

**Examining Committee Members:**

Prof. Dr. Yusuf Çetin Kocaefe  
Medical Biology, Hacettepe University \_\_\_\_\_

Prof. Dr. Ayşe Elif Erson Bensan  
Biology, METU \_\_\_\_\_

Prof. Dr. Sreeparna Banerjee  
Biology, METU \_\_\_\_\_

Prof. Dr. Tolga Can  
Computer Engineering, METU \_\_\_\_\_

Assist. Prof. Dr. Murat Cevher  
Mol. Biology and Genetics, Bilkent University \_\_\_\_\_

Date: 15.01.2020

**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, Last name: Mustafa Çiçek

Signature:



## **ABSTRACT**

### **AN EFFORT TO IDENTIFY NOVEL PROTEINS IN ALTERNATIVE POLYADENYLATION**

Çiçek, Mustafa

Doctor of Philosophy, Biotechnology

Supervisor: Prof. Dr. Ayşe Elif Erson Bensan

Co-Supervisor: Assoc. Prof. Dr. Nurcan Tunçbağ

January 2020, 81 pages

Polyadenylation, the addition of a poly (A) tail to a nascent messenger RNA (mRNA), is a tightly regulated process occurring co-transcriptionally. Human pre-mRNA 3' processing complex is a multi-protein machinery: Core complexes of this machinery can be listed as; i) Cleavage and Polyadenylation Specificity Factor (CPSF), ii) Cleavage Stimulatory Factor (CSTF), and iii) Cleavage Factors Im and IIm (CFIm and CFIIIm).

Majority of the human genes have multiple poly(A) signals which emphasizes the significance of alternative polyadenylation (APA), which is the regulated selection of alternate polyadenylation signals on mRNAs. APA may change the post-transcriptional fate of a eukaryotic mRNA by affecting its stability, localization and translational activity through altering availability of cis-regulatory elements on mRNA molecule. It is known that proliferative factors alter poly(A) signal selection in normal and cancer cells with a tendency towards increased 3' UTR shortening. Avoiding negative regulatory trans-factors such as microRNAs, these shortened

3'UTRs have been linked to increased protein levels. These events may explain how certain proteins are upregulated in cancer cells despite the lack of activating mutations. However, it is not yet fully understood how proximal signals are selected in proliferative cells.

At this point we hypothesize that, along with core subunits of polyadenylation machinery there may be other proteins which may play a role in selection of poly (A) signals. Aim of this thesis was to identify new proteins that may play a role in APA. For this purpose, we chose proximity-dependent biotin identification coupled with mass spectrometry analysis to identify interacting partners of CSTF2 to better understand how specific poly (A) sites are selected.

Keywords: mRNA, Poly (A) Tail, Alternative Polyadenylation, Proximity-Dependent Biotin Identification

## ÖZ

### **ALTERNATİF POLİADENİLASYONDA ROL OYNAYAN YENİ PROTEİNLERİN KEŞFEDİLMESİ ÜZERİNE BİR ÇALIŞMA**

Çiçek, Mustafa

Doktora, Biyoteknoloji

Tez Yöneticisi: Prof. Dr. Ayşe Elif Erson Bensan

Ortak Tez Yöneticisi: Doç. Dr. Nurcan Tunçbağ

Ocak 2020, 81 sayfa

Poliadenilasyon, yeni sentezlenmekte olan bir mesajcı RNA (mRNA) molekülüne poli (A) kuyruğunun eklenmesi olarak tanımlanan, oldukça sıkı bir biçimde kontrol edilen ve transkripsiyon ile eş zamanlı gerçekleşen bir süreçtir. Öncül mRNA'ların 3'-Uç bölgelerinin işlenmesini kapsayan bu süreç çok sayıda proteinin katılımıyla oluşan büyük bir protein kompleksi tarafından koordine edilmektedir.

Alternatif poliadenilasyon (APA) ise bir mRNA molekülü üzerinde yer alan alternatif poli (A) sinyallerinden birinin seçilmesi olarak tanımlanmaktadır. İnsan genomunda yer alan genlerin büyük çoğunluğunun birden fazla poli (A) sinyali taşıdığı gerçeği alternatif poliadenilasyonun (APA) fonksiyonel önemini vurgulamaktadır. APA, farklı cis-regülatör bölgelere sahip 3' UTR izoformları oluşturarak mRNA'ların stabilitesini, lokalizasyonunu ve transkripsiyonel aktivitesini etkilemekte ve sonuç olarak söz konusu mRNA moleküllerinin post-transkripsiyonel kaderini değiştirmektedir. Proliferatif sinyallerin normal hücrelerde ve kanser hücrelerinde poli (A) sinyal seçimini etkilediği ve genel olarak transkriptlerin 3'UTR uzunluklarında kısalmaya neden olduğu bilinmektedir. Daha kısa 3'UTR bölgesine sahip olan bu transkriptler, mikroRNA'lar gibi negatif-

düzenleyici etkiye sahip trans-faktörlerden kaçabilmekte ve bu durum kodlanan protein miktarının artmasına neden olabilmektedir. Gerçekleşen bu olaylar, kanser hücrelerinde, aktivasyon mutasyonlarından bağımsız bir şekilde anlatımı artan kimi proteinler için bir açıklama oluşturabilmektedir. Öte yandan, proksimal poli (A) sinyal seçiliminin ardında yatan mekanizma henüz tam anlamı ile izah edilememiştir.

Poliadenilasyon sürecinde rol oynadığı bilinen temel proteinlerin yanı sıra poli (A) sinyal seçiliminde rol oynayan, henüz tanımlanamamış yeni proteinlerin de mevcut olabileceği düşünülmektedir. Bu tezin amacı alternatif poliadenilasyonda rol oynaması muhtemel, yeni aday proteinlerin tespit edilmesidir. Bu hedefle, yakınlık-temelli biyotin tanımlamasını mümkün kılan yöntemlerin kütle spektrometresi ile birleştirilmesinden faydalanılarak CSTF2 proteini ile etkileşim halinde olan yeni proteinlerin tespit edilmesi amaçlanmıştır. Elde edilecek sonuçların spesifik poli (A) sinyallerinin seçilmesinin arkasında yatan mekanizmayı aydınlatacağı düşünülmektedir.

Anahtar Kelimeler: mRNA, Poli (A) Kuyruğu, Alternatif Poliadenilasyon,  
Yakınlık-Temelli Biotin Tanımlaması

To my family

## ACKNOWLEDGMENTS

First and foremost, I would like to express my sincerest gratitude and appreciation to my supervisor Prof. Dr. A. Elif Erson Bengan, for her endless patience, guidance, encouragement, and continuous support. Without her vision and immense knowledge, it would not be possible to find out a way to complete this thesis.

I would like to thank all my thesis committee members; Prof. Dr. Yusuf Çetin Kocafe, Prof. Dr. Sreeparna Banerjee, Prof. Dr. Tolga Can, and Assist. Prof. Dr. Murat Cevher.

I am deeply grateful to Prof. Dr. Mesut Muyan for his invaluable suggestions and inspiration. Also, I would like to express my sincere gratitude to Prof. Dr. Tolga Can for his endless support in bioinformatic analyses.

I thank all current and previous Erson Lab members for being amazing friends and sharing their knowledge. I would like to thank especially Gülten Tuncel for the priceless mass spectrometry data she provided, which created a basis for my thesis subject.

I would also like to thank all members of Muyan Lab and Banerjee Lab for their friendship and for sharing their resources whenever I needed.

I must express my very profound gratitude to my wife Esra Çiçek for always being by my side with her endless love, moral support, trust and also, for easing my burden through all ups and downs of this journey.

Last but not least I would like to thank my family, extended with Feray Yavuz and Sıdkı Yavuz, to whom this thesis is dedicated to, for being a constant source of love, support and strength throughout my thesis study. I am grateful to them just for being there for me.

This work is partially funded by Scientific and Technological Research Council of Turkey under grant number TUBİTAK 118Z745.

## TABLE OF CONTENTS

ABSTRACT .....	v
ÖZ.....	vii
ACKNOWLEDGMENTS.....	x
TABLE OF CONTENTS .....	xii
LIST OF TABLES .....	xv
LIST OF FIGURES.....	xvi
LIST OF ABBREVIATIONS .....	xviii
CHAPTERS	
1 INTRODUCTION.....	1
1.1 Alternative Polyadenylation .....	1
1.1.1 3'-End Processing of pre-mRNA in Eukaryotes .....	1
1.1.2 Mechanism of Cleavage and Polyadenylation .....	2
1.1.3 Types of Alternative Polyadenylation.....	5
1.1.4 Molecular Consequences of APA .....	7
1.1.5 Possible Induction Mechanisms of APA.....	10
1.2 Aim of the Study.....	13
2 MATERIALS AND METHODS .....	15
2.1 Culture Conditions for MCF7 Cells.....	15
2.2 Preparation of CD-FBS for Hormone Treatment Experiments .....	15
2.3 Expression Analysis.....	16
2.3.1 RNA Isolation.....	16



2.3.2	cDNA Synthesis from RNA Samples .....	17
2.3.3	Expression Analysis for <i>TFF1</i> Gene.....	18
2.4	Immunocytochemistry .....	19
2.5	Cloning <i>CSTF2</i> Gene into pcDNA3.1 mycBioID Vector.....	20
2.6	Cloning <i>CSTF2</i> Gene into pcDNA3.1 MCS-BirA(R118G)-HA Vector..	22
2.7	Cloning <i>CSTF2</i> Gene into pcDNA3.1 (-) Vector.....	23
2.8	Cloning TurboID into <i>CSTF2</i> _pcDNA3.1 (-) Vector.....	24
2.9	Transformation of Competent <i>E. coli</i> TOP10 Cells.....	25
2.10	Transient Transfection of MCF7 Cells.....	26
2.11	Proximity-Dependent Biotinylation and Streptavidin-Biotin Capture .	26
2.12	Western Blot Analysis .....	28
2.12.1	Nuclear Lysate Isolation .....	28
2.12.2	Western Blotting .....	29
3	RESULTS AND DISCUSSIONS.....	31
3.1	An Investigation of APA in ER+ Breast Cancer Cells.....	31
3.2	Investigation on Interacting Partners of <i>CSTF2</i> via Proximity-Dependent Biotin Identification .....	31
3.2.1	Expression of BirA*-Fused <i>CSTF2</i> Protein in HEK293 Cells.....	31
3.2.2	Expression of TurboID-Fused <i>CSTF2</i> Protein in HEK293 and MCF7 Cells	42
3.2.3	Identification of Potential Interacting Partners of <i>CSTF2</i> with Mass Spectrometry .....	52
4	CONCLUSION AND FUTURE DIRECTIONS.....	59
	REFERENCES .....	61
A.	Plasmid Maps .....	73

B. Markers .....	76
C. Buffers .....	78
CURRICULUM VITAE .....	81

## LIST OF TABLES

### TABLES

Table 2.1 DNase I Treatment Reaction Conditions .....	16
Table 2.2 cDNA Synthesis Reaction Conditions .....	18
Table 2.3 Primers for Cloning <i>CSTF2</i> Gene into pcDNA3.1 mycBioID Vector ...	20
Table 2.4 PCR Mixture Preparation for Amplification of Desired Sequences .....	21
Table 2.5 Primers for Cloning <i>CSTF2</i> gene into pcDNA3.1 MCS-BirA(R118G)- HA Vector .....	23
Table 2.6 Primers for Cloning <i>CSTF2</i> Gene into pcDNA3.1 (-) Vector .....	24
Table 2.7 Primers for Cloning TurboID into <i>CSTF2</i> _pcDNA3.1(-) Vector .....	25
Table 2.8 Compositions of Lysis Buffers .....	29
Table 2.9 List of the antibodies used in western blot analyses throughout this study .....	30

## LIST OF FIGURES

### FIGURES

Figure 1.1. A schematic representation of the polyadenylation complex .....	5
Figure 1.2. Types of alternative polyadenylation events.....	6
Figure 1.3. Summary of potential consequences of APA .....	10
Figure 1.4. Potential APA regulators .....	13
Figure 3.1. Cloning <i>CSTF2</i> coding sequence into pcDNA3.1 MCS-BirA(R118G)- HA vector .....	32
Figure 3.2. Cloning <i>CSTF2</i> coding sequence into pcDNA3.1 mycBioID vector ...	32
Figure 3.3. Transfection of HEK293 cells with <i>CSTF2</i> and BirA* fusion constructs .....	33
Figure 3.4. Transfection of HEK293 cells with <i>CSTF2</i> and BirA* fusion constructs .....	34
Figure 3.5. Transfection of HEK293 cells with <i>CSTF2</i> and BirA* fusion constructs .....	35
Figure 3.6. Biotinylation of potentially interacting partners of <i>CSTF2</i> .....	36
Figure 3.7. Streptavidin affinity capture of biotinylated proteins .....	38
Figure 3.8. Streptavidin affinity capture of biotinylated proteins .....	39
Figure 3.9. Streptavidin affinity capture of biotinylated proteins .....	40
Figure 3.10. Streptavidin affinity capture of biotinylated proteins .....	41
Figure 3.11. Cloning <i>CSTF2</i> coding sequence and <i>TurboID</i> sequence into pcDNA3.1 (-) vector.....	42
Figure 3.12. Transfection of HEK293 cells with <i>CSTF2</i> -TurboID fusion construct .....	43
Figure 3.13. Time-dependent biotinylation of potentially interacting partners of <i>CSTF2</i> .....	44
Figure 3.14. Streptavidin affinity capture of biotinylated proteins .....	45

Figure 3.15. Effects of buffer composition on streptavidin affinity capture.....	46
Figure 3.16. Cellular localization of fusion CSTF2-TurboID proteins.....	47
Figure 3.17. Streptavidin affinity capture of potentially interacting partners of CSTF2 .....	49
Figure 3.18. Streptavidin affinity capture of biotinylated CSTF2 proteins .....	50
Figure 3.19. Streptavidin affinity capture of biotinylated CSTF77 proteins .....	51
Figure 3.20. Separation of nuclear and cytoplasmic lysates .....	52
Figure 3.21. Characterization of Mass Spectrometry analysis samples.....	54
Figure 3.22. Characterization of Mass Spectrometry analysis samples.....	55
Figure 3.23. Proteins identified by nLC-MS/MS.....	57

## LIST OF ABBREVIATIONS

### ABBREVIATIONS

APA	Alternative polyadenylation
bp	Base pairs
BSA	Bovine serum albumin
cDNA	Complementary Deoxyribonucleic Acid
CPSF	Cleavage and polyadenylation-specific factor
CSTF	Cleavage stimulation factor
DNA	Deoxyribonucleic Acid
DNase I	Deoxyribonuclease I
dNTP	Deoxyribonucleotide triphosphate
E2	Estrogen
EtOH	Ethanol
FBS	Fetal bovine serum
g	Centrifuge gravity force
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
HRP	Horseradish peroxidase
MIQE	Minimum information for publication of quantitative real-time PCR experiments
miRNA	microRNA
mRNA	messenger RNA
NCBI	National Center for Biotechnology Information
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
Pre-mRNA	Precursor messenger RNA
RBP	RNA binding protein
RNA	Ribonucleic acid
SDS	Sodium dodecyl sulfate
TBS-T	Tris buffered saline- Tween
<i>TFF1</i>	Trefoil factor 1
UTR	Untranslated Region

# CHAPTER 1

## INTRODUCTION

### 1.1 Alternative Polyadenylation

#### 1.1.1 3'-End Processing of pre-mRNA in Eukaryotes

Mature eukaryotic messenger RNAs (mRNA) are produced from pre-mRNAs, the primary products of RNA polymerase II. The generation of mRNAs from these precursors requires a number of processing events, all occurring in the nucleus. 3'-end processing is one of these essential processes. With the exception of replication-dependent histone mRNAs, all eukaryotic pre-mRNAs undergo an endonucleolytic cleavage and lose sequences from 3'-ends and subsequently receive a poly (A) tail (Krieg and Melton, 1984; Wahle and Rügsegger, 1999).

The poly (A) tail affects the fate of the mRNA molecule in various aspects: 1) It can affect translation rate of an mRNA because a protein-mediated interaction between 5'-cap and the poly (A) tail is essential for initiation of translation (Sachs *et al.*, 1997), 2) Since the rate-limiting stage of the mRNA decay process is deadenylation, the poly (A) tail also has effects on mRNA stability (Beelman and Parker, 1995). In addition, the poly(A) tails aid the transportation of the mRNAs from the nucleus to the cytoplasm (Huang and Carmichael, 1996).

## **1.1.2 Mechanism of Cleavage and Polyadenylation**

### **1.1.2.1 Sequence Elements of Cleavage and Polyadenylation**

The cleavage and polyadenylation reactions require the presence of three sequence elements in pre-mRNA. The first sequence element is the poly (A) signal, which is located nearly 10 – 30 nucleotides upstream of the actual cleavage sites. The canonical hexanucleotide AAUAAA is highly conserved in mammalian species. Most of the other mRNA molecules that do not have a perfect match to this canonical signal differ by only an A-to-U nucleotide conversion at the second position. There are also other signal variants (e.g. AUAAAA, CAAUAA, AUAAAC) that are utilized throughout the genome with lower frequencies (Colgan and Manley, 1997; Derti *et al.*, 2012).

The second sequence element is located generally 20 – 40 nucleotides downstream of the cleavage site but also can be located further downstream (Takagaki and Manley, 1997). In contrast to poly (A) signal, there is no consensus for this downstream-element and the sequence is generally referred as “uridine-rich” (U-rich) and/or “guanine/uridine-rich” (GU-rich) since it is mostly characterized by stretches of five sequential U-residues, generally interrupted by single G nucleotides (MacDonald *et al.* 1994; Takagaki and Manley, 1997). The sequence composition of these two elements and the distance between them determines the strength of poly (A) signal and also specifies the site of cleavage and poly (A) addition. Changing the distance between these two core elements causes a shift in the precise site of the endonucleolytic attack (MacDonald *et al.* 1994; Chen *et al.*, 1995; Colgan and Manley, 1997).

Another sequence element is located 10–35 nucleotides upstream of the poly (A) signal and its function is help to position the polyadenylation machinery. Recognition of the poly (A) signal primarily depends on the presence of this



upstream U-rich (e.g. UGUA) element (Brown and Gilmartin, 2003; Millevoi and Vagner, 2009).

### **1.1.2.2 The Polyadenylation Machinery**

The recognition of the poly (A) signal, cleavage and polyadenylation of the nascent mRNA is coordinated by a multi-protein machinery. This machinery (Figure 1.1) is mainly composed of three protein complexes including i) Cleavage and Polyadenylation Stimulatory Factor (CPSF), ii) Cleavage Stimulatory Factor (CSTF), and iii) Cleavage Factors Im and IIm (CFIm and CFIIIm). In addition to these complexes, symplekin functions as a scaffold protein. The untemplated adenosines are added by poly(A) polymerases (PAPs) but the stimulation of PAPs and regulation of the length of the poly (A) tail is performed by poly (A)-binding proteins (PABPs) (Kühn *et al.*, 2009).

The sequence specificity of the process is provided by the recognition of polyadenylation signal AAUAAA by the CPSF complex. This complex is composed of CPSF30, CPPSF73, CPSF100, CPSF160, hFip1 and WRD33 subunits (Kaufmann *et al.*, 2004; Mandel *et al.*, 2007). WRD33 is the binding subunit to the AAUAAA signal (Chan *et al.*, 2014; Schönemann *et al.*, 2014) and CPSF73 is the endonuclease which cleaves the pre-mRNA at poly (A) site (Mandel *et al.*, 2006).

CSTF complex consists of CSTF1, CSFT2 and CSTF3 subunits (Millevoi and Vagner, 2009). This complex binds to the downstream region of the cleavage site and is necessary for the accuracy of cleavage reaction (Wilusz and Shenk, 1990). The CSTF2 subunit of the complex interacts with downstream GU-rich regions directly (Takagaki and Manley, 2000; Yao *et al.*, 2002) whereas CSTF1 and CSTF3 subunits interact with C-terminus of RNA pol II (Cevher and Kleiman, 2010).

CFIm complex, which consists of CFIm25, CFIm59, and CFIm68 subunits, binds to the upstream elements (Brown and Gilmartin, 2003). Binding of CFIm to upstream UGUA elements is thought to role in poly (A) site recognition (Yang and Doubleie,

2011). On the other hand, the CFIm complex, consisting of CLP1 and PCF11 subunits, may have a role in RNA Pol II-mediated termination of the transcription (Birse *et al.*, 1998).

Cleavage reaction is immediately followed with polyadenylation of the nascent mRNA. This situation suggests that these two processes are closely related and binding of the CPSF complex to the polyadenylation signal recruits PAPs (Kaufmann *et al.*, 2004). PAPs have multiple alternatively spliced isoforms (Colgan and Manley, 1997) and their activity is linked to poor prognosis in breast cancer and leukemia (Scorilas, 2002).

Besides the mentioned complexes of polyadenylation machinery, RNA-binding protein (RBPs) can also affect the cleavage and polyadenylation of target mRNAs. These proteins can enhance the binding of polyadenylation machinery to their targets or can compete with this machinery (Erson-Bensan and Can, 2016).

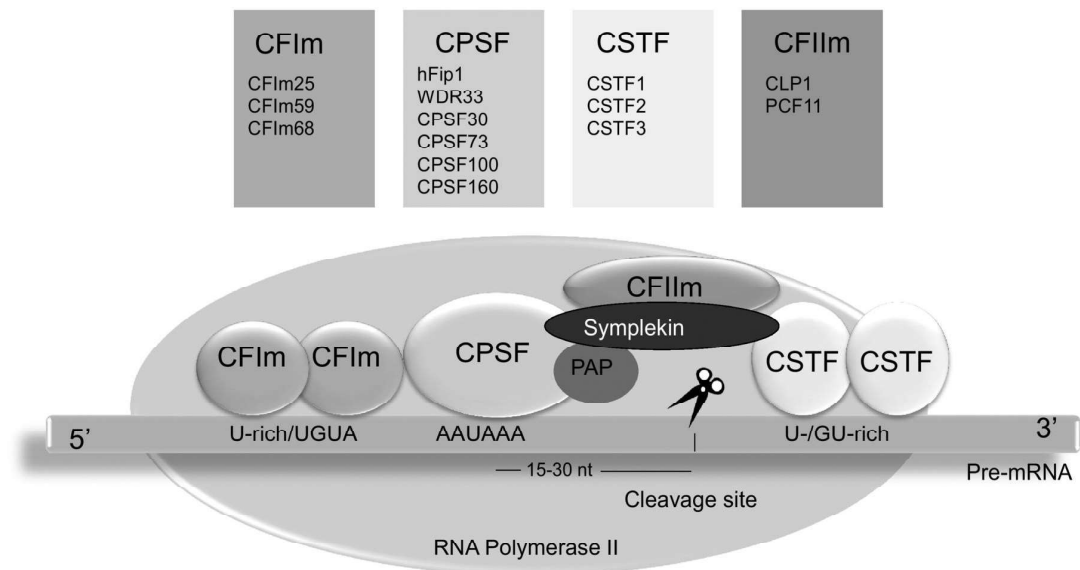


Figure 1.1. A schematic representation of the polyadenylation complex. Boxed regions are polyadenylation signal sequences (upstream U-rich/UGUA, AAUAAA poly (A) signal, and downstream U-/GU-rich region). Figure retrieved from (Erson-Bensan and Can, 2016).

### 1.1.3 Types of Alternative Polyadenylation

Most of the human genes harbor multiple poly (A) signals which emphasize the significance of alternative polyadenylation (APA), the selection of alternate polyadenylation signals. APA has been reported to alter the post-transcriptional fate of a eukaryotic mRNA by affecting its degradation rate, localization, nuclear export and translational efficiency (Moore, 2005). According to the genomic location of the selected poly (A) signal APA can be classified into four groups (Figure 1.2). The first group, named “Tandem 3’UTR APA” can be defined as the simplest and most common form of APA events (Figure 1.2 A). In this type of APA, the mRNA harbors multiple poly (A) signals in its terminal exon and the selection of alternate signals do not alter the coding potential, instead, it generates isoforms that differ in only 3’UTR length. The second most common class is named “Alternative Terminal Exon APA” (Figure 1.2 B). In this group, alternative splicing results in skipping of an internal exon and the use of a distinct poly (A) signal located within the new terminal

exon. “Intronic APA” constitutes the third group of APA events. During intronic APA (Figure 1.2 C), mRNA is cleaved upon selection of a cryptic intronic poly (A) signal which results in the extension of an internal exon and making it the new terminal one.

The final group of APA events, named “Internal Exon APA” (Figure 1.2 D), involves premature polyadenylation within the coding region which can lead to the production of truncated proteins (Elkon *et al.*, 2013; Akman & Erson-Bensan, 2014; Yuan *et al.*, 2019).

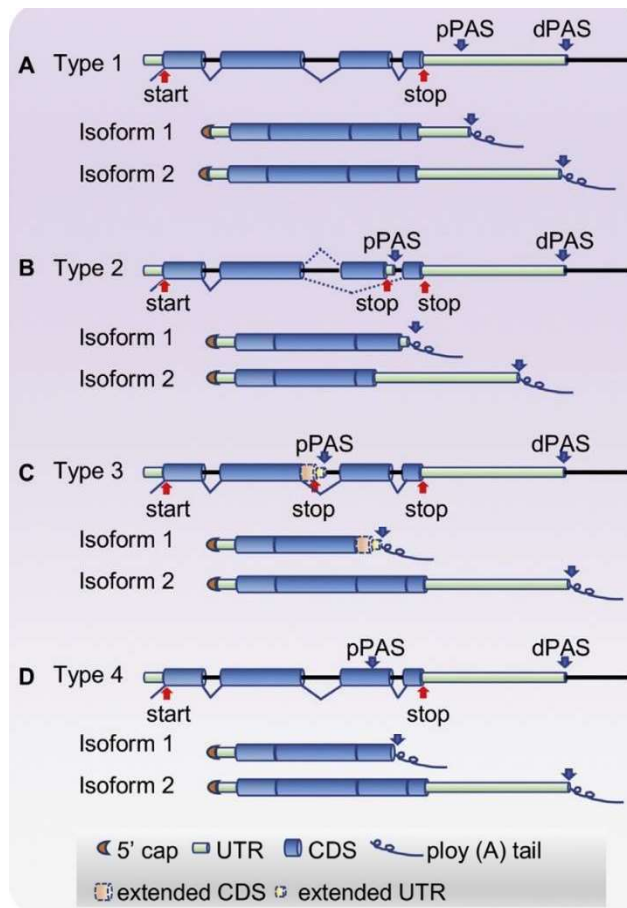


Figure 1.2. Types of alternative polyadenylation events. Type 1: Tandem 3' UTR APA, Type 2: Alternative terminal exon APA, Type 3: Intronic APA, Type 4: Internal exon APA. Figure retrieved from (Yuan *et al.*, 2019).

#### 1.1.4 Molecular Consequences of APA

Approximately 70% of known human genes have multiple poly(A) sites in their 3'-UTRs (Derti *et al.*, 2012) and APA may generate multiple mRNA isoforms from these genes. However, the relationship between the usage of alternate signals and the functional importance of generated isoforms still remains unclear. To clarify the consequences of APA, it would be convenient to examine under four categories (Figure 1.3):

- a) **Impact of APA on mRNA stability:** Cis-regulatory elements localized in 3'-UTRs of mRNA specifically are recognized by RNA-binding proteins (RBPs), microRNAs (miRNAs) and long non-coding RNAs (lncRNAs). miRNAs, which are ~22 nt small RNA molecules, can affect the stability and/or translation efficiency of complementary mRNAs. Most miRNA binding sites are located in 3'UTR regions of mRNAs and nearly 50% of miRNA target sites in mRNAs are affected by APA (Sandberg *et al.*, 2008; Ji *et al.*, 2009). There are numerous studies creating a link between increased mRNA half-life and 3'UTR shortening, especially in cancer cells (Mayr and Bartel, 2009; Xia *et al.*, 2014). On the other hand, there are some recent studies showing that a small fraction of shortened mRNA isoforms exhibits expression differences (Nam *et al.*, 2014; Brumbaugh *et al.*, 2018). This situation can be explained by alterations in miRNA binding efficiency: the selection of proximal APA signal results in 3'UTR shortening, which in turn affects the availability of some miRNA binding elements, but also can increase the binding efficiency of some miRNAs that bind upstream of the proximal poly (A) signal. In conclusion, shortening of 3'UTR may cause loss of some miRNA target sites and increased stability of mRNA or may enhance binding efficiency of miRNAs that are targeting sites upstream of the proximal PAS and promote degradation of the 3'UTR shortened mRNA (Figure 1.3 B) (Yuan *et al.*, 2019).

- b) **Impact of APA on mRNA translation:** Regulation of translation can be also modulated by changes in 3' UTR length of target mRNAs. A study that focuses on APA events during T cell activation demonstrated that shortened 3'-UTR length, resulting from selection of proximal poly (A) signal, gives rise to increased expression of an array of genes. This upregulated expression profile is reversed upon forced selection of distal poly (A) signals (Sandberg *et al.*, 2008). According to another study performed by Chang *et al.* (2015) activation of mammalian target of rapamycin (mTOR) results in 3' UTR shortening of mRNA molecules. A group of these transcripts exhibits increased translation profiles. Fu *et al.* (2018) also revealed that mRNA isoforms with shorter 3' UTRs were bound with more polysomes in a variety of cell lines and lead to a higher gene translational efficiency. On the contrary, there are some studies indicating that mRNA isoforms with longer 3'-UTRs are often associated with polysome fractions and display a higher translation efficiency. For instance, longer 3'UTR isoform of *Drosophila* transcript *polo*, encoding a Polo-like kinase, has a higher translational efficiency in comparison to its shorter isoform. Also, during development of the fly loss of the distal poly (A) signal of *polo* shows lethal effects at the pupa stage due to a failure in the proliferation of the precursor cells of the abdomen, the histoblasts, whereas loss of the proximal signal does not result in a phenotypic effect (Pinto *et al.*, 2011). These conflicting results suggest that APA affects mRNA translation in a gene- and cell-specific manner.
- c) **Impact of APA on mRNA nuclear export and localization:** 3'UTR regions of mRNAs also include sequence elements that are targeted by motor proteins and RNA-binding proteins (RBPs) which facilitates mobilization of these transcripts to different cellular compartments (An *et al.*, 2008; Bertrant *et al.*, 1998; Niedner *et al.*, 2014; Ephrussi *et al.*, 1991). Recent studies indicated that nearly 10% of APA isoforms display localization differences between nucleus and cytoplasm, with long 3'-UTR isoforms generally enriched in the

nucleus relative to cytoplasm (Chen and Carmichael, 2009; Neve *et al.*, 2015; Yuan *et al.*, 2019). However, it is difficult to assess whether this is due to the differences in 3'UTR length or the cytoplasmic degradation of longer isoforms. Length of 3'UTR can also affect the localization of transcripts within the cytoplasm. For example, lengthened 3'UTRs may facilitate localization of mRNAs in endoplasmic reticulum which in turn results in increased expression profiles of membrane proteins (Loya *et al.*, 2008; Reid and Nicchitta, 2015). A specific example which represents the effect of 3' length on mRNA localization comes from neural cells: The short 3'UTR isoform of brain-derived neurotrophic factor (BDNF) localizes in the soma of neurons, whereas the longer isoform is found in dendrites (Figure 1.3 C) (Lau *et al.*, 2010). These examples may indicate the importance of 3'UTR length in subcellular localization of mRNAs but further investigations are needed to elucidate mechanisms behind the process.

- d) **Impact of APA on protein localization:** The results obtained in recent studies indicated that alterations in 3'UTR lengths of transcripts may affect localizations of encoded proteins independently of localization of mRNAs. For instance, protein products coded by different 3'UTR isoforms of “cluster of differentiation 47” (CD47) exhibits difference in localization: Protein that is encoded by shorter 3'UTR isoform localizes to the ER, while the longer 3'UTR isoform encoded protein localizes to plasma membrane via interactions of RNA-binding protein Hur and effector protein SET with RAC1 (Figure 1.3 D) (Berkovits and Mayr, 2015). Integrin A1 (ITGA1) and TNF receptor superfamily member 13C (TNFRSF13C) generated by 3'UTR isoforms are also other examples that exhibit differential cellular localization (Berkovits and Mayr, 2015). There is a requirement of further studies for investigating the global effects of APA on protein localization which may lead to altered function of proteins according to their cellular settlement.

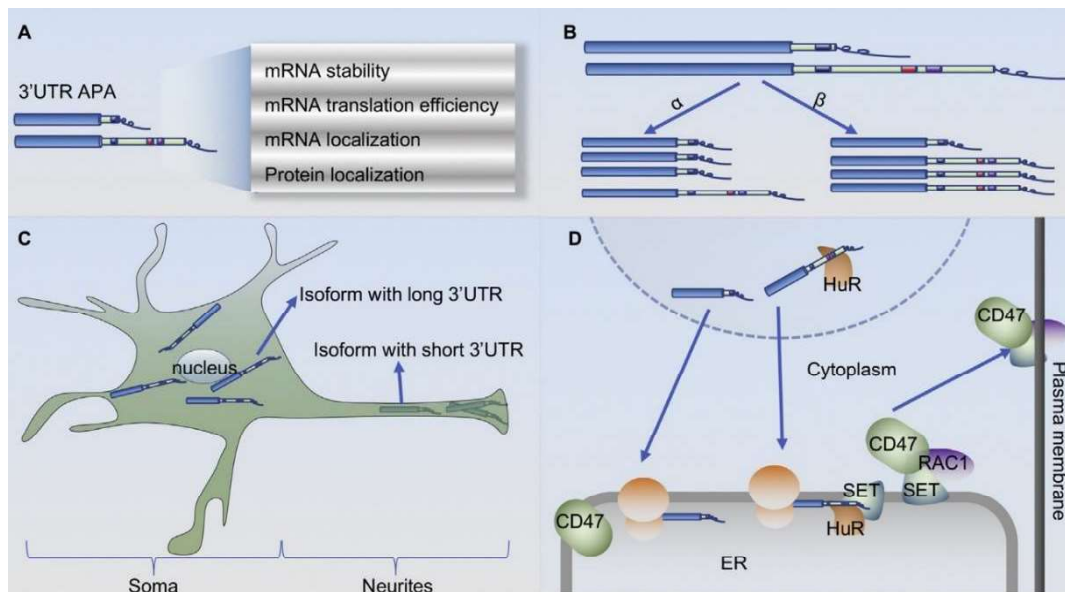


Figure 1.3. Summary of potential consequences of APA. A) APA can affect the post-transcriptional fate of transcripts in various ways. B) Long 3' UTR isoforms generally harbor more cis-regulatory elements in comparison to shorter 3' UTR isoforms which may alter translation rates and stabilities of transcripts. In  $\alpha$  pathway presence of more cis-regulatory elements (i.e. miRNA binding sites) promotes degradation of long 3' UTR isoforms. In  $\beta$  pathway 3' UTR shortening results in increased miRNA binding efficiency, thus decreasing the stability of shorter isoform. Also, the presence of cis-regulatory elements in 3' UTRs may alter the translation efficiency of shorter and longer isoforms. C) APA can affect mRNA localization within the cell. D) Cis-regulatory elements of 3'UTRs can alter the subcellular localization of the encoded protein. Figure retrieved from (Yuan *et al.*, 2019).

### 1.1.5 Possible Induction Mechanisms of APA

The exact mechanism of APA regulation in normal and cancer cells is unknown but there are some mechanistic explanations. These explanations are illustrated in Figure 1.4 and summarized below.

Some subunits of polyadenylation machinery have active roles in APA. CSTF2 subunit of the CSTF complex can be mentioned as the strongest candidate to be an APA regulator. The depletion of this protein induces the selection of distal poly (A) sites in HeLa cells (Yao *et al.*, 2012). Another study performed with mouse embryonic stem cells (ESCs) that do not express the *Cstf2* gene resulted in the loss



of differentiation potential toward the endodermal lineage while some other differentiation pathways were unaffected (Youngblood *et al.*, 2014). This result suggests that CSTF2 does not basically alter 3' UTR lengths of all genes in a cell type. CFI<sub>m</sub> complex seems to be another regulator of APA. According to the studies performed by Gruber *et al.* (2012, 2013) and Martin *et al.* (2012) CFI<sub>m</sub>25 and CFI<sub>m</sub>68 depletion resulted in the selection of proximal poly (A) sites in the HEK293 cell line. Besides, RNA-sequencing studies showed that CFI<sub>m</sub>25 knockdown resulted in shortening of mRNAs of HeLa cells (Masamha *et al.*, 2014). Another study revealed a link between the Fip1 subunit of CPSF complex and APA where the distance between the poly (A) signals and the level of Fip1 activity affects the selection of proximal and distal poly (A) sites over each other (Lackford *et al.*, 2014). Since polyadenylation occurs co-transcriptionally, transcriptional machinery may also have a regulatory role in APA. This machinery effects efficiency and specificity of polyadenylation by controlling the initiation, the rate of RNA polymerase and splicing processes (Proudfoot *et al.*, 2002). Some recent studies have linked the RNA-binding proteins (RBPs) to the APA. One of these studies is performed by Hilgers *et al.* (2012) and showed that ELAV (embryonic-lethal abnormal visual system) hinders short isoform formation by directly binding to the proximal poly (A) signals.

Oktaba *et al.* (2015) showed that inhibition of short isoform formation by ELAV may be related to the promoter sequence of target genes. Their data suggest that promoter regions of target genes carrying Trl-binding GAGA element tend to halt RNA Pol II and recruit ELAV to the site of proximal poly (A) signal on the nascent mRNA.

In addition, studies focusing on the analysis of different human tissues and cell line transcriptomes showed a strong relation between APA and alternative splicing patterns which suggest that these processes may be regulated coordinately (Wang *et al.*, 2008). For example, the depletion of U1 snRNP (small nuclear ribonucleoprotein), which is an important component of the spliceosome, leads to

the activation of intronic poly (A) signals and causes APA in a genome-wide manner (Spraggon and Cartegni, 2013).

Sequence composition of the poly (A) motif may also affect APA. For example, a single nucleotide polymorphism (SNP) modifies the AATAAA to AATACA in the 3' UTR of a known tumor suppressor, TP53. This nucleotide substitution results in the lengthening of the corresponding transcript. This SNP is associated with various cancers including; prostate cancer, colorectal adenoma, and glioma (Stacey *et al.*, 2011). There are potentially many other SNPs disrupting poly (A) signals, to affect the 3' UTR length, miRNA regulation and mRNA levels (Thomas and Saetrom, 2012).

Chromatin structure and epigenetic marks may also be listed among the regulators of APA. The relation between histone positioning and polyadenylation was initially reported in *S. cerevisiae*. Studies showed that the 3' region near the poly (A) site was depleted of nucleosomes (Mavrich *et al.*, 2008; Shivaswamy *et al.*, 2008; Spies *et al.*, 2009). This may be due to; i) the lower intrinsic affinity of the nucleotide sequence of interest for nucleosomes, or ii) the presence of a nucleosome-excluding DNA-binding protein that binds near poly(A) site (Di Giammartino *et al.*, 2011).

On the other hand, there is also a link between the proliferative signaling pathways and APA in breast cancer cells. A study performed by our group has indicated that epidermal growth factor (EGF) and estrogen (E2) treatment of EGF- and E2-responsive breast cancer cells, respectively, resulted in increased 3' UTR shortening of several mRNAs (Akman *et al.*, 2012 and 2015).

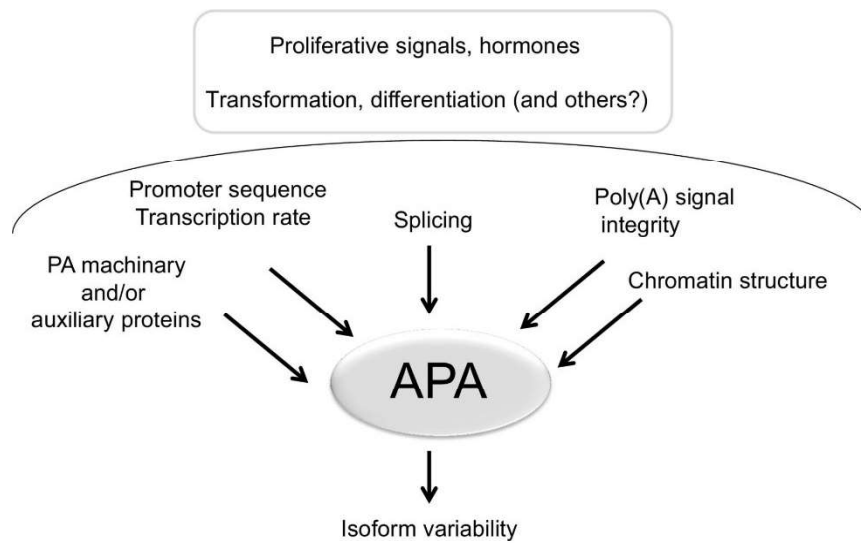


Figure 1.4. Potential APA regulators. APA can be regulated by changes in expression profiles of polyadenylation machinery proteins, sequence compositions of promoter regions, rate of transcription, integrities of poly (A) signal sequences, splicing patterns of transcripts and local chromatin structures. Figure retrieved from (Erson-Bensan and Can, 2016).

## 1.2 Aim of the Study

Alternative polyadenylation is gaining importance in regulation of gene expression. In this thesis, to better understand how APA operates in context specific manner, we aimed to identify CSTF2-interacting proteins in a breast cancer cell model.



## CHAPTER 2

### MATERIALS AND METHODS

#### 2.1 Culture Conditions for MCF7 Cells

MCF7 breast cancer cell line was cultured in Dulbecco's Modified Eagle Medium (DMEM) (01-052-1A, BI) supplemented with 10% Fetal Bovine Serum (FBS) (S1810-500, Biowest), 2% L-Glutamine (01-115-1A, BI), 1% Sodium Pyruvate Solution (03-042-1B, BI) and 1% Penicillin/Streptomycin (Pen/Strep) (03-031-1B, BI). Cells were incubated at 37 °C with 95% humidified air and 5% CO<sub>2</sub> and were grown as monolayers.

Before hormone stimulation experiments, cells were grown in phenol red free medium (01-053-1A, BI) supplemented with 10% charcoal-dextran treated FBS (CD-FBS), 2% L-Glutamine, 1% Sodium Pyruvate solution and 1% Pen/Strep solution for 72 hours. MCF7 cells were treated with 100 nM 17  $\beta$ -Estradiol (E2) (Sigma-Aldrich) or with ethanol (vehicle control) for 45 minute, 3 hours and 12 hours after reaching 70 – 80 % confluency (Lee and Sheen, 1997).

#### 2.2 Preparation of CD-FBS for Hormone Treatment Experiments

Preparation of CD-FBS were performed according to the protocol provided by Dr. Mesut Mutan (METU, Ankara). Briefly; 10 g of dextran-coated charcoal (C6241, Sigma-Aldrich) were added to 500 ml FBS and incubated overnight at 4°C on a magnetic stirrer. The following day FBS-charcoal mixture were divided into two sterile centrifuge tubes and centrifuged at 10800xg for 30 minutes at 4°C. In a biological safety cabinet, the supernatants were carefully transferred into a 0.45  $\mu$ m filter unit (CLS430770, Sigma-Aldrich) and filtered. The filtrate again mixed with 10 g dextran-coated charcoal and incubated at 4°C for 4 – 6 hours on a magnetic stirrer. The mixture again divided into two sterile centrifuge tubes and centrifuged at

10800xg for 30 minutes at 4°C. In a biological safety cabinet, the supernatants again were carefully transferred into a 0.45 µm filter unit and filtered. The filtrate aliquoted into 50 ml sterile falcon tubes and stored at -20°C.

### 2.3 Expression Analysis

MIQE Guidelines and checklist (Bustin *et al.*, 2009) were followed during expression analysis experiments including RNA isolation, RNA quantification, cDNA synthesis and RT-qPCR analysis.

#### 2.3.1 RNA Isolation

Total RNA was isolated with High Pure RNA Isolation Kit (11828665001, Roche) according to the instructions of the manufacturer. To eliminate DNA contamination, the isolated total RNA samples were incubated with DNase I (04716728001, Roche) at 37°C for 2 hours according to the given conditions in Table 2.1.

Table 2.1 DNase I Treatment Reaction Conditions

Components	Amount
Total RNA	75 µl
10x Incubation Buffer	10 µl
DNase I recombinant, RNase-free (10 units/ul)	5 µl
mgH <sub>2</sub> O	10 µl
Total Reaction Volume	100 µl

At the end of the incubation period for DNase treatment, equal amount of (100 µl) Phenol:Chloroform:Isoamyl alcohol (25:24:1) (pH 4.0) were added on the samples. The tubes were vortexed at highest speed for 30 seconds and incubated 10 minutes on ice. Following centrifugation at 14000xg for 20 minutes at 4°C, 80 µl of the upper phase, containing RNAs, were transferred into a fresh tube. 240 µl of molecular

grade ethanol (EtOH) and 8 µl of 3M Sodium Acetate (NaAc) were added to samples and the mixtures were incubated at -20°C, overnight. Following the centrifugation at 14000xg for 30 minutes at 4°C, the supernatants were removed. The pellets were washed with 600 µl of 70% molecular grade EtOH. Following the centrifugation at 14000xg for 15 minutes at 4°C the supernatants were removed and the pellet, containing extracted RNAs, were dissolved in 25 µl mgH<sub>2</sub>O.

Following DNase I treatment and phenol-chloroform extraction, elimination of DNA contamination confirmed by Glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) specific Polymerase Chain Reaction (PCR) with *GAPDH\_F*: 5'-GGGAGCCAAAAGGGTCATCA-3' and *GAPDH\_R*: 5'-TTTCTAGACGGCAGGTCAGGT-3' primer pair. PCR performed according to the given conditions: initial denaturation at 95°C for 3 minutes, 40 cycles of 95°C for 30 seconds, 56°C for 30 seconds and 72°C for 60 seconds, and a final extension at 72°C for 5 minutes. MCF7 genomic DNA was used as positive control for PCR reactions.

Following the confirmation of DNA contamination elimination, RNA samples were spectrophotometrically quantified by using NanoDrop (MN-913, Maestrogen). Purity of RNA samples were determined by evaluation of A260/A280 and A260/A230 ratios. A representative result is given in Appendix showing a proper RNA samples with appropriate A260/A280 and A260/A230 ratios. All RNA samples that were used in this study were fitting the purity criteria.

### **2.3.2 cDNA Synthesis from RNA Samples**

cDNA synthesis performed by using RevertAid First Strand cDNA Synthesis Kit (K1622, Thermo Fisher Scientific) using oligo<sub>d</sub>T primers and 1000 ng total RNA. Reaction conditions given in Table 2.2 were followed for cDNA synthesis.

Table 2.2 cDNA Synthesis Reaction Conditions

<b>Components</b>	<b>Amount</b>
Total RNA	1000 ng
Oligo(dT) <sub>18</sub> Primer, 100 μM	1 μl
mgH <sub>2</sub> O	upto 12 μl
Mix thoroughly, spin down, incubate at 65 °C for 5 minutes	
10 mM dNTP Mix	2 μl
5X Reaction Buffer (250 mM Tris-HCl (pH 8.3), 250 mM KCl, 20 mM MgCl <sub>2</sub> , 50 mM DTT)	4 μl
Ribolock RNase Inhibitor (20U/μl)	1 μl
Revertaid Reverse Transcriptase (200U/μl)	1 μl
Total Volume	20 μl
Incubate at 42 °C for 60 minutes and 70 °C for 5 minutes. Store synthesized cDNA samples at -20 °C.	

### 2.3.3 Expression Analysis for *TFF1* Gene

Trefoil factor 1 (*TFF1*, also known as *pS2*) is a known E2 responsive gene (Brown *et al.*, 1984). E2 stimulated expression of this gene was evaluated by PCR as a positive control for E2 treatment experiments. The following primer pair was used for PCR analysis *TFF1\_F*: 5'-CCATGGAGAACAAGGTGATCTGC-3' and *TFF1\_R*: 5'-GTCAATCTGTGTTGTGAGCCGAG-3' (product size: 399 bp). PCR performed according to the given conditions: initial denaturation at 95°C for 3 minutes, 18 cycles of 95°C for 30 seconds, 61°C for 30 seconds and 72°C for 60 seconds, and a final extension at 72°C for 5 minutes.



## 2.4 Immunocytochemistry

For immunocytochemistry (ICC) experiments, MCF7 cells grown on coverslips. Following the incubation period (24 hours for transfected cells, 45 minutes and 3 hours for E2/H<sub>2</sub>O<sub>2</sub> treated cells), the coverslips were washed three times with D-PBS. Fixation of the samples were performed with 3.7 % formaldehyde with 30 minutes incubation at room temperature. The cells were again washed with DPBS for 3 times and permeabilized by incubation with 0,4% Triton X-100 in DPBS at room temperature for 10 minutes. Following the permeabilization step, 10% BSA (11943.01, Serva) in PBS was used as a blocking reagent. The cells were incubated in this solution for 1 hour at room temperature. Primary antibodies diluted in 1x D-PBS supplemented with 3% BSA were added on cells and incubated for 2 hours at room temperature. Concentrations of the primary antibodies were as follows: 2µg/ml for anti-HA antibody (ab9110, Abcam) and 1µg/ml for anti-CSTF2 antibody (ab72297, Abcam). Subsequently, cells were again washed 3 times with 1x D-PBS and incubated with Alexa Fluor 488-conjugated secondary antibodies (ab150077, Abcam; for anti-CSTF2 and anti-HA tag antibodies) for 30 minutes at room temperature. Secondary antibodies were 1:1000 diluted in 1xD-PBS supplemented with 3% BSA. The cells were washed 3 times with 1xD-PBS and mounted on slides using DAPI-containing mounting medium (ab104139, Abcam) for nuclear staining and visualized under a Leica Laser Scanning Confocal Microscope equipped with a Leica 63×/1.32 HCX PL APO OilDIC objective, VT-Hawk 2D array laser scanning confocal microscopy system (Visitech International, UK) with an Olympus IX-83 inverted microscope equipped with 100×/1.49 Oil UAPO lens (Olympus) and EM-CCD digital camera (Hamamatsu) (Department of Microbiology and Department of Biochemistry, Cell, and Molecular Biology, University of Tennessee, Knoxville, Tennessee) and Leica DMI4000 with Andor DSD2Spinning Disc Confocal Microscope equipped with a Leica 100×/1.49Oil DIC objective (Department of Biological Sciences, Middle East Technical University, Ankara, Turkey) (Cevheroglu *et al.*, 2017)

## 2.5 Cloning Coding Sequence of *CSTF2* Gene into pcDNA3.1 mycBioID Vector

Coding sequence of *CSTF2* gene was retrieved from NCBI (National Center for Biotechnology) with the accession number: NM\_001325.2. Specific forward and reverse primers harboring NotI and KpnI restriction enzyme recognition sites, respectively, were designed for PCR amplification of 1734 bp long full coding sequence and cloning into pcDNA 3.1 mycBioID (#35700, Addgene) to generate myc\_BirA\*\_*CSTF2* fusion plasmid. The generated plasmid encodes for *CSTF2* protein fused to C-terminus of promiscuous biotin ligase BirA\* [or BirA(R118G)], a mutated form of bacterial “DNA-binding transcriptional repressor/biotin-[acetyl-CoA-carboxylase] ligase *BirA*, for proximity dependent biotin identification (BioID) of potential interacting partners of *CSTF2*. Vector map of pcDNA 3.1 mycBioID is given in Appendix. Sequences of the cloning primers are shown in the Table 2.3: the underlined letters indicate recognition aiding sites, bold letters indicate restriction enzyme recognition sites, lowercase letters in forward primer are additional nucleotides to maintain the open reading frame of *CSTF2*.

Table 2.3 Primers for Cloning *CSTF2* Gene into pcDNA3.1 mycBioID Vector

BirA (R118G)- <i>CSTF2</i> Fusion Forward	<u>CGCATGCGGCCGC</u> aaATGGCGGGTTTGACTGTGA
BirA (R118G)- <i>CSTF2</i> Fusion Reverse	<u>CGCATGGTACCTCAAGGTGCTCCAGTGGATTT</u>

PCR mixture was prepared as indicated in Table 2.4. Amplification of the gene was performed using Phusion High-Fidelity DNA Polymerase (F530L, Thermo Fisher Scientific) and the conditions for amplification were as follows: initial denaturation

at 98 °C for 3 minutes; 35 cycles of 98 °C for 30 seconds, 61 °C for 25 seconds, 72 °C for 45 seconds; and final extension at 72 °C for 5 minutes.

Table 2.4 PCR Mixture Preparation for Amplification of Desired Sequences

<b>1x PCR Mixture</b>	
5x Phusion HF Buffer	4 µl
10 mM dNTPs	2 µl
5 µM Forward Primer	2 µl
5 µM Reverse Primer	2 µl
Template DNA	variable
Phusion DNA Polymerase (2U/µl)	0.2 µl
Molecular grade water (mgH <sub>2</sub> O)	to 20 µl

PCR products were loaded on 1% agarose gel. After running at 100V and visualization under UV-light, the DNA fragments with expected band sizes were cut and isolated from agarose gel by using Zymoclean Gel DNA Recovery Kit (D4008, Zymo Research) according to the instructions of the manufacturer. Purified DNA fragments and pcDNA 3.1 mycBioID empty vector were quantified with NanoDrop (MN-913, Maestrogen). Eluted PCR products and empty pcDNA 3.1 mycBioID vector (5 µg) were double digested with NotI and KpnI restriction enzymes (respectively, FD0594 and FD0524, Thermo Fisher Scientific) for 4 hours at 37 °C. Digested PCR fragments and empty vector were loaded on 1% agarose gel and the end of the running period, purified from agarose gel. Eluted restriction fragments were quantified by NanoDrop again.

Digested pcDNA 3.1 mycBioID empty vector and PCR products were mixed with different ratios and ligated with using T4 DNA Ligase (EL0011, Thermo Fisher Scientific). Calculations for different mixture ratios were performed according to the following formula:

$$\frac{\text{vector amount (ng)} \times \text{insert size (kb)}}{\text{vector size (kb)}} \times \frac{\text{insert}}{\text{vector}}$$

$$= \text{insert amount to be used (ng)}$$

For example, to get 3:1 ratio;

$$\frac{100 \text{ ng vector} \times 1734 \text{ kb}}{6410 \text{ kb}} \times \frac{3}{1} = \sim 82 \text{ ng insert to be used.}$$

Ligation mixtures were incubated at 16 °C for 16h. At the end of the incubation period, ligation products were transformed into competent *E. coli* TOP10 cells as explained in Section 2.10.

## 2.6 Cloning *CSTF2* Gene into pcDNA3.1 MCS-BirA(R118G)-HA Vector

Specific forward and reverse primers harboring NheI (FD0974, Thermo Scientific Fisher) and AflII (FD0834, Thermo Scientific Fisher) restriction enzyme recognition sites, respectively, were designed for PCR amplification of *CSTF2* gene (NM\_001325.2) and cloning into pcDNA3.1 MCS-BirA(R118G)-HA Vector (#36047, Addgene) to generate *CSTF2*\_BirA\*\_HA fusion plasmid. The full coding sequence of the gene was 1734 bp long, but during PCR amplification last 2 nucleotides of the gene was deleted for disruption of stop codon and also conservation of the open reading frame for production of *CSTF2* protein fused to N-terminus of promiscuous biotin ligase BirA\* (*CSTF2*-BirA\* fusion). The 1732 bp long coding sequence of *CSTF2* was PCR amplified with cloning primers using Phusion High-Fidelity DNA Polymerase. Vector map of pcDNA3.1 MCS-BirA(R118G)-HA is given in Appendix. Sequences of the cloning primers are given in the Table 2.5: as in Table 2.3 recognition enhancer sites are given with underlined letters and restriction enzyme recognition sites are given in bold letters.

Table 2.5 Primers for Cloning *CSTF2* gene into pcDNA3.1 MCS-BirA(R118G)-HA Vector

CSTF2-BirA Fusion Forward	<u>CGCAT</u> <b>GCTAGCATGGCGGGTTT</b> GACTGTGA
CSTF2-BirA Fusion Reverse	<u>CGCAT</u> <b>CTTAAGCAAGGTGCTCCAGTGGATTT</b>

The conditions used for PCR amplification were as follows: initial denaturation at 98 °C for 3 minutes; 35 cycles of 98 °C for 30 seconds, 64 °C for 25 seconds, 72 °C for 45 seconds; and final extension at 72 °C for 5 minutes.

After PCR amplification, agarose gel electrophoresis, gel extraction, double digestion and ligation steps were performed as described in Section 2.1, except for minor differences in used restriction enzymes and empty vectors. The ligation products were transformed into competent *E. coli* TOP10 cells as explained in Section 2.10.

## 2.7 Cloning *CSTF2* Gene into pcDNA3.1 (-) Vector

Specific forward and reverse primers harboring XbaI (FD0684, Thermo Fisher Scientific) and EcoRV (FD0303, Thermo Fisher Scientific) restriction enzyme recognition sites, respectively, were designed for PCR amplification of *CSTF2* gene (NM\_001325.2) with Phusion High-Fidelity DNA Polymerase. During PCR, last 3 nucleotides of the gene was deleted to disrupt stop codon for subsequent cloning experiments and 1731 bp of the coding sequence was cloned into mammalian expression vector pcDNA 3.1(-) (Life Technologies) to generate *CSTF2*\_pcDNA3.1 (-) plasmid. Vector map of pcDNA 3.1 (-) was given in Appendix. Sequences of the cloning primers are given in the Table 2.6: as in Table 2.3, the underlined letters indicate recognition enhancer sites, the bold letters indicate restriction enzyme recognition sites, and the lowercase letters in forward primer indicates Kozak's sequence added to increase expression levels.

Table 2.6 Primers for Cloning CSTF2 Gene into pcDNA3.1 (-) Vector

CSTF2- TurboID-HA Fusion Forward	5'- <u>CGCATTCTAGA</u> aaaatgATGGCGGGTTTGACTGTGAG- 3'
CSTF2- TurboID-HA Fusion Reverse	5'- <u>CGCATGATATCAGGTGCTCCAGTGGATTTCTG</u> -3'

Following conditions were used for PCR amplification *CSTF2* gene: initial denaturation at 98 °C for 3 minutes; 35 cycles of 98 °C for 10 seconds, 57 °C for 25 seconds, 72 °C for 45 seconds; and final extension at 72 °C for 5 minutes.

After PCR amplification, agarose gel electrophoresis, gel extraction, double digestion and ligation steps were performed as described in Section 2.1, except for minor differences in used restriction enzymes and empty vectors. The ligation products were transformed into competent *E. coli* TOP10 cells as explained in Section 2.10.

## 2.8 Cloning TurboID into CSTF2\_pcDNA3.1 (-) Vector

Promiscuous biotin ligase TurboID (Branon *et al.*, 2018), another mutated form of bacterial “DNA-binding transcriptional repressor/biotin-[acetyl-CoA-carboxylase] ligase *BirA*” with a higher catalytic activity in comparison to *BirA*\*, was PCR amplified with cloning primers using Phusion High-Fidelity DNA Polymerase and cloned downstream of *CSTF2* in pcDNA 3.1 (-) to generate CSTF2\_TurboID\_HA plasmid. Plasmid used as template for PCR amplification of TurboID was a kind gift from Dr. Mesut Muyan (METU, Ankara). Sequences of forward and reverse cloning primers harboring KpnI and AflIII recognition sites, respectively are given in Table

2.7: the underlined letters indicate recognition enhancer sites and the bold letters indicate restriction enzyme recognition sites.

Table 2.7 Primers for Cloning TurboID into CSTF2\_pcDNA3.1(-) Vector

TurboID Cloning Forward	5'– <u>CGCAT</u> <b>GGTACCGCGGCCCGCAGGCGGCA</b> –3'
TurboID Cloning Reverse	5'– <u>CGCAT</u> <b>CTTAAGTTTATTAAGCGTAATCTGGAACATC</b> GTATGGGTAAC–3'

Following PCR conditions were used for amplification of TurboID: initial denaturation at 98 °C for 3 minutes; 35 cycles of 98 °C for 10 seconds, 59 °C for 25 seconds, 72 °C for 30 seconds; and final extension at 72 °C for 5 minutes.

After PCR amplification, agarose gel electrophoresis, gel extraction, double digestion and ligation steps were performed as described in Section 2.1, except for minor differences in used restriction enzymes and empty vectors. The ligation products were transformed into competent *E. coli* TOP10 cells as explained in Section 2.10.

## 2.9 Transformation of Competent *E. coli* TOP10 Cells

Transformation of competent *E. coli* TOP10 cells were performed as described (Green and Sambrook, 2012). Briefly; 50 µl competent cells were thawed on ice, mixed with 2 µl ligation product and incubated on ice for 30 minutes. Following heat shock step performed at 42 °C for 5 minutes, the cells again incubated on ice for 5 minutes. 500 µl Nutrient Broth (NB) medium was added on reaction mixture and cells were incubated at 37 °C for 90 minutes. NB agar plates supplemented with 100 µg/ml ampicillin were inoculated with 100 µl of transformed bacteria and overnight incubated at 37 °C.

Overnight grown colonies were screened by colony PCR using cloning primers. After PCR confirmation, positive colonies carrying the desired vector were selected for inoculation of NB media and incubated overnight at 37 °C with 180 rpm shaking. The following day, plasmids were isolated using GeneJET Plasmid Miniprep Kit (K0502, Thermo Fisher Scientific) according to the instructions of the manufacturer.

The transformation and colony PCR steps explained above were performed for all cloning/ligation products mentioned in Sections 2.6 – 2.9.

## **2.10 Transient Transfection of MCF7 Cells**

Before transfection of MCF7 breast cancer cells with myc\_BirA\*\_CSTF2, CSTF2\_BirA\*\_HA or CSTF2\_TurboID\_HA plasmids and control vectors of these plasmids the cells were seeded in flasks and incubated at 37 °C with 95% humidified air and 5% CO<sub>2</sub> to reach 70-80% confluency. TurboFect Transfection Reagent (R0532, Thermo Fisher Scientific) was used 1:2 (plasmid:TurboFect) ratio. For instance, for transfecting cells grown in a T25 flask; 5 µg plasmid was diluted in 500 µl DMEM and mixed with 10 µl transfection reagent. Whole mixture was incubated at room temperature for 15-20 minutes and then was added to each flask drop wise. The flasks were gently rocked and cells were incubated at 37 °C with 95% humidified air and 5% CO<sub>2</sub>.

At the end of the incubation period, cells were either collected directly for RNA and protein isolation or treated with 50 µg D-Biotin (B-4639, Sigma Aldrich) and 1mM ATP (A2383, Sigma-Aldrich) for 3 hours before harvesting (Branon *et al.*, 2018).

## **2.11 Proximity-Dependent Biotinylation and Biotin Affinity Capture**

Proximity-dependent biotinylation approach, which utilizes a promiscuous biotin ligase to detect protein-protein interactions within living cells, was utilized for identification of interacting partners of CSTF2 protein in MCF7 cells. Biotinylated proteins were obtained with streptavidin-affinity capture and identified by nano-liquid chromatography mass spectrometry (n-LC MS/MS). BioID experiments were



performed using myc\_BirA\_CSTF2 and CSTF2\_BirA\_HA plasmids, whereas CSTF2\_TurboID\_HA plasmid was used for TurboID experiments. MCF7 cells were transfected with vectors carrying *CSTF2*-coding sequence fused with BirA\*/TurboID or empty vector carrying only the BirA\*/TurboID. 24 hours after transfection 50 µg D-Biotin (B-4639, Sigma Aldrich) and 1mM ATP (adenosine 5'-triphosphate disodium salt hydrate) (A2383, Sigma-Aldrich) were included to the medium. At the end of the incubation period (3 hours for TurboID plasmid, 18 hours for BioID plasmids) existing medium were removed, cells were collected and washed twice with ice-cold DPBS (Dulbecco's Phosphate Buffered Saline) (02-023-1A, Biological Industries). Nuclear lysate isolation from these cells were performed using NE-PER Nuclear and Cytoplasmic Extraction Reagents (78835, Thermo Scientific). Obtained lysates were incubated with adjusted volumes of high capacity neutravidin agarose resin (29204, Thermo Fisher Scientific) overnight. The following day, beads were collected with centrifugation at 2500 rpm for 5 minutes and washed twice with Wash Buffer I [2 % Sodium Dodecyl Sulfate (SDS)], once with Wash Buffer II (2% deoxycholate, 1% Triton-X, 50 mM NaCl, 50 mM Hepes, pH=7.5, 1 mM EDTA), once with Wash Buffer III (0.5% NP-40, 0.5% deoxycholate, 1% Triton-X, 500 mM NaCl, 1 mM EDTA, 10mM Tris, pH=8.1), and once with Wash Buffer IV (50mM Tris, pH=7.4, 50 mM NaCl). All washing steps were carried out room temperature for 10 minutes. 10% of the beads were separated and mixed with 50 µl of Laemmli buffer containing 500 nM D-Biotin and incubated at 100 °C for 10 minutes. Following the incubation, supernatant containing the bound proteins were used for western blot analysis.

## **2.12 Mass Spectrometry Analysis**

The remaining 90% of the beads were sent to Mass Spectrometry analysis at Koç University Proteomics Facility (Istanbul, Turkey). The bead-bound proteins were digested with trypsin (Pierce™ Trypsin Protease, MS Grade, Thermo Scientific). Peptide analysis was performed with Thermo Scientific Q Exactive Quadrupole-Orbitrap Mass Spectrometer with nano Liquid Chromatography, UltiMate 3000

NCS-3500RS (n-LC MS/MS). Thermo Discoverer 2.3 software was used to identify proteins.

## **2.13 Western Blot Analysis**

### **2.13.1 Nuclear Lysate Isolation**

Nuclear lysates of E2 treated MCF7 cells were isolated either by using NE-PER Nuclear and Cytoplasmic Extraction Reagents (78835, Thermo Scientific) according to the instructions of the manufacturer or following the protocol described (Akman *et al.*, 2012). Briefly; MCF7 cells were washed two times with D-PBS, collected into 1.5 ml tubes and centrifuged at 1500 rpm for 5 minutes at 4°C. Supernatants were removed and the cell pellets were resuspended in Lysis Buffer A (composition given in Table 2.8) and incubated on ice for 30 minutes. Cells were vortexed at the highest speed for 15 seconds. NP-40 was added on cell suspensions at a final concentration of 0,15% and cells were again incubated on ice for 5 minutes. Cell suspensions were vortexed at the highest speed for 5 seconds and centrifuged at 12000xg for 1 minute at 4°C. Supernatants containing cytoplasmic proteins were transferred into a clean tube and stored on ice for concentration measurement. Pelleted nuclei were washed 3 times with Lysis Buffer A for removal of residual cytoplasmic proteins and resuspended in Lysis Buffer B (composition given in Table 2.8). Following the incubation on ice for 30 minutes, the tubes were centrifuged at 12000xg for 10 minutes at 4°C. BCA Protein Assay Kit (23225, Pierce) was used for quantification of lysates. Isolated nuclear and cytoplasmic lysates were stored at -80°C.

Table 2.8 Compositions of Lysis Buffers

	<b>Lysis Buffer A</b>	<b>Lysis Buffer B</b>
HEPES*	30mM	30mM
KCl*	10mM	10mM
MgCl <sub>2</sub> *	1.5mM	1.5mM
EDTA*	0.1mM	0.1mM
DTT*	1mM	1mM
NaCl*	-	500mM
Glycerol	-	25%

\* HEPES, 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid; KCl, potassium chloride; MgCl<sub>2</sub>, magnesium chloride; EDTA, Ethylenediaminetetraacetic acid; DTT, 1,4-dithiothreitol; NaCl, sodium chloride.

\*\*Both buffers were supplemented with a tablet of protease inhibitor cocktail (11873580001, Sigma-Aldrich).

### 2.13.2 Western Blotting

Nuclear lysates were mixed with 6X Laemmli Buffer (60% Glycerol, 12% SDS, 30% 2-mercaptoethanol, 0.375 M Tris and 0.012% bromophenol blue) and denatured by incubating at 100 °C for 10 minutes. Denatured lysates were separated on 10% polyacrylamide gel and transferred to a nitrocellulose membrane (Bio-Rad). The membranes were blocked with 5% skimmed milk or bovine serum albumin (BSA) (11943, Serva) in Tris Buffered Saline supplemented with 0.1% Tween 20 (0.1% TBS-T) for 1 hour at room temperature and overnight incubated with appropriate primary antibodies (antibodies used western blot analyses performed in this study are collectively listed in Table 2.9.). At the end of the incubation period, membranes were washed 3 times with 0,1% TBS-T for 10 minutes and incubated with appropriate HRP-conjugated secondary antibodies at room temperature for 1 hour. Proteins were visualized using western blotting detection kit (WestBright ECL, Advansta) according to the manufacturer's instructions.

Table 2.9 List of the antibodies used in western blot analyses throughout this study

<b>Antibody Name and Catalog Information</b>	<b>Dilution Information</b>	<b>Blockin Reagent</b>
Rabbit Anti-Cstf2 antibody (ab72297, Abcam)	1:1000 in 0.1% TBS-T	5% Non-Fat Dry Milk (170-6404, Bio-Rad)
Rabbit Anti-CFIm25 antibody (10322-1-AP, Proteintech)	1:200 in 0.1% TBS-T	5% Non-Fat Dry Milk (170-6404, Bio-Rad)
Mouse Anti-HDAC1 antibody (sc-81598, Santa Cruz Biotechnology)	1:1000 in 0.1% TBS-T	5% Non-Fat Dry Milk (170-6404, Bio-Rad)
Mouse HRP-Conjugated Alpha Tubulin antibody (HRP-66031, Proteintech)	1:2000 in 0.1% TBS-T	5% Non-Fat Dry Milk (170-6404, Bio-Rad)
Rabbit Anti-Biotin antibody (ab53494, Abcam)	1:200 in 0.05% TBS-T	5% Non-Fat Dry Milk (170-6404, Bio-Rad)
Rabbit HA-Tag antibody (ab9110, Abcam)	1:2000 in 0.1%TBS-T	5% Non-Fat Dry Milk (170-6404, Bio-Rad)
Mouse Anti-Cstf-77 Antibody (sc-376553, Santa Cruz Biotechnology)	1:500 in 0.1% TBS-T	5% Albumin Bovine Modified Cohn Fraction V, pH 7.0 (11943.01, Serva)
Goat-anti-rabbit HRP secondary antibody (R-05072-500, Advansta)	1:2000 in 0.1% TBS-T	Depending on primary antibody used
Goat-anti-mouse HRP secondary antibody (R-05071-500, Advansta)	1:2000 in 0.1% TBS-T	Depending on primary antibody used

## CHAPTER 3

### RESULTS AND DISCUSSIONS

#### 3.1 An Investigation of APA in ER+ Breast Cancer Cells

This section will summarize findings on how APA may operate in response to E2. To begin understanding APA, we wanted to identify proteins that may interact with the polyadenylation machinery. We took the proximity-dependent biotin identification and mass spectroscopy (MS) approach to discover poly (A) machinery interacting proteins.

#### 3.2 Investigation on Interacting Partners of CSTF2 via Proximity-Dependent Biotin Identification

##### 3.2.1 Expression of BirA\*-Fused CSTF2 Protein in HEK293 Cells

Proximity-dependent biotin identification (BioID) method utilizes a promiscuous biotin ligase for biotinylation of proteins in a proximity-dependent fashion. A protein of interest is fused to the promiscuous biotin ligase and expressed in cells, where it biotinylates proteins in close-proximity, with a ~10nm of labelling radius (Kim *et al.*, 2014). Following the biotinylation process, affinity-capture method enables selective isolation and identification of the biotinylated proteins. These proteins are considered as potential interacting partners of the protein of interest. The biotin retention protein A (BirA, also known as biotin protein ligase) of *E. coli* is a bi-functional protein which acts as the biotin operon repressor and as the enzyme catalyzing the biotinylation of a subunit of acetyl-CoA carboxylase (Buoncristiani *et al.*, 1986; Satiaputra *et al.*, 2019). The wild-type BirA enzyme catalyzes biotin addition in two steps: first, generates reactive biotinoyl-5'-AMP (bioAMP) from

ATP and biotin and second, attaches this bioAMP to a specific lysine residue on acetyl-CoA carboxylase (Roux *et al.*, 2018). BirA mutant (R118G, also known as BirA\*), which is defective in both self-association and DNA-binding, displays an affinity for bioAMP two orders of magnitude less than its wild-type form (Kwon *et al.*, 2009; Kwon and Beckett, 2000). This results in a premature release of bioAMP, which will covalently react with adjacent primary amines (Roux *et al.*, 2018). Humanized form of the BirA\* (Mechold *et al.*, 2005) is used for biotinylation of proteins in live mammalian cells in BioID approaches (Roux *et al.*, 2012).

To perform proximity dependent biotinylation of potential interactors of CSTF2 protein we genetically fused CSTF2 cDNA to BirA\*. Figure 3.1 and 3.2 shows the schematic representation of constructs generated for production of C-terminally and N-terminally fused proteins.

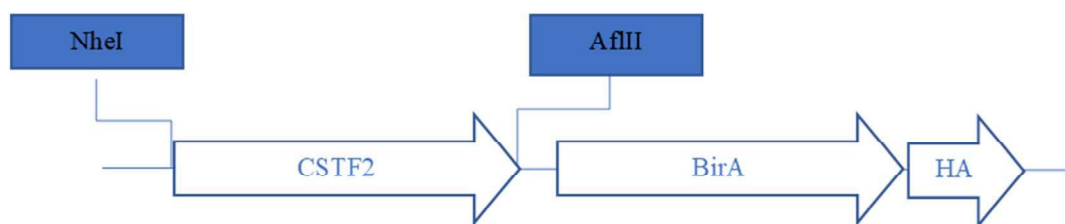


Figure 3.1. Cloning CSTF2 coding sequence into pcDNA3.1 MCS-BirA(R118G)-HA (addgene, plasmid #36047) vector for creating *CSTF2* C-terminal fusion with *BirA*.

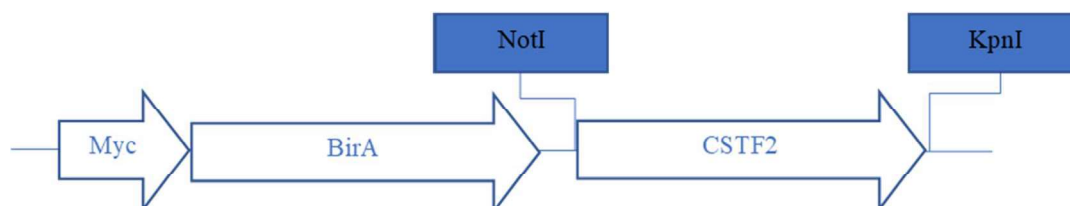


Figure 3.2. Cloning CSTF2 coding sequence into pcDNA3.1 mycBioID (addgene, plasmid #35700) vector for creating *CSTF2* N-terminal fusion with *BirA*.

After cloning the CSTF2 coding sequence into pcDNA3.1 MCS-BirA(R118G)-HA and pcDNA3.1 mycBioID plasmids, these constructs were used for transfection into HEK293 cells (Figure 3). Transfected HEK293 cells were collected after 24 and 72 hours of transfection and used for nuclear and cytoplasmic protein extraction with NE-PER™ Nuclear and Cytoplasmic Extraction Reagents (78835, Thermo Scientific). Isolated lysates were subjected to western blot analysis for evaluation of expression patterns and investigation of subcellular localization of the fused proteins (Figure 3.3).

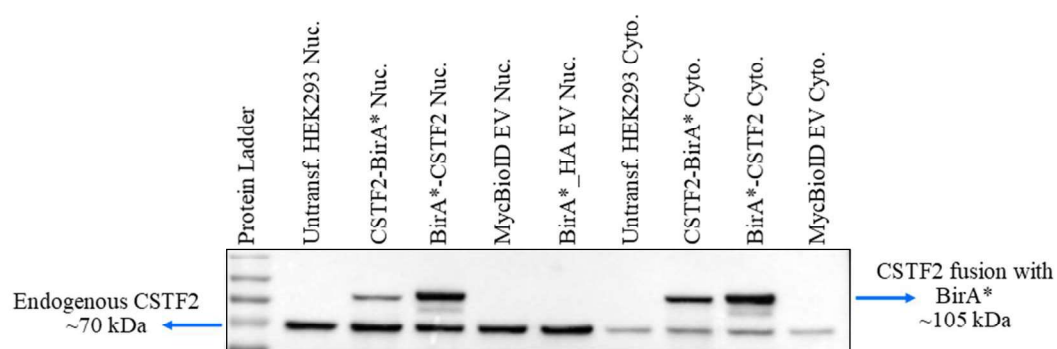


Figure 3.3. Transfection of HEK293 cells with CSTF2 and BirA\* fusion constructs. Lysates isolated from transfected HEK293 cells 24-hour post-transfection and western blotting performed by using anti-CSTF2 antibody (ab72297, Abcam). CSTF2-BirA: CSTF2 in pcDNA3.1 MCS-BirA(R118G)-HA; BirA-CSTF2: CSTF2 in pcDNA3.1 mycBioID; MycBioID: pcDNA3.1 mycBioID Empty Vector; BirAHA: pcDNA3.1 MCS-BirA(R118G)-HA.

This result indicated that transfected cells are capable of using these recombinant constructs for producing fusion CSTF2-BirA\* and/or BirA\*-CSTF2 proteins. Then, the membranes were stripped for visualization using anti-HDAC1 antibody to control the loaded amounts of proteins (Figure 3.4)

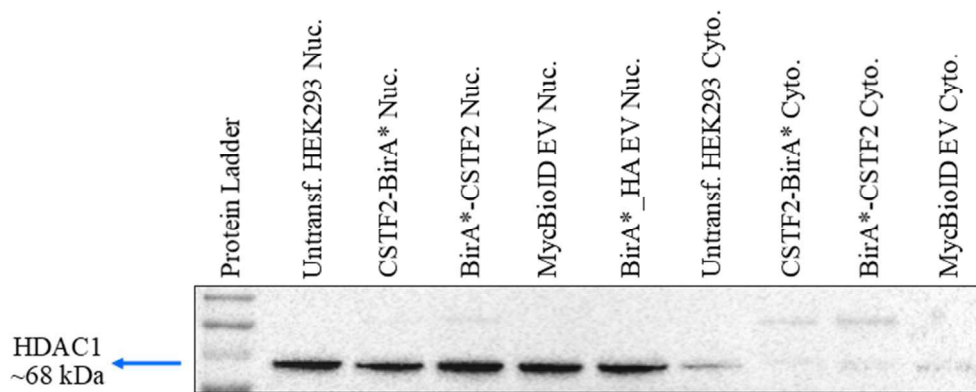


Figure 3.4. Transfection of HEK293 cells with CSTF2 and BirA\* fusion constructs. Lysates isolated from transfected HEK293 cells 24-hour post-transfection and western blotting performed by using anti-HDAC1 antibody (Santa Cruz, sc-81598). CSTF2-BirA: CSTF2 in pcDNA3.1 MCS-BirA(R118G)-HA; BirA-CSTF2: CSTF2 in pcDNA3.1 mycBioID; MycBioID: pcDNA3.1 mycBioID Empty Vector; BirAHA: pcDNA3.1 MCS-BirA(R118G)-HA.

This result showed that the loaded amounts of nuclear proteins isolated from HEK293 cells transfected with different plasmids were nearly equal and also indicated that there is a minimal nuclear-to-cytoplasmic leakage during fractionated lysate isolation.

After taking western blot results of 24h post-transfection samples we performed the same experiments with lysates isolated 72-hour transfection set of HEK293 cells (Figure 3.5).



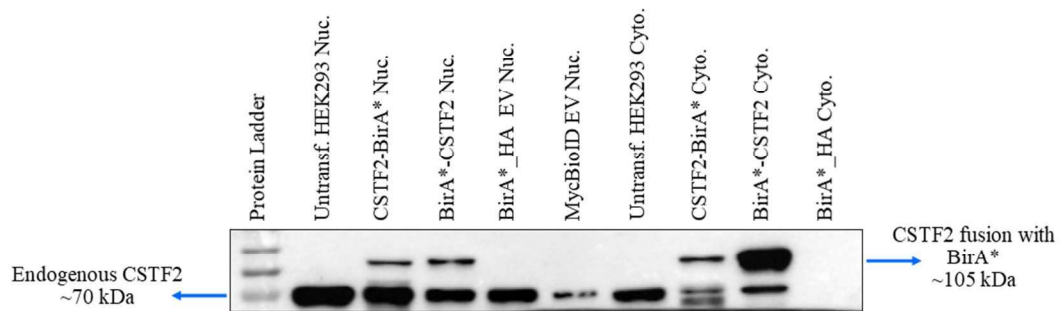


Figure 3.5. Transfection of HEK293 cells with CSTF2 and BirA\* fusion constructs. Lysates isolated from transfected HEK293 cells 72-hour post-transfection and western blotting performed by using anti-CSTF2 antibody (ab72297, Abcam). CSTF2-BirA: CSTF2 in pcDNA3.1 MCS-BirA(R118G)-HA; BirA-CSTF2: CSTF2 in pcDNA3.1 mycBioID; MycBioID: pcDNA3.1 mycBioID Empty Vector; BirAHA: pcDNA3.1 MCS-BirA(R118G)-HA.

The results shown in Figure 3.3 and Figure 3.5 indicated that some of the fusion CSTF2 protein remained in the cytoplasmic fraction of cell lysates. This situation can be explained with the structure of CSTF2 protein. CSTF2 protein does not harbor a nuclear localization signal which targets it for nuclear localization. Nuclear transport of CSTF2 protein is achieved by interaction of CSTF2 protein with CSTF77 (aka, CSTF3). The two plasmids (pcDNA3.1 mycBioID, pcDNA3.1 and MCS-BirA(R118G)-HA) are mammalian expression plasmids harboring cytomegalovirus (CMV) promoter for overexpression of recombinant proteins. In our case, CSTF2 protein is over-produced but CSTF77 protein is not. In this case, it is possible that the endogenous CSTF77 protein can carry some of the over-produced CSTF2 into the nucleus, and it is possible for the rest to remain in the cytosol.

Then, the biotin addition activity of CSTF2-fused BirA\* protein was evaluated to determine if the fused protein was capable of biotinylating potential interacting partners of CSTF2 or not. For this purpose, HEK293 cells were transfected with recombinant vectors and divided into two groups as: “+ Biotin” and “- Biotin” groups. 24 hours after transfection, existing media of the “+Biotin” group replaced with fresh media including 50  $\mu$ M Biotin and 1mM ATP. Media of “-Biotin” group is replaced with fresh media including only 1mM ATP. 18 hours after addition of

biotin these cells are harvested and used for nuclear and cytoplasmic lysate isolation with a manual protocol described elsewhere (Akman *et al.*, 2012). Isolated nuclear lysates used for western blot with anti-Biotin (ab52494, Abcam) antibody to evaluate if the biotinylation process worked or not (Figure 3.6).

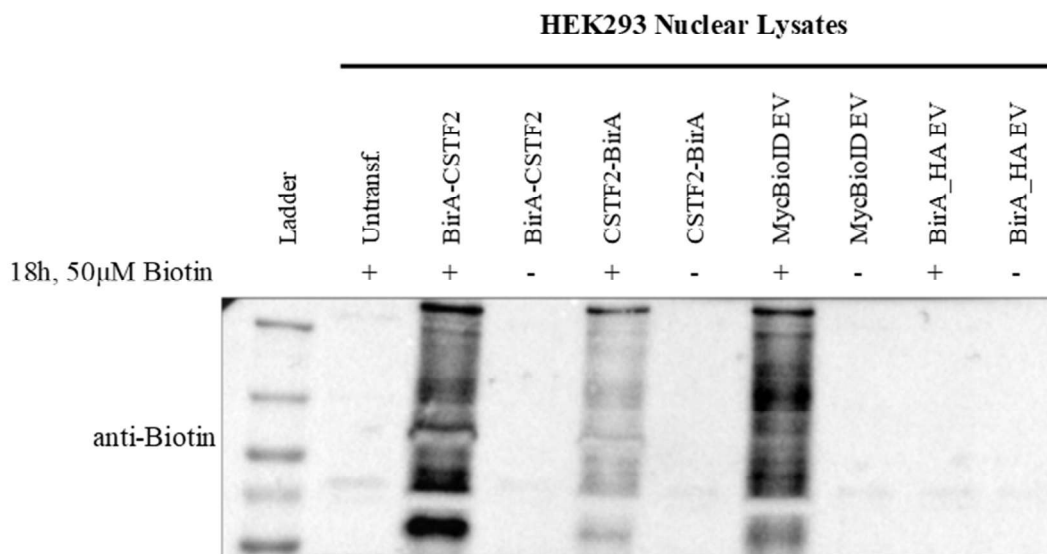


Figure 3.6. Biotinylation of potentially interacting partners of CSTF2. HEK293 cells were transfected with CSTF2 and BirA\* fusion constructs. After 24 hours, transfected cells incubated with 1 mM ATP in the presence/absence of 50 µM Biotin for 18 hours and western blotting performed by using anti-Biotin antibody (ab53494, Abcam). CSTF2-BirA\*: CSTF2 in pcDNA3.1 MCS-BirA(R118G)-HA; BirA\*-CSTF2: CSTF2 in pcDNA3.1 mycBioID; MycBioID: pcDNA3.1 mycBioID Empty Vector; BirAHA: pcDNA3.1 MCS-BirA(R118G)-HA.

According to the result, the promiscuous biotin ligase “BirA” protein is capable of covalently adding biotin molecules to the proteins depending on proximity even it is produced in fusion with CSTF2 protein. Especially myc\_CSTF2\_pcDNA3.1 plasmid (BirA\*\_CSTF2 fusion) showed a higher efficiency of biotinylation in comparison to CSTF2\_BirA\*\_HA plasmid. On the other hand, empty vector control of this plasmid pcDNA3.1 mycBioID (designated as MycBioID EV) resulted an unexpected and strong biotinylation of nuclear proteins. BirA\* protein coded by

pcDNA 3.1 mycBioID empty vector does not have a nuclear localization signal and the produced protein was not expected to localize in the nucleus and biotinylate nuclear proteins. This situation could be a result of leakage of biotinylated proteins from cytoplasm to nucleus during fractionated lysate isolation. This situation suggested that NE-PER Nuclear and Cytoplasmic Extraction Reagents (78835, Thermo Scientific) provided a better fractionation with lower amounts of leakage of cytoplasmic proteins and manual extraction protocol (Akman *et al.*, 2012) needed more optimization for decreasing the cross-contamination between cytoplasmic and nuclear proteins.

Despite the leaky isolation, streptavidin pull-down efficiency of the biotinylated proteins was tested. For this purpose, lysates isolated from “+Biotin” and “-Biotin” groups of MCF7 cells transfected with BirA\*-CSTF2 fusion plasmids were used and pull-down was performed as described (Dong *et al.*, 2015). 200 µg of lysates were diluted in lysis buffer by completing the volume up to 400 µl and incubated with 35 µl (1ml/6mg lysate, as indicated in the manual of the product) of Streptavidin Sepharose High Performance (GE Life Sciences, 17511301) for 16 hours with gentle agitation. Then streptavidin beads were collected by centrifugation (7000 × g) and washed 4 times with lysis buffer at 4 °C. Bead-bound proteins are eluted via incubation at 95 °C in 2x Laemmli sample buffer (Dong *et al.*, 2015). Eluted proteins were resolved in 10% SDS-PAGE gel for western blot analysis (Figure 3.7 and 3.8).

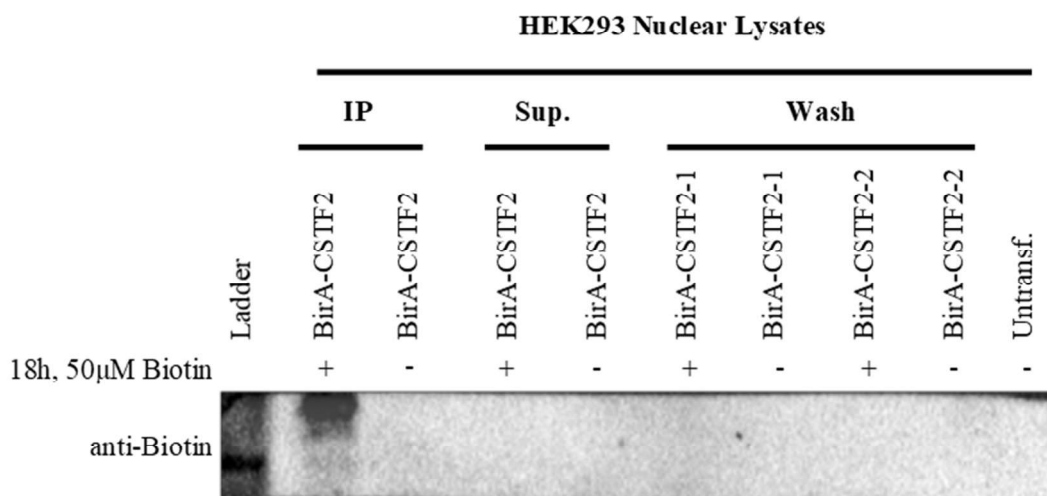


Figure 3.7. Streptavidin affinity capture of biotinylated proteins. HEK293 cells were transfected with CSTF2 and BirA fusion constructs. After 24 hours, transfected cells incubated with 1 mM ATP in the presence/absence of 50 μM Biotin for 18 hours. Isolated lysates used for affinity capture with streptavidin beads. Eluted proteins used for western blot analysis with anti-Biotin antibody (ab53494, Abcam). BirA-CSTF2: CSTF2 in pcDNA3.1 mycBioID.

This blot showed that all of the biotinylated proteins were probably captured and eluted efficiently since there wasn't any biotinylated protein in "supernatant" and "wash" fractions of "+Biotin" group. Also, in "-Biotin" group there were no biotinylated protein as expected.

To test whether biotinylation and streptavidin pull-down was successful, CSTF2 and CFIm25, 25 kDA subunit of CFIm complex of polyadenylation machinery, antibodies were used for immunblotting (Figure 3.8).

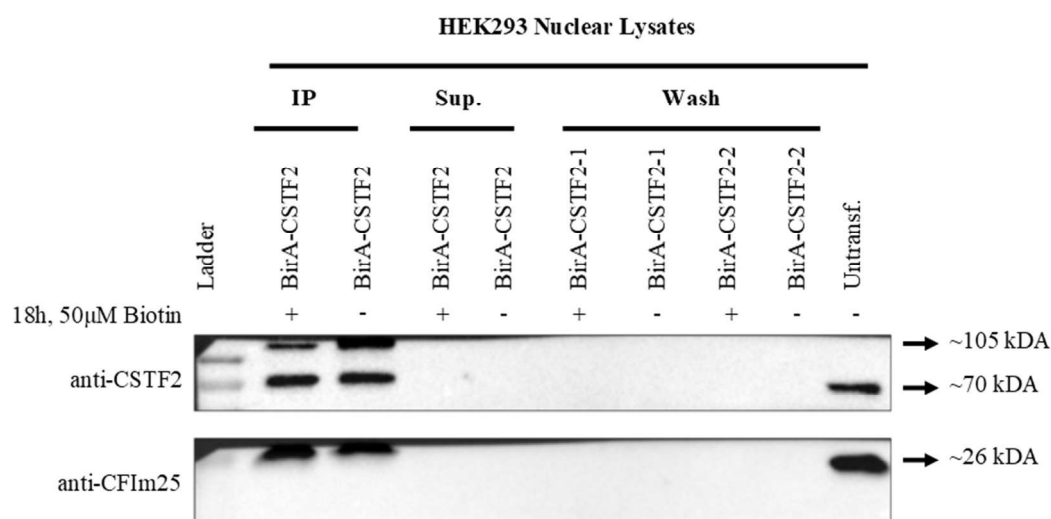


Figure 3.8. Streptavidin affinity capture of biotinylated proteins. HEK293 cells were transfected with CSTF2 and BirA\* fusion constructs. After 24 hours, transfected cells incubated with 1 mM ATP in the presence/absence of 50 µM Biotin for 18 hours. Isolated lysates used for affinity capture with streptavidin beads. Eluted proteins used for western blot analysis with anti-CSTF2 (ab72297, Abcam) and anti-CFIm25 (10322-1-AP, Proteintech).

However, as results clearly showed; conditions for streptavidin affinity capture was not suitable, because we can also observe CSTF2 and CFIm25 bands on “-Biotin” group which means unbiotinylated proteins also pulled-down under these conditions.

This situation suggested that centrifugation speed (7000 x g) indicated in the reference article (Dong *et al.*, 2015) may be too high for collecting streptavidin beads (high speed may cause settling of unbound proteins) and decided to lower it. The experiment repeated with exactly under same conditions except the centrifugation speed (Figure 3.9). For the repeated pull-down experiment, I set the centrifugation speed to 2500 x g, as indicated manual of a different manufacturer (20347, Thermo Fischer Scientific).

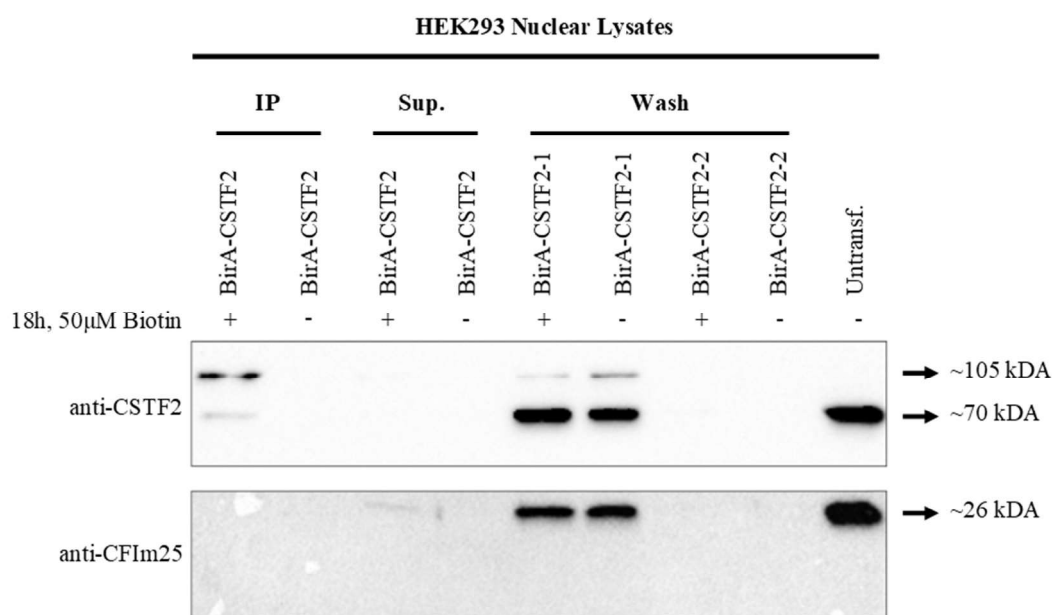


Figure 3.9. Streptavidin affinity capture of biotinylated proteins. HEK293 cells were transfected with CSTF2 and BirA fusion constructs. After 24 hours, transfected cells incubated with 1 mM ATP in the presence/absence of 50 µM Biotin for 18 hours. Isolated lysates used for affinity capture with streptavidin beads. Eluted proteins used for western blot analysis with anti-CSTF2 (ab72297, Abcam) and anti-CFIm25 (10322-1-AP, Proteintech).

According to the results, a significant portion of BirA\*-CSTF2 fusion protein, along with some endogenous CSTF2 was captured via streptavidin pull-down. On the other hand, CFIm25, a component of polyadenylation machinery was not pulled-down with CSTF2 protein. This may be due to inefficient streptavidin-capture resulting from insufficient amount of bead usage. To evaluate this possibility membrane used for visualization of CSTF2 protein was stripped and immunoblotted with anti-Biotin antibody (Ab53494, Abcam) (Figure 3.10).

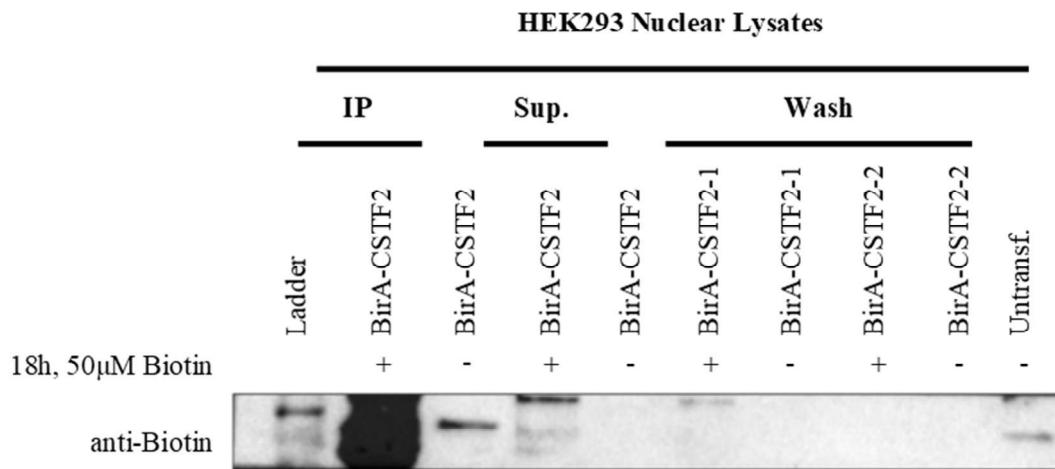


Figure 3.10. Streptavidin affinity capture of biotinylated proteins. HEK293 cells were transfected with CSTF2 and BirA fusion constructs. After 24 hours, transfected cells incubated with 1 mM ATP in the presence/absence of 50 µM Biotin for 18 hours. Isolated lysates used for affinity capture with streptavidin beads. Eluted proteins used for western blot analysis with anti-Biotin (Ab53494, Abcam).

The results indicated that for “+Biotin” group, most of the biotinylated proteins were pulled-down during affinity capture, and only a driblet of biotinylated proteins was remained in “supernatant” fraction. For “-Biotin” group and “untransfected” control group some bands were observed, which may be resulting from non-specific binding of the antibody.

When evaluated together these results (Figure 3.9 and 3.10) indicated three important points: 1) decreased centrifugation speed (2500 x g) was high enough for settling of beads and bound proteins. 2) for “+Biotin” group, most of the BirA\*-CSTF2 fusion proteins were precipitated during affinity-capture but a small fraction of fusion proteins was lost during first washing step and visualizing the same blot with anti-Biotin antibody suggested the same conclusion (Figure 3.10). This could be due to insufficient amount of streptavidin beads so bead/lysate ratio were decided to be increased to 1ml/4mg for later affinity-capture experiments. The last and most important point was that CFIm25, a protein which is a component of polyadenylation

machinery, was not biotinylated and pulled-down with CSTF2 protein. This situation suggested that BioID approach may not be efficient enough for our case to detect dynamic and transient interactions of CSTF2 and led the idea to utilize a more rapid method for identification of potential interacting partners of CSTF2.

### 3.2.2 Expression of TurboID-Fused CSTF2 Protein in HEK293 and MCF7 Cells

As indicated before, BioID approach requires 18 hour of incubation time in the presence of 50  $\mu$ M D-Biotin. This long duration is required because of the relatively low catalytic activity of promiscuous biotin protein ligase BirA (R118G), but dynamic and transient interactions of CSTF2 with its interacting partners were not detected due to this low catalytic activity. A mutant of BirA called TurboID is capable of adding biotin molecules depending on proximity in a much shorter duration (i.e. 10 min) (Brannon *et al.*, 2018). Hence, to capture transient interactions of CSTF2 protein, fusion CSTF2 and TurboID was generated in pcDNA3.1 (-) (Invitrogen, V79520), a mammalian expression vector. Figure 3.11 shows the schematic representation of construct generated for production of CSTF2 C-terminally fused with TurboID.

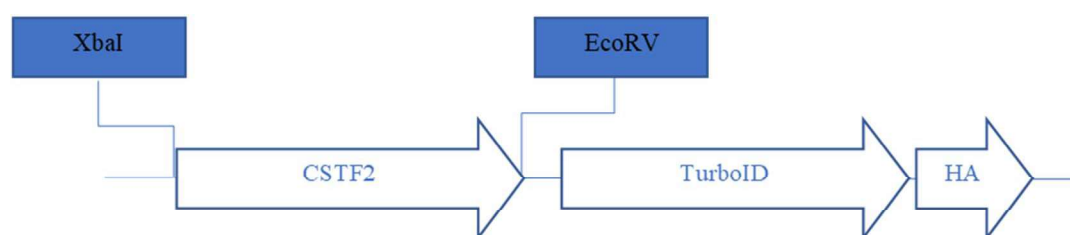


Figure 3.11. Cloning CSTF2 coding sequence and *TurboID* sequence into pcDNA3.1 (-) (Invitrogen, V79520) vector for creating *CSTF2* C-terminal fusion with *TurboID*.

After cloning the coding sequences of both genes into pcDNA 3.1 (-) vector and verification via sequence analysis, we used this construct for transfection of HEK293



cells. Transfected HEK293 cells were collected after 24 and 48 hours of transfection and used for nuclear and cytoplasmic protein extraction following the protocol described (Akman *et al.*, 2012). Isolated lysates resolved in 10% SDS-PAGE gel for western blot analysis (Figure 3.12).

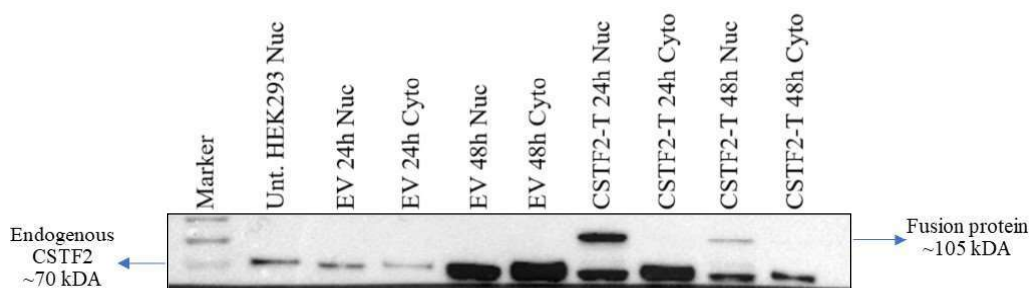


Figure 3.12. Transfection of HEK293 cells with CSTF2-TurboID fusion construct. Lysates isolated from transfected HEK293 cells 24- and 48-hour post-transfection and western blotting performed by using anti-CSTF2 antibody (ab72297, Abcam). EV: TurboID in pcDNA 3.1 (-); CSTF2-T: CSTF2-TurboID in pcDNA 3.1 (-).

The results showed that the CSTF2-TurboID fusion protein amount was quite low in nuclear lysate isolated from 48-hour transfection samples of HEK293 cells in comparison to those of 24-hour transfection samples. Since it is known that the extraction method followed for nuclear/cytoplasmic protein isolation (Akman *et al.*, 2012) was not effective enough for separation nuclear and cytoplasmic proteins, it was thought that this difference may be a result of leaky isolation.

Despite the leaky protein isolation problem, we wanted to test whether CSTF2-fused TurboID protein was capable of biotinylating potential interacting partners of CSTF2. For this purpose, HEK293 cells were transfected with recombinant vectors and divided into two groups as “+ Biotin” and “- Biotin” groups. 24 hours after transfection, existing media of the “+ Biotin” group was replaced with fresh media including 50  $\mu$ M Biotin and 1mM ATP. Media of “- Biotin” group is replaced with fresh media including only 1mM ATP. 30-, 60- and 360-minutes after addition of biotin these cells were harvested and used for manual nuclear and cytoplasmic lysate

isolation according to the protocol described elsewhere (Akman *et al.*, 2012). Isolated nuclear lysates used for western blot with anti-Biotin antibody to evaluate biotinylation levels of the samples (Figure 3.13).

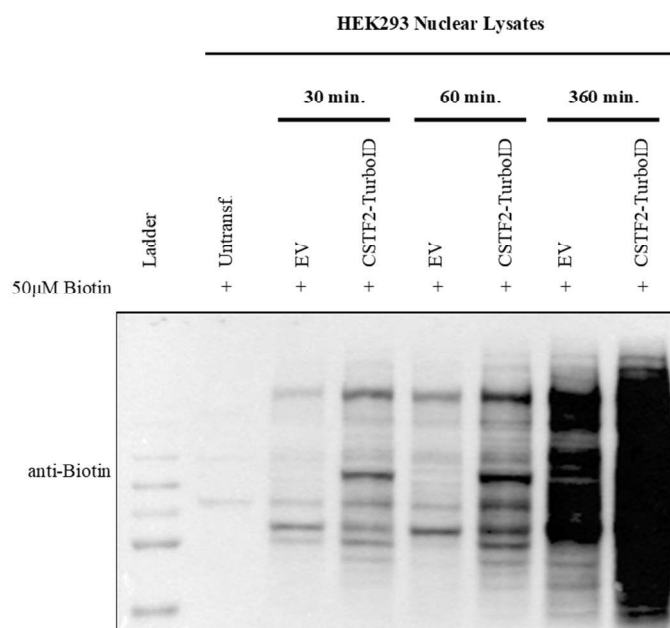


Figure 3.13. Time-dependent biotinylation of potentially interacting partners of CSTF2. HEK293 cells were transfected with CSTF2-TurboID fusion constructs. After 24 hours, transfected cells incubated with 1 mM ATP in the presence/absence of 50 µM Biotin for 30-, 60- and 360-minutes and western blotting performed by using anti-Biotin antibody (ab53494, Abcam).

Results showed that the CSTF2-fused TurboID protein can indeed efficiently add biotin within minutes as opposed to BirA. However, we detected biotinylated protein in nuclear lysates isolated from empty vector (only TurboID in pcDNA 3.1 (-)), suggesting a continuing issue of leakage during nuclear/cytoplasmic lysate isolation.

Then, streptavidin affinity capture was performed to separate biotinylated proteins from others. Lysates isolated from “+Biotin” and “-Biotin” groups of transfection experiment used for this purpose. 200 µg of lysates were diluted in lysis buffer by completing the volume up to 400 µl and incubated with 50 µl of Streptavidin Sepharose High Performance (GE Life Sciences, 17511301) for 16 hours with gentle

agitation. Then streptavidin beads were collected by centrifugation (2500 rpm) and washed 4 times with lysis buffer. Bead-bound proteins were eluted via incubation at 100 °C in 2x Laemmli sample buffer (Dong *et al.*, 2015). Eluted protein used for western blot analysis with anti-Biotin antibody (ab53494, Abcam) to evaluate the efficiency of streptavidin pull-down (Figure 3.14).

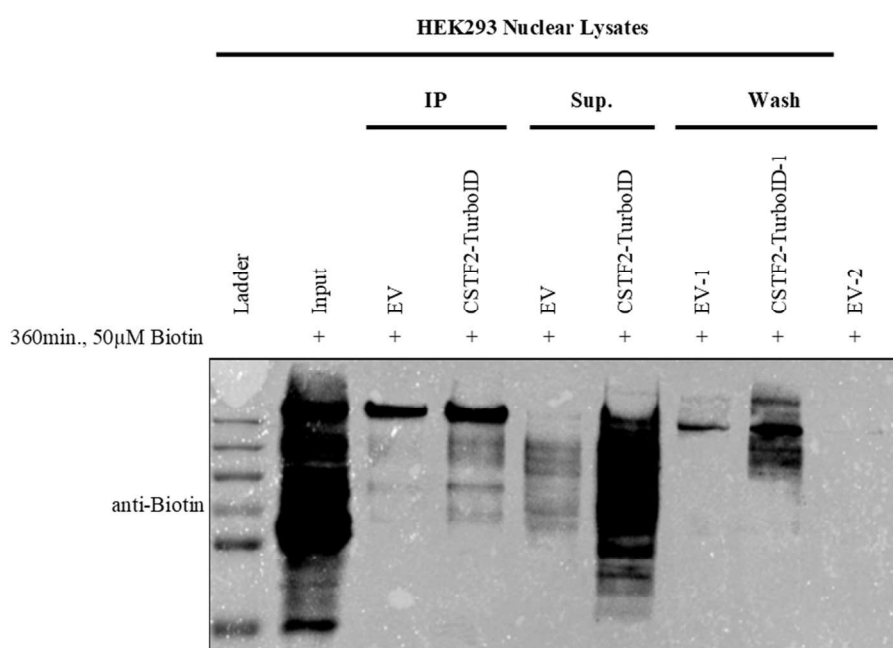


Figure 3.14. Streptavidin affinity capture of biotinylated proteins. HEK293 cells were transfected with CSTF2-TurboID fusion constructs. After 24 hours, transfected cells incubated with 1 mM ATP in the presence/absence of 50 µM Biotin for 360-minutes. Isolated lysates used for affinity capture with streptavidin beads. Eluted proteins used for western blot analysis with anti-Biotin antibody (ab53494, Abcam).

This result showed that increasing bead amount 1.5-fold (in comparison to BioID pull-down) was not enough for capturing all of the biotinylated proteins since most of these proteins remained in supernatant fraction. We decided to repeat this experiment by even increasing the bead amount by 2-folds (adjusted according to the lysate amount) and trying different buffer compositions. For this set of experiment; 50 µg of lysates diluted in lysis buffer (Buffer B), SSHP Buffer, 1xPBS or 0,1%

TBST by completing the volume to 400  $\mu$ l and incubated with 25  $\mu$ l of Streptavidin Sepharose High Performance (1ml/2mg lysate) (GE Life Sciences, 17511301) for 16 hours with gentle agitation. Then streptavidin beads were collected by centrifugation (2500 rpm) and washed twice with Wash Buffer I [2 % Sodium Dodecyl Sulfate (SDS)], once with Wash Buffer II (2% deoxycholate, 1% Triton-X, 50 mM NaCl, 50 mM Hepes, pH=7.5, 1 mM EDTA), once with Wash Buffer III (0.5% NP-40, 0.5% deoxycholate, 1% Triton-X, 500 mM NaCl, 1 mM EDTA, 10mM Tris, pH=8.1), and once with Wash Buffer IV (50mM Tris, pH=7.4, 50 mM NaCl). (Figure 3.15).

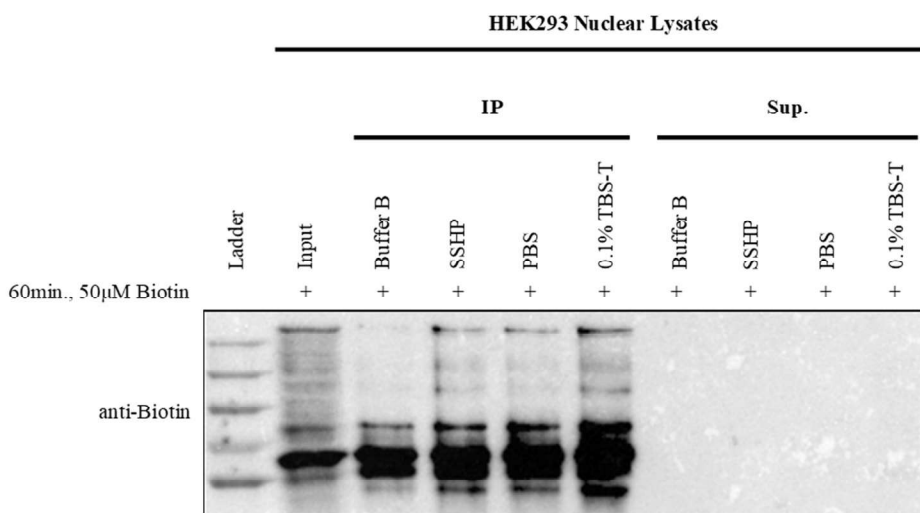


Figure 3.15. Effects of buffer composition on streptavidin affinity capture. HEK293 cells were transfected with CSTF2-TurboID fusion constructs. After 24 hours, transfected cells incubated with 1 mM ATP in the presence/absence of 50  $\mu$ M Biotin for 60-minutes. Isolated lysates used for affinity capture with streptavidin beads. Eluted proteins used for western blot analysis with anti-Biotin antibody (ab53494, Abcam).

This result showed that changing the buffer compositions and increasing bead amounts was a promising strategy for optimization of streptavidin affinity capture. Using especially 0.1% TBST instead of lysis buffer (Buffer B) provided an important improvement for capturing most of the biotinylated proteins.

Another important point was localization of produced CSTF2-TurboID protein. So, we also checked nuclear localizations of CSTF2-TurboID fusion protein along with

those of TurboID protein coded by control empty vector and endogenous CSTF2 protein. For this purpose, HEK293 cells seeded on coverslips were transfected with CSTF2\_TurboID\_HA or TurboID\_HA empty vector. Also, untransfected MCF7 cells were used for visualization of endogenous CSTF2 protein. 24-hours and 48-hours after the transfection coverslips were used for visualization under microscope (Figure 3.16).

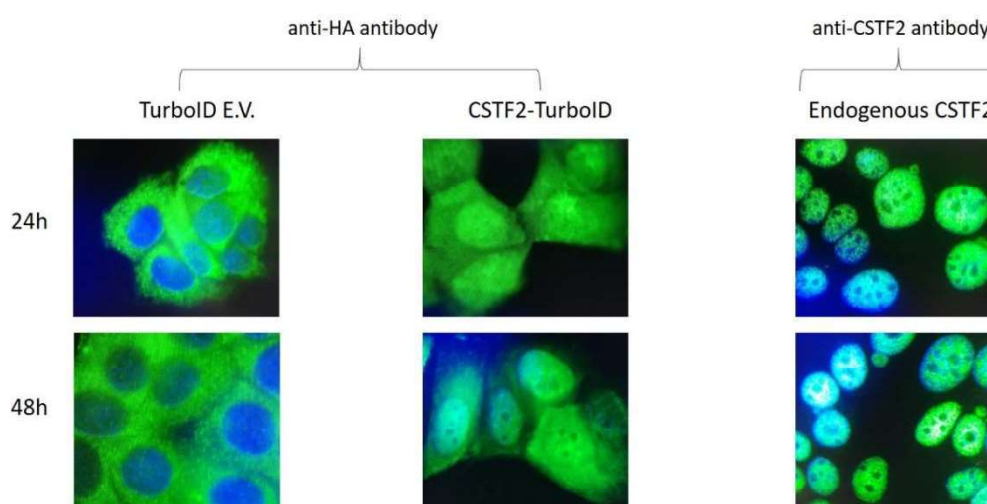


Figure 3.16. Cellular localization of fusion CSTF2-TurboID proteins. HEK293 cells transfected with CSTF2\_TurboID\_HA or TurboID\_HA plasmids were used for investigating nuclear localizations of the encoded proteins. Untransfected HEK293 cells used as control group for visualization of endogenous CSTF2 protein. Anti-HA primary antibody (ab9119, Abcam) and anti-CSTF2 antibody (ab72297, Abcam) used as primary antibodies. Alexa Fluor 488-conjugated goat anti-rabbit IgG (ab150077, Abcam) used as secondary antibody.

Immunocytochemistry results indicated that TurboID protein coded by control vector (designated as TurboID E.V.) was stacking in the cytosol as expected since the encoded protein did not have a nuclear localization signal peptide. Most of the CSTF2-TurboID fusion protein coded by CSTF2\_TurboID\_HA plasmid (designated as CSTF<sup>2</sup>-TurboID) remained in the cytoplasm and a small fraction of the protein was translocated to nucleus as expected since interacting partner of CSTF2, CSTF77, which was responsible for nuclear localization of CSTF2, was not over-expressed.

The results also showed that endogenous CSTF2 protein is localized in the nucleus of the cells.

Then we performed the same experiment with MCF7 breast cancer cells. For this purpose, cells were transfected with CSTF2\_TurboID\_HA or TurboID\_HA empty vector. 24 hours after transfection cells were divided into two sub-groups as: “+Biotin” and “-Biotin” and existing media of the “+Biotin” group replaced with fresh media including 50  $\mu$ M Biotin and 1mM ATP. Media of “-Biotin” group is replaced with fresh media including only 1mM ATP. 3 hours after addition of biotin these cells are harvested and used for nuclear and cytoplasmic lysate isolation according to the protocol described (Akman *et al.*, 2012). 200  $\mu$ g of lysates were diluted in 0.1 % TBST by completing the volume up to 400  $\mu$ l and incubated with 100  $\mu$ l (1ml/2mg lysate) of Streptavidin Sepharose High Performance (GE Life Sciences, 17511301) for 16 hours with gentle agitation. Then streptavidin beads were collected by centrifugation (2500 rpm) and washed with wash buffers. Bead-bound proteins were eluted via incubation at 100 °C in 2x Laemmli sample buffer (Dong *et al.*, 2015). Eluted protein used for western blot analysis with anti-Biotin antibody (ab53494, Abcam) to evaluate the efficiency of streptavidin pull-down (Figure 3.17).

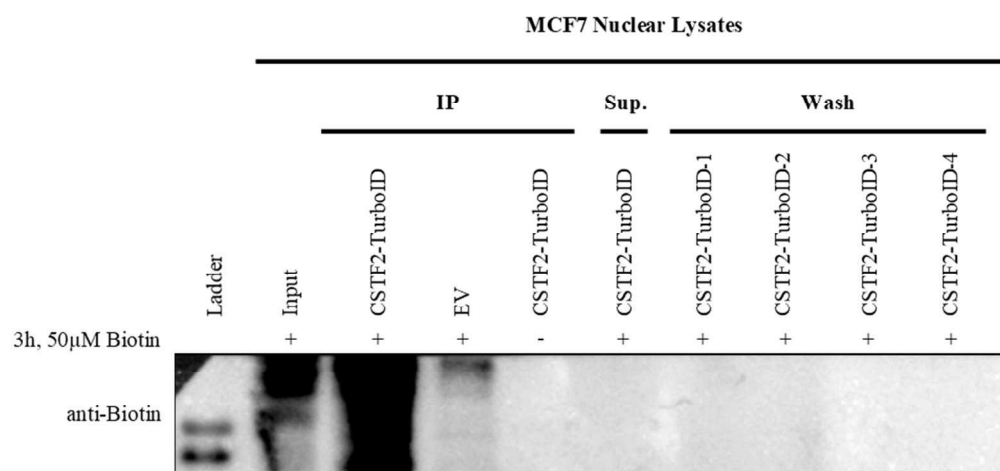


Figure 3.17. Streptavidin affinity capture of potentially interacting partners of CSTF2. MCF7 cells were transfected with CSTF2\_TurboID fusion and TurboID empty vector constructs. After 24 hours, transfected cells incubated with 1 mM ATP in the presence/absence of 50 µM Biotin for 3 hours and western blotting performed by using anti-Biotin antibody (ab53494, Abcam). CSTF2-TurboID + Biotin, transfected with CSTF2\_TurboID\_HA plasmid and incubated with 50 µM Biotin and 1mM ATP; TurboID EV + Biotin, transfected with TurboID\_HA plasmid control plasmid and incubated with 50 µM Biotin and 1mM ATP, CSTF2-TurboID – Biotin, transfected with CSTF2\_TurboID\_HA plasmid and incubated with 1mM ATP only.

As the results indicated, MCF7 cells were transfected with CSTF2\_TurboID\_HA and TurboID empty vector plasmids and the biotinylation process was working very efficiently in 3 hours incubation duration. In addition, all biotinylated proteins of “CSTF2-TurboID +Biotin” group (cells were transfected with CSTF2\_TurboID\_HA plasmid and incubated with 50 µM Biotin, 1mM ATP) were captured in the elution fraction during pull-down experiment and no biotinylated protein was observed in “Supernatant” and “Wash” fractions. Again, there was leakage of biotinylated proteins form cytoplasm to nucleus in “TurboID EV +Biotin” group (cells were transfected with TurboID\_HA control plasmid and incubated with 50 µM Biotin, 1mM ATP). Also, there was no considerable amount of biotinylated and captured proteins in “CTSF2-TurboID –Biotin” group as expected.

Next, we tested whether TurboID fused to CSTF2 was capable of biotinylating itself and captured during affinity capture (Figure 3.18).

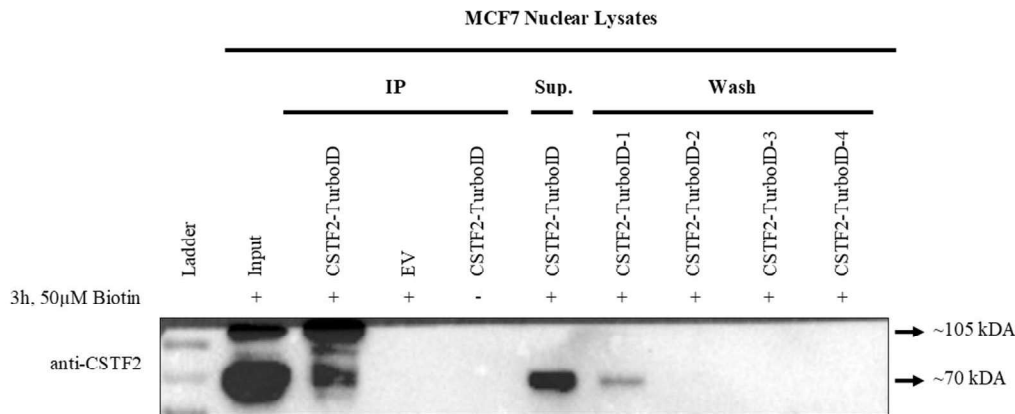


Figure 3.18. Streptavidin affinity capture of biotinylated CSTF2 proteins. MCF7 cells were transfected with CSTF2-TurboID fusion constructs. After 24 hours, transfected cells incubated with 1 mM ATP in the presence/absence of 50 µM Biotin for 3 hours. Isolated lysates used for affinity capture with streptavidin beads. Eluted proteins used for western blot analysis with anti-CSTF2 antibody (ab72297, Abcam). CSTF2-TurboID + Biotin, transfected with CSTF2\_TurboID\_HA plasmid and incubated with 50 µM Biotin and 1mM ATP; TurboID EV + Biotin, transfected with TurboID\_HA plasmid control plasmid and incubated with 50 µM Biotin and 1mM ATP, CSTF2-TurboID – Biotin, transfected with CSTF2\_TurboID\_HA plasmid and incubated with 1mM ATP only.

As the results indicated, for “CSTF2-TurboID +Biotin” sample group nearly all of the CSTF2-TurboID fusion proteins were biotinylated and captured with streptavidin affinity capture. This results also proved that followed pull-down protocol was efficient enough to capture a remarkable portion of the biotinylated proteins since we did not observe any unbiotinylated CSTF2-TurboID fusion in “Supernatant” and “Wash” fractions. Also, it appears that some of the endogenous CSTF2 proteins cannot be captured and remained in “Supernatant” and “Wash” fractions.

Next, we checked the pulldown status of CSTF77 protein, a known interacting partner of CSTF2, to evaluate if the CSTF2-TurboID fusion protein was capable of conserving natural interactions of endogenous CSTF2 during dynamic polyadenylation. For this purpose, the same experimental conditions were followed and at the end of the affinity capture process, eluted proteins were used for western blot analysis (Figure 3.19).



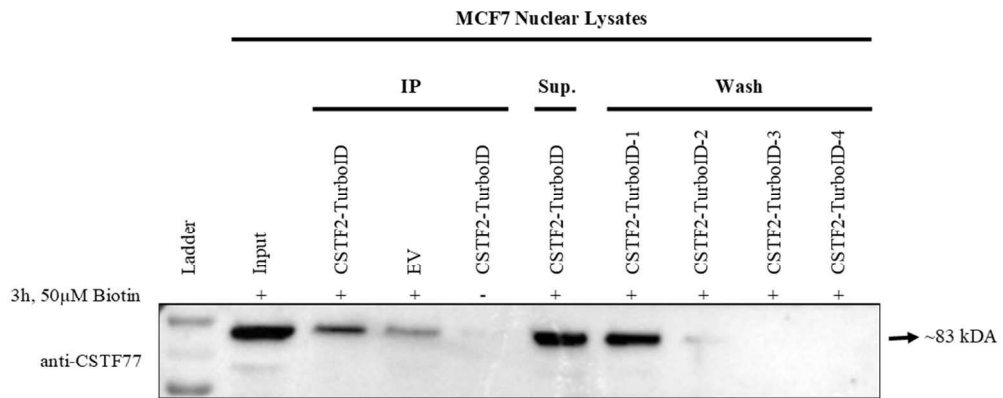


Figure 3.19. Streptavidin affinity capture of biotinylated CSTF77 proteins. MCF7 cells were transfected with CSTF2-TurboID fusion constructs. After 24 hours, transfected cells incubated with 1 mM ATP in the presence/absence of 50 µM Biotin for 3 hours. Isolated lysates used for affinity capture with streptavidin beads. Eluted proteins used for western blot analysis with anti-CSTF77 antibody (sc-376553, Santa Cruz Biotechnology). CSTF2-TurboID + Biotin, transfected with CSTF2\_TurboID\_HA plasmid and incubated with 50 µM Biotin and 1mM ATP; TurboID EV + Biotin, transfected with TurboID\_HA plasmid control plasmid and incubated with 50 µM Biotin and 1mM ATP, CSTF2-TurboID – Biotin, transfected with CSTF2\_TurboID\_HA plasmid and incubated with 1mM ATP only.

As the result indicated that some of the CSTF77 proteins, a known interacting partner of endogenous CSTF2, was biotinylated by CSTF2-TurboID fusion protein and pulled-down by streptavidin affinity capture. On the other hand, a significant portion of the proteins were remained in the “Supernatant” and “Wash” fractions of the samples. It is possible that some of the endogenous CSTF77 proteins mediate nuclear translocation of the CSTF2-TurboID fusion protein, while other CSTF77 proteins are likely to mediate transport of the endogenous CSTf2 protein to the nucleus. In this case, it is possible that the CSTF77 proteins converging with the endogenous CSTF2 proteins could not be biotinylated and could not be captured by downstream streptavidin pull-down. There are some CSTF77 proteins biotinylated by TurboID protein encoded by TurboID\_HA control plasmid but when we compared the band intensities with CSTF2\_TurboID\_HA sample group it is clear that the amount of CSTF77 protein specifically biotinylated by CSTF2-TurboID fusion protein is higher with proteins that were biotinylated by random interactions of TurboID alone.

In light of these findings; since the CSTF2-TurboID fusion protein capable of biotinylating potential interacting partners of CSTF2 and the followed streptavidin-affinity pulldown protocol was finally optimized to capture nearly all biotinylated proteins, we performed these experiments in a larger scale for Mass Spectrometry analysis. Also, to optimize our lysate isolation, we tried a commercial kit, NE-PER Nuclear and Cytoplasmic Extraction Reagents (78835, Thermo Scientific), for the next experiments.

### 3.2.3 Identification of Potential Interacting Partners of CSTF2 with Mass Spectrometry

Before performing large scale experiments for Mass Spectrometry analysis we evaluated the efficiency of commercial NE-PER Nuclear and Cytoplasmic Extraction Reagents (78835, Thermo Scientific) for separation of nuclear and cytoplasmic lysates. For this purpose, nuclear and cytoplasmic lysates isolated from MCF7 cells using NE-PER kit, used for western blot analysis with HDAC1 and alpha-Tubulin antibodies (Figure 3.20).

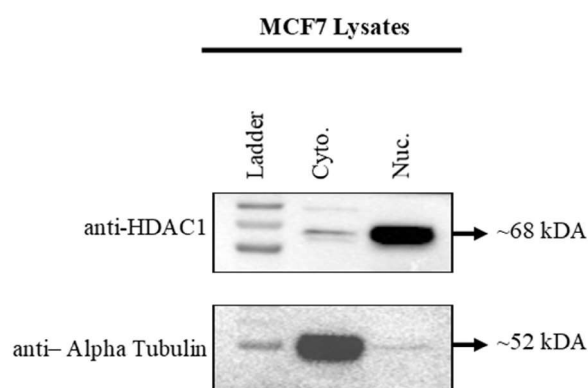


Figure 3.20. Separation of nuclear and cytoplasmic lysates. MCF7 cells were used for nuclear and cytoplasmic lysate isolation using NE-PER Nuclear and Cytoplasmic Extraction Reagents (78835, Thermo Scientific). Isolated lysates used for western blot analysis with anti-HDAC1 antibody (sc-81598, Santa Cruz) and anti-Alpha Tubulin antibody (HRP-66031, Proteintech).

Results showed that the used commercial kit was performing a better separation of cytoplasmic and nuclear lysates in comparison to manual protocol described (Akman *et al.*, 2012). However, a weak band of HDAC1 was observed in cytoplasmic fraction of lysates and a very slight band of alpha-Tubulin was observed in nuclear fraction. This situation was also expected since a 10% of leakage was stated to be normal in the manual of the kit. This issue was noted for the future mass spectrometry experiments as this group of biontynlated proteins would have to be subtracted from the actual sample list of proteins.

Then we performed large scale experiments for mass spectrometry analysis. To perform TurboID assay MCF7 cells were grown until they reach to the optimal confluency. Then, these cells were incubated in phenol-red free DMEM supplemented with 10% CD-FBS, 2% L-Glutamine, 1% Sodium Pyruvate solution and 1% Pen/Strep solution for 72 hours. At 48<sup>th</sup> hour of incubation, cells were divided into 3 subgroups. 2 subgroups were transfected with CSTF2\_TurboID\_HA plasmid and 1 subgroup with TurboID\_HA control plasmid. At the end of the 72-hour incubation duration; one of the CSTF2\_TurboID\_HA transfected subgroups were treated with 100nM E2 along with 50  $\mu$ M Biotin and 1mM ATP. On the other hand, remaining CSTF2\_TurboID\_HA subgroup and TurboID\_HA control vector subgroup were treated with 50  $\mu$ M Biotin, 1mM ATP and EtOH instead of E2. After 3-hour incubation duration cells were collected for RNA and nuclear/cytoplasmic lysate isolation. RNAs isolated from these cells were used for evaluation of efficiency of E2 treatment by comparison of *TFF1* expression between sample groups, while nuclear lysates were used for neutravidin (29204, Thermo Scientific) affinity capture followed by nanoLC-MS/MS analysis. Before sending the samples to Koç University for mass spectrometry analysis, efficiencies of E2 treatment and transient transfection of MCF7 cells were evaluated (Figure 3.21 A, B).

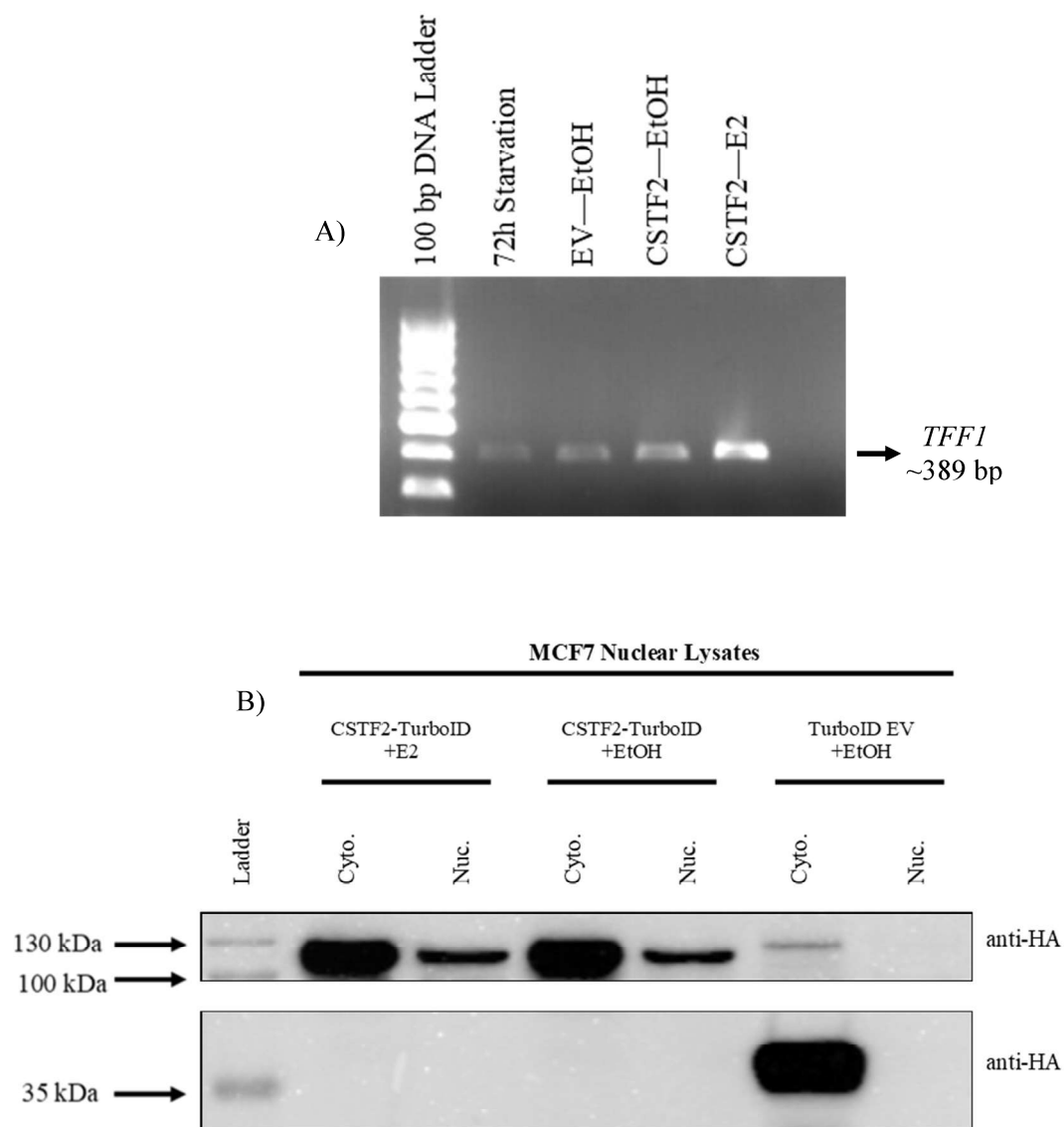


Figure 3.21. Characterization of Mass Spectrometry analysis samples. A) *TFF1* PCR of cDNA samples synthesized from RNAs isolated from transfected and E2 or EtOH treated MCF7 cells B) Expression of fusion CSTF2-TurboID protein encoded by fusion plasmid and TurboID protein encoded by control vector evaluated with western blot analysis using anti-HA antibody (ab9110, Abcam).

The efficiency of E2 treatment was confirmed with the observation of a significant upregulation in *TFF1* expression in E2 treated sample group in comparison to EtOH treated and 72h starvation control groups (Figure 3.21 A). Also, detecting CSTF2-TurboID fusion proteins in nuclear and cytoplasmic lysates indicated that these cells

were efficiently transfected with the plasmid encoding for the fusion protein. Besides, detecting TurboID protein encoded by the control vector in cytoplasmic lysates, but not in the nuclear fraction of the control group indicated that; i) these cells were efficiently transfected with the control plasmid and ii) the encoded protein stuck in the cytosol, which proves the successful fractionation of the lysates with NE-PER kit (Figure 3.21 B).

Then, we performed neutravidin affinity capture with the isolated nuclear proteins and used 10% of the captured proteins to check the pulldown status of biotinylated proteins, CSTF2 and CSTF77 (Figure 3.22).

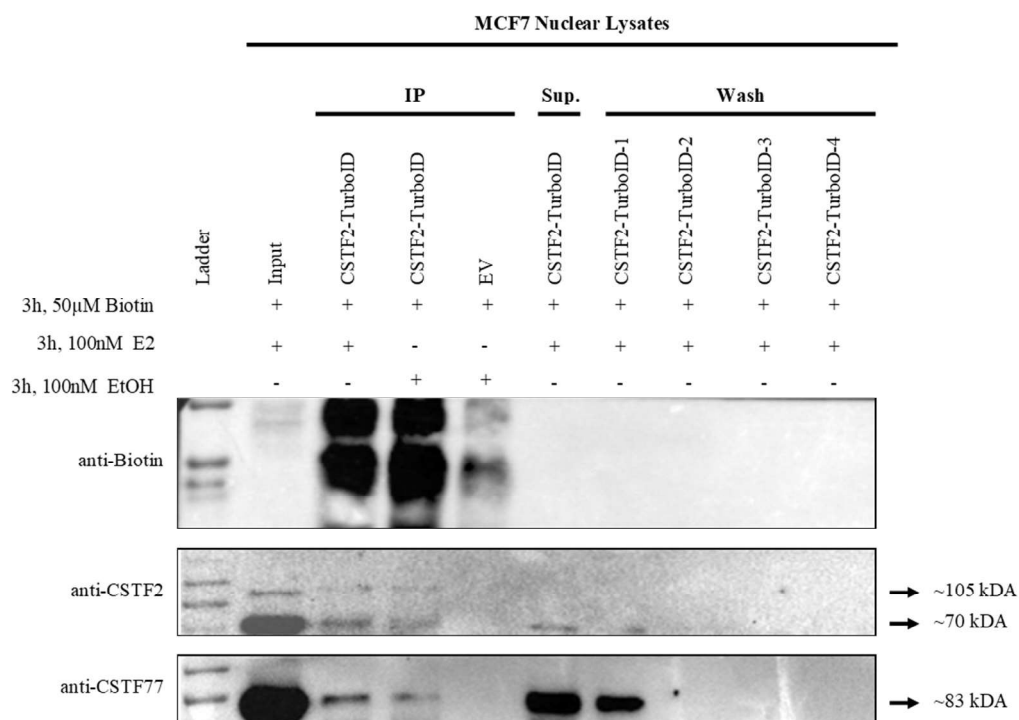


Figure 3.22. Characterization of Mass Spectrometry analysis samples. Pull-down status of biotinylated proteins, CSTF2 and CSTF77 visualized with anti-Biotin (ab-53494), anti-CSTF2 (ab-72297) and anti-CSTF77 (sc-376553) antibodies, respectively.

As the results indicated, nearly all of the biotinylated proteins were captured in the elution fraction during pull-down experiment and no biotinylated protein was

observed in “Supernatant” and “Wash” fractions. When we checked the pull-down status of CSTF2 protein; we detected a CSTF2-TurboID fusion protein band in elution fractions of CSTF2\_TurboID\_HA transfected and E2/EtOH treated sample groups. Since we did not observe any unbiotinylated CSTF2-TurboID fusion protein in “Supernatant” and “Wash” fractions this result suggested that CSTF2-TurboID fusion protein was biotinylated and efficiently captured during pulldown experiment. Also, it appears that some of the endogenous CSTF2 proteins could not be captured and remained in “Supernatant” and “Wash” fractions as observed in the previous experiments. Also, some of the CSTF77 proteins was biotinylated by CSTF2-TurboID fusion protein and pulled-down by neutravidin affinity capture. On the other hand, a significant portion of the proteins were remained in the “Supernatant” and “Wash” fractions of the samples which supports the earlier results. This time we did not detect CSTF77 band in elution fraction of empty vector (TurboID\_HA) control group which further proves the successful fractionation of lysates during protein isolation.

For the mass spectrometry analysis there were two technical replicates for each group of samples. The common proteins obtained from two different technical replicates of “CSTF2\_TurboID\_HA transfected and E2 treated” group were compared with overall protein lists of other two groups (sum of both technical replicates). Based on results there were 16 proteins that were specific for “CSTF2\_TurboID\_HA + E2” group (Figure 3.23 A). Then I decided to utilize STRING network database to evaluate if there is a potential network that links these 16 proteins to known polyadenylation machinery subunits. For this purpose, 6 proteins are listed which were precipitated with CSTF2\_TurboID fusion protein and identified with mass spectrometry (Figure 3.23 B).

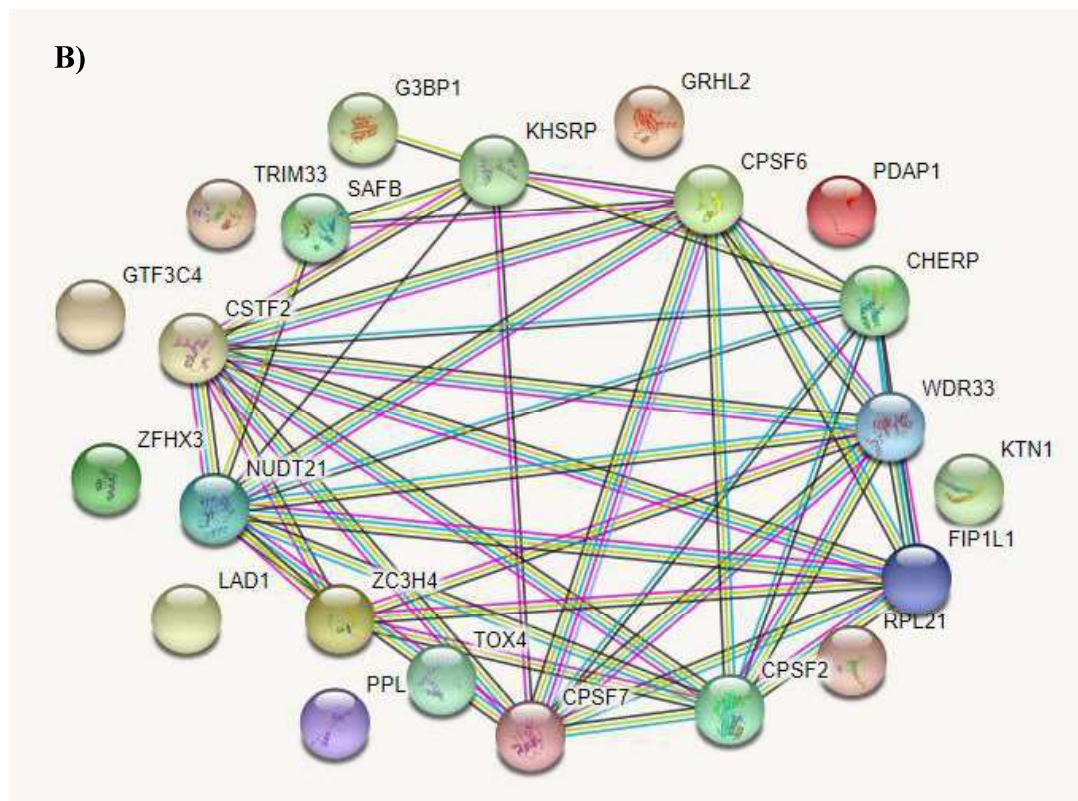
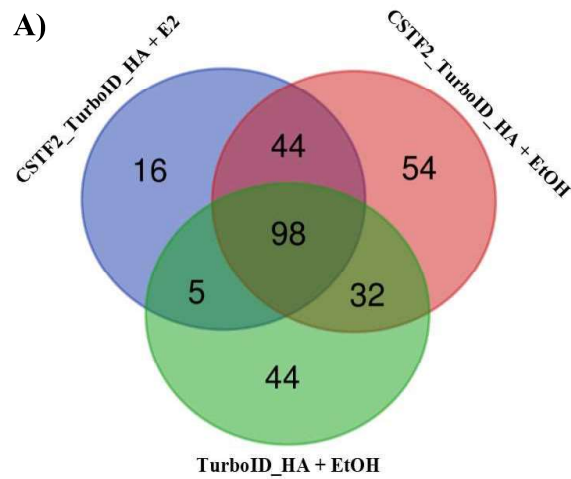


Figure 3.23. Proteins identified by nLC-MS/MS. A) Venndiagram showing the overlap of proteins from “CSTF2\_TurboID\_HA+E2”, “CSTF2\_TurboID\_HA+EtOH” and “TurboID\_HA+EtOH” transfected groups. B) STRING network analysis of 15 CSTF2\_TurboID\_HA+E2-specific proteins along with nLC-MS/MS-detected polyadenylation machinery subunits. Red line-the presence of fusion evidence; green line-neighborhood evidence; Blue line-cooccurrence evidence; Purple line-experimental evidence; Yellow line-textmining evidence; Light blue line-database evidence; Black line-coexpression evidence.

As a result of STRING network, we identified two candidate proteins as KHSRP (KH-Type Splicing Regulatory Protein) and SAFB (Scaffold attachment factor B) for evaluating their potential interactions between CSTF2 and also their effects of on polyadenylation machinery with the presence of E2. Future experiments will have to be performed to verify the interactions with these candidates and to test whether these interactions do affect the polyadenylation choices.



## CHAPTER 4

### CONCLUSION AND FUTURE DIRECTIONS

Alternative polyadenylation (APA), selection of alternative poly (A) signals on a single mRNA, contributes the transcript diversity via production of isoforms exhibiting variations either in their 3'UTR region and/or in their coding sequence, thus regulating the stability, function, translation efficiency and localization of the target RNAs. The exact mechanism behind regulation of APA in normal cells and cancer cells is not fully uncovered, but there are some mechanistic explanations. Among the components of polyadenylation machinery, CSTF2, a subunit of CSTF complex which directly interacts with the downstream UG-rich sequence elements, is considered as the strongest candidate to be an APA regulator. For example; depletion of this protein results in preferential usage of the distal poly (A) signals in HeLa cells (Yao *et al.*, 2012). There are reports to suggest altered expression of poly (A) machinery proteins to change poly (A) site usage decisions. However, changes in poly (A) machinery proteins may not always explain context-dependent and specific poly (A) site usage cases. Therefore, we aimed to identify novel interactors/regulators of poly (A) machinery proteins. To do this, we focused on potential interacting partners of CSTF2 to better understand the regulation of APA in breast cancer cells.

To identify interacting partners of CSTF2 we performed proximity-dependent biotin identification (BioID) and TurboID methods. Promiscuous biotin ligase BirA\* used in BioID, has a lower catalytic activity in comparison to that of the mutant biotin ligase used in TurboID. As a consequence, the biotinylation duration is very long in BioID (18 hours) when compared to TurboID (i.e 10 minutes). Our first biotin-labeling trials were performed with BioID. BirA\* fused to CSTF2 was capable of biotinylating some nuclear proteins during 18-hour incubation period. However,

CFIm25, a subunit of polyadenylation machinery, was not among the biotinylated proteins. This led to the idea that low catalytic activity of the BirA\* may not be enough for labeling potential interacting partners CSTF2, if the interaction was in a transient and rapid nature. Then we switched to a more rapid labeling method, TurboID, for detecting these transient interactions. Following the optimization steps for streptavidin affinity capture mass spectrometry analysis, we observed that both CFIm25 (based on Mass Spectrometry data) and CSTF77 (western blot proved) were biotinylated in TurboID, with 3-hour incubation period. Besides these known components of polyadenylation machinery, we identified 2 candidate proteins as KHSRP and SAFB that seem to be E2-stimulation specific interacting partners of CSTF2. In the future studies we will be focusing on these proteins for verification of their interaction with CSTF2 in the presence/absence of E2-stimulation. Also RNA interference studies may be performed for evaluation of these potential interactions on poly (A) site selections.

## REFERENCES

- Akman, B. H., Can, T., & Erson-Bensan, A. E. (2012). Estrogen-induced upregulation and 3'-UTR shortening of CDC6. *Nucleic Acids Research*, 40(21), 10679–10688. doi: 10.1093/nar/gks855
- Akman, H. B., & Erson-Bensan, A. E. (2014). Alternative Polyadenylation and Its Impact on Cellular Processes. *MicroRNA*, 3(1), 2–9. doi: 10.2174/2211536602666131210001152
- Akman, H. B., Oyken, M., Tuncer, T., Can, T., & Erson-Bensan, A. E. (2015). 3'UTR shortening and EGF signaling: implications for breast cancer. *Human Molecular Genetics*, ddv391. doi: 10.1093/hmg/ddv391
- 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. doi: 10.1016/j.cell.2008.05.045
- Beelman, C. A., & Parker, R. (1995). Degradation of mRNA in eukaryotes. *Cell*, 81(2), 179–183. doi: 10.1016/0092-8674(95)90326-7
- Berkovits, B. D., & Mayr, C. (2015). Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature*, 522(7556), 363–367. doi: 10.1038/nature14321
- Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S. M., Singer, R. H., & Long, R. M. (1998). Localization of ASH1 mRNA Particles in Living Yeast. *Molecular Cell*, 2(4), 437–445. doi: 10.1016/s1097-2765(00)80143-4
- Birse, C. E., Minvielle-Sebastia L., Lee B. A., Keller W., Proudfoot N. J. (1998). Coupling Termination of Transcription to Messenger RNA Maturation in Yeast. *Science*, 280(5361), 298–301. doi: 10.1126/science.280.5361.298

- Branon, T. C., Bosch, J. A., Sanchez, A. D., Udeshi, N. D., Svinkina, T., Carr, S. A., ... Ting, A. Y. (2018). Efficient proximity labeling in living cells and organisms with TurboID. *Nature Biotechnology*, 36(9), 880–887. doi: 10.1038/nbt.4201
- Brown, K. M., & Gilmartin, G. M. (2003). A Mechanism for the Regulation of Pre-mRNA 3' Processing by Human Cleavage Factor Im. *Molecular Cell*, 12(6), 1467–1476. doi: 10.1016/s1097-2765(03)00453-2
- Brumbaugh, J., Stefano, B. D., Wang, X., Borkent, M., Forouzmand, E., Clowers, K. J., ... Hochedlinger, K. (2018). Nudt21 Controls Cell Fate by Connecting Alternative Polyadenylation to Chromatin Signaling. *Cell*, 172(3), 629–631. doi: 10.1016/j.cell.2017.12.035
- Buoncristiani, M. R., Howard, P. K., & Otsuka, A. J. (1986). DNA-binding and enzymatic domains of the bifunctional biotin operon repressor (BirA) of *Escherichia*. *Gene*, 44(2-3), 255–261. doi: 10.1016/0378-1119(86)90189-7
- 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
- Cevher, M. A., & Kleiman, F. E. (2010). Connections between 3'-end processing and DNA damage response. *Wiley Interdisciplinary Reviews: RNA*, 1(1), 193–199. doi: 10.1002/wrna.20
- Cevheroğlu, O., Kumaş, G., Hauser, M., Becker, J. M., & Son, Ç. D. (2017). The yeast Ste2p G protein-coupled receptor dimerizes on the cell plasma membrane. *Biochimica Et Biophysica Acta (BBA) - Biomembranes*, 1859(5), 698–711. doi: 10.1016/j.bbamem.2017.01.008
- Chang, J.-W., Zhang, W., Yeh, H.-S., Jong, E. P. D., Jun, S., Kim, K.-H., ... Yong, J. (2015). mRNA 3'-UTR shortening is a molecular signature of mTORC1 activation. *Nature Communications*, 6(1). doi: 10.1038/ncomms8218

- 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 & Development*, 28(21), 2370–2380. doi: 10.1101/gad.250993.114
- Chen, F., Macdonald, C. C., & Wilusf, J. (1995). Cleavage site determinants in the mammalian polyadenylation signal. *Nucleic Acids Research*, 23(14), 2614–2620. doi: 10.1093/nar/23.14.2614
- Chen, L.-L., & Carmichael, G. G. (2009). Altered Nuclear Retention of mRNAs Containing Inverted Repeats in Human Embryonic Stem Cells: Functional Role of a Nuclear Noncoding RNA. *Molecular Cell*, 35(4), 467–478. doi: 10.1016/j.molcel.2009.06.027
- Colgan, D. F., & Manley, J. L. (1997). Mechanism and regulation of mRNA polyadenylation. *Genes & Development*, 11(21), 2755–2766. doi: 10.1101/gad.11.21.2755
- Derti, A., Garrett-Engle, 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. doi: 10.1101/gr.132563.111
- Di Giammartino, D. C., Nishida, K., & Manley, J. L. (2011). Mechanisms and Consequences of Alternative Polyadenylation. *Molecular Cell*, 43(6), 853–866. doi: 10.1016/j.molcel.2011.08.017
- Dong, A., Wodziak, D., & Lowe, A. W. (2015). Epidermal Growth Factor Receptor (EGFR) Signaling Requires a Specific Endoplasmic Reticulum Thioredoxin for the Post-translational Control of Receptor Presentation to the Cell Surface. *Journal of Biological Chemistry*, 290(13), 8016–8027. doi: 10.1074/jbc.m114.623207
- Elkon, R., Ugalde, A. P., & Agami, R. (2013). Alternative cleavage and polyadenylation: extent, regulation and function. *Nature Reviews Genetics*, 14(7), 496–506. doi: 10.1038/nrg3482

Ephrussi, A., Dickinson, L. K., & Lehmann, R. (1991). oskar organizes the germ plasm and directs localization of the posterior determinant nanos. *Cell*, 66(1), 37–50. doi: 10.1016/0092-8674(91)90137-n

Erson-Bensan, A. E., & Can, T. (2016). Alternative Polyadenylation: Another Foe in Cancer. *Molecular Cancer Research*, 14(6), 507–517. doi: 10.1158/1541-7786.mcr-15-0489

Fu, Y., Chen, L., Chen, C., Ge, Y., Kang, M., Song, Z., ... Xu, A. (2018). Crosstalk between alternative polyadenylation and miRNAs in the regulation of protein translational efficiency. *Genome Research*, 28(11), 1656–1663. doi: 10.1101/gr.231506.117

Green, M. R., & Sambrook, J. (2012). *Molecular cloning: a laboratory manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Gruber, A. R., Martin, G., Keller, W., & Zavolan, M. (2012). Cleavage factor Im is a key regulator of 3' UTR length. *RNA Biology*, 9(12), 1405–1412. doi: 10.4161/rna.22570

Gruber, A. R., Martin, G., Keller, W., & Zavolan, M. (2013). Means to an end: mechanisms of alternative polyadenylation of messenger RNA precursors. *Wiley Interdisciplinary Reviews: RNA*, 5(2), 183–196. doi: 10.1002/wrna.1206

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

Huang, Y., & Carmichael, G. G. (1996). Role of polyadenylation in nucleocytoplasmic transport of mRNA. *Molecular and Cellular Biology*, 16(4), 1534–1542. doi: 10.1128/mcb.16.4.1534

Ji, Z., Lee, J. Y., Pan, Z., Jiang, B., & Tian, B. (2009). Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse

embryonic development. *Proceedings of the National Academy of Sciences*, 106(17), 7028–7033. doi: 10.1073/pnas.0900028106

Kaufmann, I., Martin, G., Friedlein, A., Langen, H., & Keller, W. (2004). Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. *The EMBO Journal*, 23(3), 616–626. doi: 10.1038/sj.emboj.7600070

Kim, D. I., Kc, B., Zhu, W., Motamedchaboki, K., Doye, V., & Roux, K. J. (2014). Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proceedings of the National Academy of Sciences*, 111(24). doi: 10.1073/pnas.1406459111

Krieg, P. A., & Melton, D. A. (1984). Formation of the 3' end of histone mRNA by post-transcriptional processing. *Nature*, 308(5955), 203–206. doi: 10.1038/308203a0

Kühn, U., Gündel, M., Knoth, A., Kerwitz, Y., Rüdell, 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. doi: 10.1074/jbc.m109.018226

Kwon, K., & Beckett, D. (2000). Function of a conserved sequence motif in biotin holoenzyme synthetases. *Protein Science*, 9(8), 1530–1539. doi: 10.1110/ps.9.8.1530

Kwon, K., Streaker, E. D., & Beckett, D. (2009). Binding specificity and the ligand dissociation process in the E. coli biotin holoenzyme synthetase. *Protein Science*, 11(3), 558–570. doi: 10.1110/ps.33502

Lackford, B., Yao, C., Charles, G. M., Weng, L., Zheng, X., Choi, E.-A., ... Shi, Y. (2014). Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. *The EMBO Journal*, 33(8), 878–889. doi: 10.1002/embj.201386537

- Lau, A. G., Irier, H. A., Gu, J., Tian, D., Ku, L., Liu, G., ... Feng, Y. (2010). Distinct 3UTRs differentially regulate activity-dependent translation of brain-derived neurotrophic factor (BDNF). *Proceedings of the National Academy of Sciences*, 107(36), 15945–15950. doi: 10.1073/pnas.1002929107
- Lee, H. O., & Sheen, Y. Y. (1997). Estrogen modulation of human breast cancer cell growth. *Archives of Pharmacal Research*, 20(6), 566–571. doi: 10.1007/bf02975213
- Loya, A., Pnueli, L., Yosefzon, Y., Wexler, Y., Ziv-Ukelson, M., & Arava, Y. (2008). The 3-UTR mediates the cellular localization of an mRNA encoding a short plasma membrane protein. *Rna*, 14(7), 1352–1365. doi: 10.1261/rna.867208
- Macdonald, C. C., Wilusz, J., & Shenk, T. (1994). The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. *Molecular and Cellular Biology*, 14(10), 6647–6654. doi: 10.1128/mcb.14.10.6647
- Mandel, C. R., Bai, Y., & Tong, L. (2007). Protein factors in pre-mRNA 3'-end processing. *Cellular and Molecular Life Sciences*, 65(7-8), 1099–1122. doi: 10.1007/s00018-007-7474-3
- Mandel, C. R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J. L., & Tong, L. (2006). Polyadenylation factor CPSF-73 is the pre-mRNA 3-end-processing endonuclease. *Nature*, 444(7121), 953–956. doi: 10.1038/nature05363
- Martin, G., Gruber, A. R., Keller, W., & Zavolan, M. (2012). Genome-wide Analysis of Pre-mRNA 3' End Processing Reveals a Decisive Role of Human Cleavage Factor I in the Regulation of 3' UTR Length. *Cell Reports*, 1(6), 753–763. doi: 10.1016/j.celrep.2012.05.003
- Masamha, C. P., Xia, Z., Yang, J., Albrecht, T. R., Li, M., Shyu, A.-B., ... Wagner, E. J. (2014). CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature*, 510(7505), 412–416. doi: 10.1038/nature13261



- Mayr, C., & Bartel, D. P. (2009). Widespread Shortening of 3'UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells. *Cell*, 138(4), 673–684. doi: 10.1016/j.cell.2009.06.016
- Mavrigh, T. N., Ioshikhes, I. P., Venters, B. J., Jiang, C., Tomsho, L. P., Qi, J., ... Pugh, B. F. (2008). A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Research*, 18(7), 1073–1083. doi: 10.1101/gr.078261.108
- Mechold, U., Gilbert, C., & Ogryzko, V. (2005). Codon optimization of the BirA enzyme gene leads to higher expression and an improved efficiency of biotinylation of target proteins in mammalian cells. *Journal of Biotechnology*, 116(3), 245–249. doi: 10.1016/j.jbiotec.2004.12.003
- Millevoi, S., & Vagner, S. (2009). Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Research*, 38(9), 2757–2774. doi: 10.1093/nar/gkp1176
- Moore, M. J. (2005). From Birth to Death: The Complex Lives of Eukaryotic mRNAs. *Science*, 309(5740), 1514–1518. doi: 10.1126/science.1111443
- Nam, J.-W., Rissland, O. S., Koppstein, D., Abreu-Goodger, C., Jan, C. H., Agarwal, V., ... Bartel, D. P. (2014). Global Analyses of the Effect of Different Cellular Contexts on MicroRNA Targeting. *Molecular Cell*, 53(6), 1031–1043. doi: 10.1016/j.molcel.2014.02.013
- Neve, J., Burger, K., Li, W., Hoque, M., Patel, R., Tian, B., ... Furger, A. (2015). Subcellular RNA profiling links splicing and nuclear DICER1 to alternative cleavage and polyadenylation. *Genome Research*, 26(1), 24–35. doi: 10.1101/gr.193995.115
- Niedner, A., Edelmann, F. T., & Niessing, D. (2014). Of social molecules: The interactive assembly ofASH1mRNA-transport complexes in yeast. *RNA Biology*, 11(8), 998–1009. doi: 10.4161/rna.29946

- Oktaba, K., Zhang, W., Lotz, T. S., Jun, D. J., Lemke, S. B., Ng, S. P., ... Hilgers, V. (2015). ELAV Links Paused Pol II to Alternative Polyadenylation in the *Drosophila* Nervous System. *Molecular Cell*, 57(2), 341–348. doi: 10.1016/j.molcel.2014.11.024
- Proudfoot, N. J., Furger, A., & Dye, M. J. (2002). Integrating mRNA Processing with Transcription. *Cell*, 108(4), 501–512. doi: 10.1016/s0092-8674(02)00617-7
- 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 polypolyadenylation signal selection. *The EMBO Journal*, 30(12), 2431–2444. doi: 10.1038/emboj.2011.156
- Reid, D. W., & Nicchitta, C. V. (2015). Diversity and selectivity in mRNA translation on the endoplasmic reticulum. *Nature Reviews Molecular Cell Biology*, 16(4), 221–231. doi: 10.1038/nrm3958
- Roux, K. J., Kim, D. I., Burke, B., & May, D. G. (2018). BioID: A Screen for Protein-Protein Interactions. *Current Protocols in Protein Science*, 91(1). doi: 10.1002/cpps.51
- Roux, K. J., Kim, D. I., Raida, M., & Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *The Journal of Cell Biology*, 196(6), 801–810. doi: 10.1083/jcb.201112098
- Sachs, A. B., Sarnow, P., & Hentze, M. W. (1997). Starting at the Beginning, Middle, and End: Translation Initiation in Eukaryotes. *Cell*, 89(6), 831–838. doi: 10.1016/s0092-8674(00)80268-8
- Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A., & Burge, C. B. (2008). Proliferating Cells Express mRNAs with Shortened 3' Untranslated Regions and Fewer MicroRNA Target Sites. *Science*, 320(5883), 1643–1647. doi: 10.1126/science.1155390

- Satiaputra, J., Sternicki, L. M., Hayes, A. J., Pukala, T. L., Booker, G. W., Shearwin, K. E., & Polyak, S. W. (2019). Native mass spectrometry identifies an alternative DNA-binding pathway for BirA from *Staphylococcus aureus*. *Scientific Reports*, 9(1). doi: 10.1038/s41598-019-39398-6
- Schönemann, L., Kühn, U., Martin, G., Schäfer, P., Gruber, A. R., Keller, W., ... Wahle, E. (2014). Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. *Genes & Development*, 28(21), 2381–2393. doi: 10.1101/gad.250985.114
- Scorilas, A. (2002). Polyadenylate Polymerase (PAP) and 3' End pre-mRNA Processing: Function, Assays, and Association with Disease. *Critical Reviews in Clinical Laboratory Sciences*, 39(3), 193–224. doi: 10.1080/10408360290795510
- Shivaswamy, S., Bhinge, A., Zhao, Y., Jones, S., Hirst, M., & Iyer, V. (2008). Dynamic Remodeling of Individual Nucleosomes Across a Eukaryotic Genome in Response to Transcriptional Perturbation. *SciVee*. doi: 10.4016/5532.01
- Spies, N., Nielsen, C. B., Padgett, R. A., & Burge, C. B. (2009). Biased Chromatin Signatures around Polyadenylation Sites and Exons. *Molecular Cell*, 36(2), 245–254. doi: 10.1016/j.molcel.2009.10.008
- Spraggon, L., & Cartegni, L. (2013). U1 snRNP-Dependent Suppression of Polyadenylation: Physiological Role and Therapeutic Opportunities in Cancer. *International Journal of Cell Biology*, 2013, 1–10. doi: 10.1155/2013/846510
- Stacey, S. N., Sulem, P., Jonasdottir, A., Masson, G., Gudmundsson, J., Gudbjartsson, D. F., *et al.* (2011) A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. *Nat Genet* 2011;43:1098–103.
- Takagaki, Y., & Manley, J. L. (1997). RNA recognition by the human polyadenylation factor CstF. *Molecular and Cellular Biology*, 17(7), 3907–3914. doi: 10.1128/mcb.17.7.3907

Takagaki, Y., & Manley, J. L. (2000). Complex Protein Interactions within the Human Polyadenylation Machinery Identify a Novel Component. *Molecular and Cellular Biology*, 20(5), 1515–1525. doi: 10.1128/mcb.20.5.1515-1525.2000

Thomas, L. F., & Sætrom, P. (2012). Single Nucleotide Polymorphisms Can Create Alternative Polyadenylation Signals and Affect Gene Expression through Loss of MicroRNA-Regulation. *PLoS Computational Biology*, 8(8). doi: 10.1371/journal.pcbi.1002621

Wahle, E., & Rügsegger, U. (1999). 3'-End processing of pre-mRNA in eukaryotes. *FEMS Microbiology Reviews*, 23(3), 277–295. doi: 10.1111/j.1574-6976.1999.tb00400.x

Wang, E. T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., ... Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature*, 456(7221), 470–476. doi: 10.1038/nature07509

Wilusz, J., & Shenk, T. (1990). A uridylyate tract mediates efficient heterogeneous nuclear ribonucleoprotein C protein-RNA cross-linking and functionally substitutes for the downstream element of the polyadenylation signal. *Molecular and Cellular Biology*, 10(12), 6397–6407. doi: 10.1128/mcb.10.12.6397

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'-UTR landscape across seven tumour types. *Nature Communications*, 5(1). doi: 10.1038/ncomms6274

Yang, Q., & Doublé, S. (2011). Structural biology of poly(A) site definition. *Wiley Interdisciplinary Reviews: RNA*, 2(5), 732–747. doi: 10.1002/wrna.88

Yao, C., Biesinger, J., Wan, J., Weng, L., Xing, Y., Xie, X., & Shi, Y. (2012). Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. *Proceedings of the National Academy of Sciences*, 109(46), 18773–18778. doi: 10.1073/pnas.1211101109

Yao, Y., Song, L., Katz, Y., & Galili, G. (2002). Cloning and characterization of Arabidopsis homologues of the animal CstF complex that regulates 3' mRNA cleavage and polyadenylation. *Journal of Experimental Botany*, 53(378), 2277–2278. doi: 10.1093/jxb/erf073

Youngblood, B. A., Grozdanov, P. N., & Macdonald, C. C. (2014). CstF-64 supports pluripotency and regulates cell cycle progression in embryonic stem cells through histone 3' end processing. *Nucleic Acids Research*, 42(13), 8330–8342. doi: 10.1093/nar/gku551

Yuan, F., Hankey, W., Wagner, E. J., Li, W., & Wang, Q. (2019). Alternative polyadenylation of mRNA and its role in cancer. *Genes & Diseases*. doi: 10.1016/j.gendis.2019.10.011



## APPENDICES

### A. Plasmid Maps

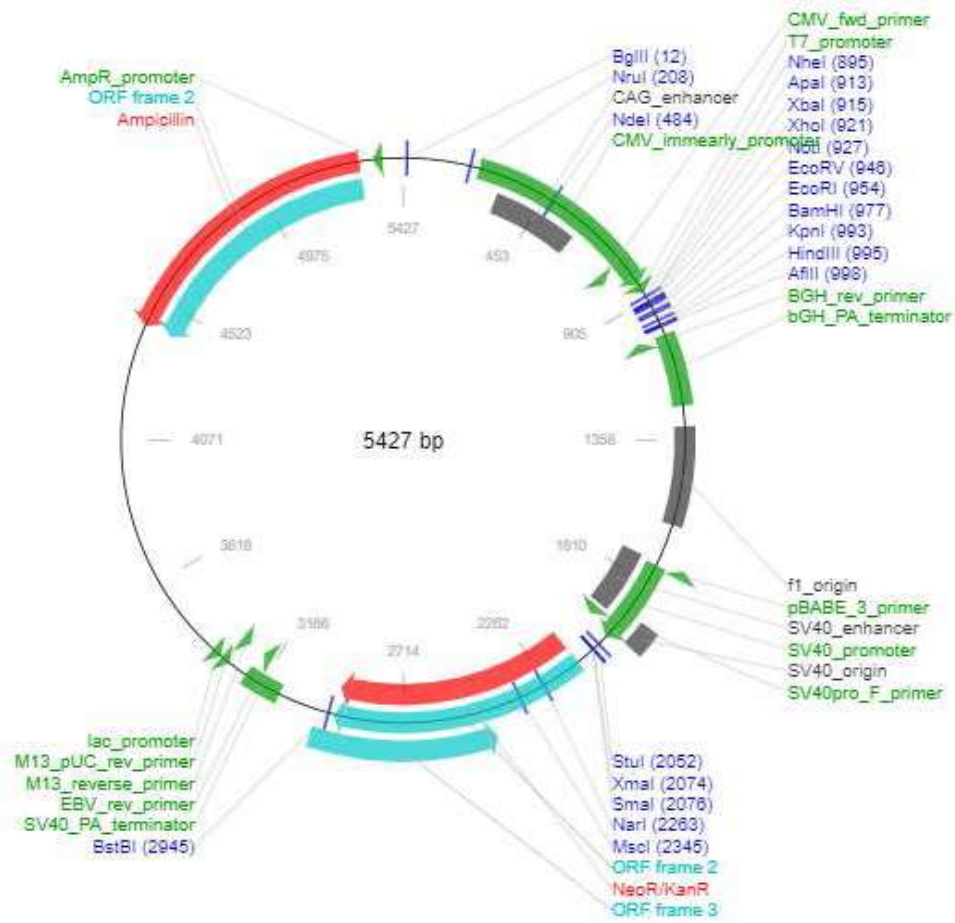


Figure A.1. pcDNA 3.1(-) Map

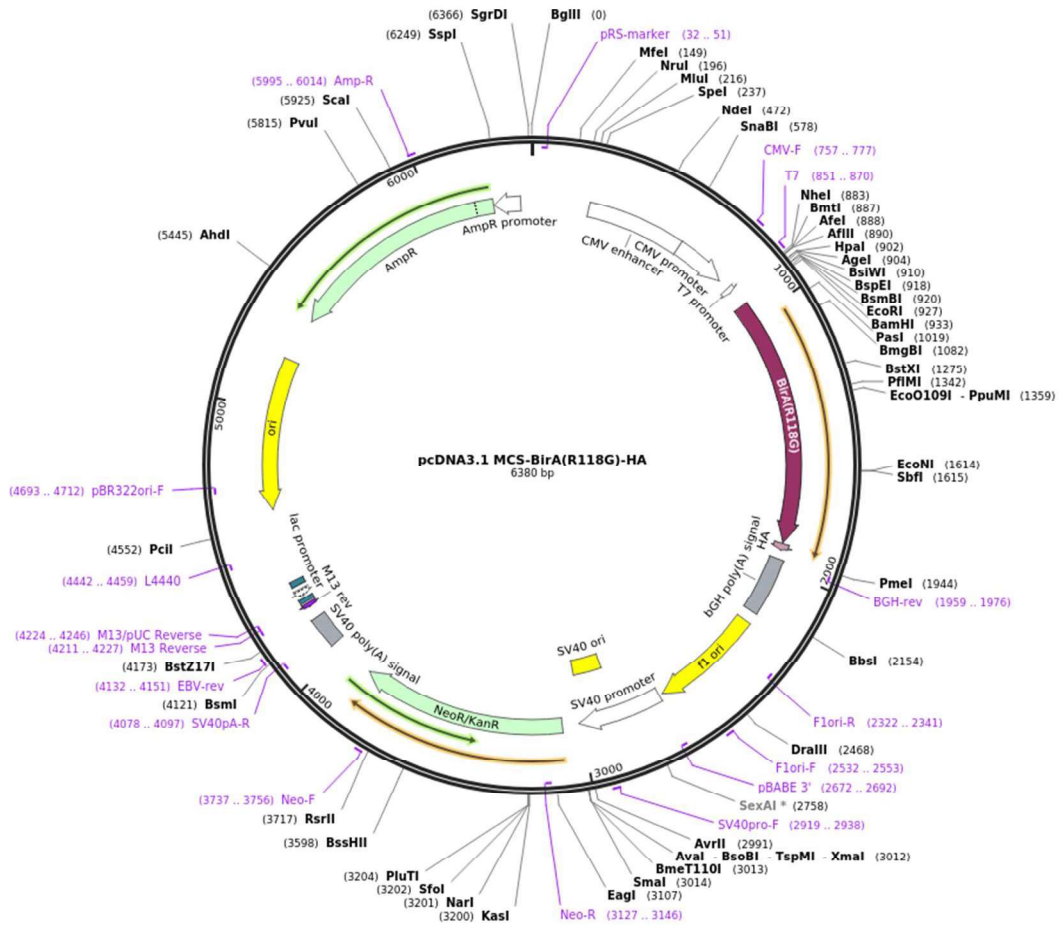


Figure A.2. pcDNA3.1 MCS-BirA(R118G)-HA Map



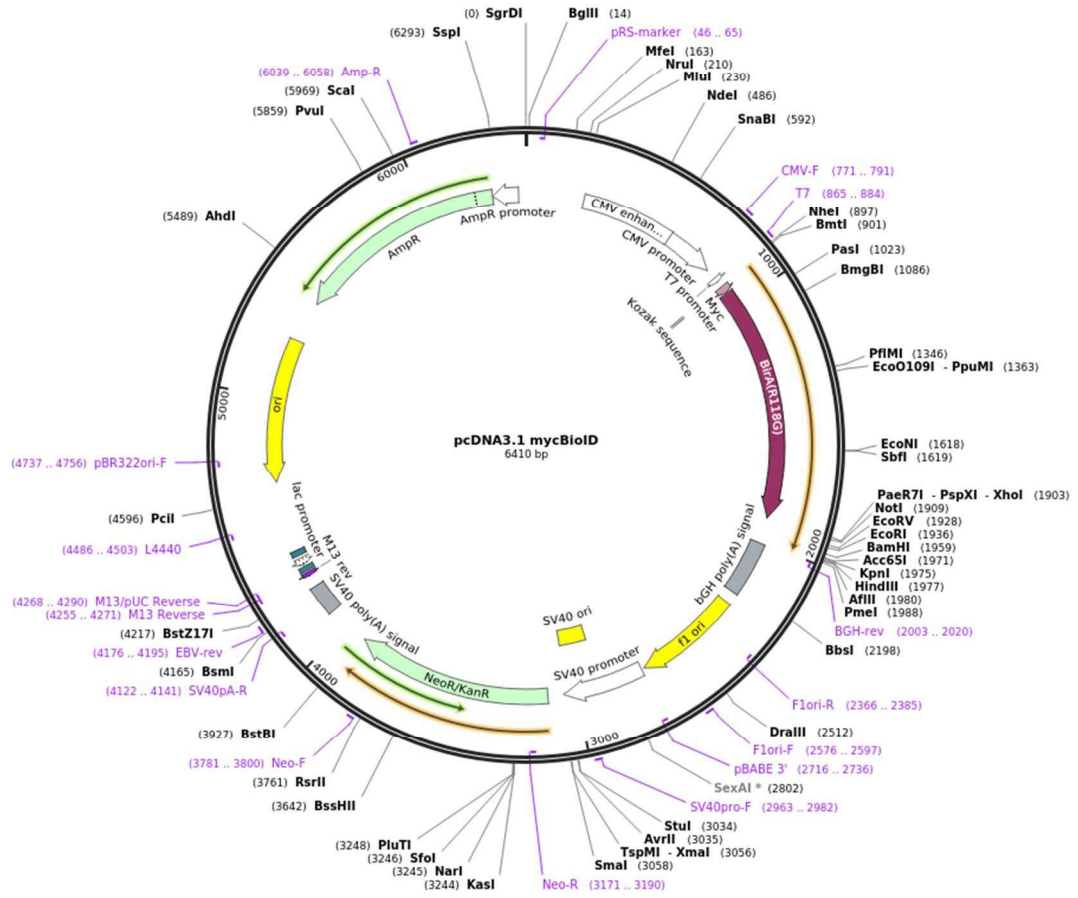


Figure A.3. pcDNA3.1 mycBioID Map

## B. Markers

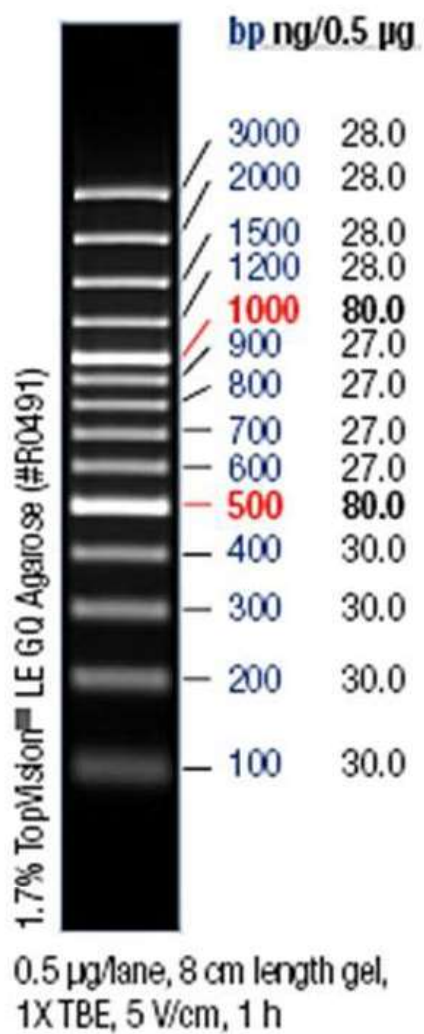


Figure B.1. GeneRuler 100 bp Plus DNA Ladder (SM0321, Thermo Scientific)

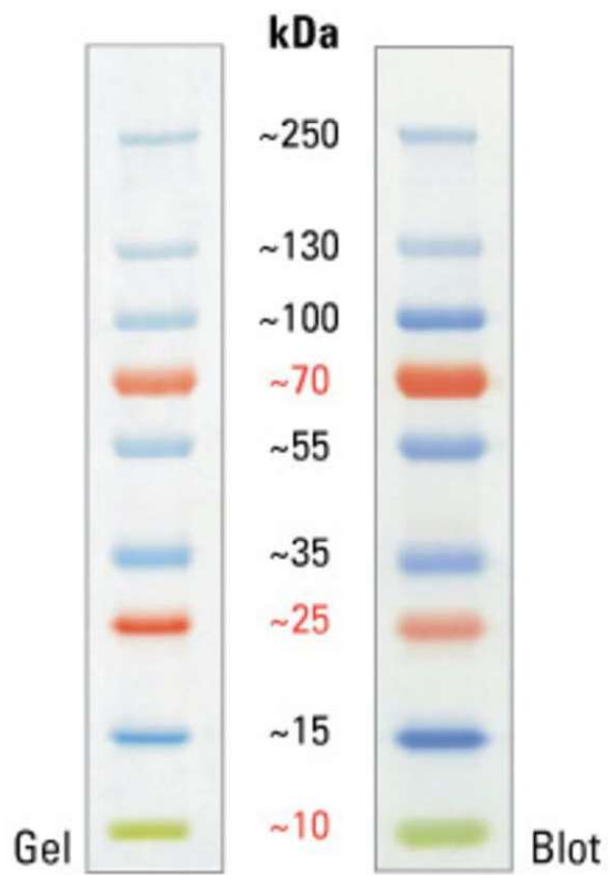


Figure B.2. PageRuler™ Plus Prestained Protein Ladder, (26619, Thermo Scientific)

## **C. Buffers**

### **6x Laemmli Buffer**

0.375 M Tris

60% Glycerol

12% SDS

0.012% bromophenol blue

30% 2-mercaptoethanol

### **0.1% TBS-T**

137 mM NaCl

20 mM Tris, pH: 7.6

0.1% Tween 20

### **Lysis Buffer A:**

HEPES        30 mM

KCl            10 mM

MgCl<sub>2</sub>        1.5 mM

EDTA         0.1 mM

DTT            1 mM

Protease inhibitor cocktail

**Lysis Buffer B:**

HEPES	30 mM
KCl	10 mM
EDTA	0.1 mM
DTT	1 mM
MgCl <sub>2</sub>	1.5 mM
NaCl	500 mM
Glycerol	25%
Protease inhibitor cocktail	



## CURRICULUM VITAE

### PERSONAL INFORMATION

Surname, Name: Çiçek, Mustafa  
Nationality: Turkish (TC)  
Date and Place of Birth: 30 August 1985, Mersin  
Marital Status: Married  
Phone: +90 312 210 50 44  
email: cicekm@metu.edu.tr

### EDUCATION

<b>Degree</b>	<b>Institution</b>	<b>Year of Graduation</b>
MSc	METU Biotechnology	2014
BS	IU Molecular Biology and Genetics	2008
High School	İçel Anatolian High School, Mersin	2003

### WORK EXPERIENCE

<b>Year</b>	<b>Place</b>	<b>Enrollment</b>
2011-Present	METU Dept of Biological Sci.	Research Assistant
2010-2011	KMU Dept. of Biology	Research Assistant

### FOREIGN LANGUAGES

Advanced English

### PUBLICATIONS

1. M. CİCEK., O. MUTLU, A. ERDEMİR, E. OZKAN, Y. SARIÇAY, D. BALIK, "Single Mutation in Shine-Dalgarno-Like Sequence Present in the Amino Terminal of Lactate Dehydrogenase of Plasmodium Effects the Production of an Eukaryotic Protein Expressed in a Prokaryotic System", MOLECULAR BIOTECHNOLOGY, 2013 vol.54, pp.602-608