

ESTROGEN-INDUCED UPREGULATION AND 3'-UTR SHORTENING OF  
*CDC6*

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

BY

H. BEGÜM AKMAN

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

MARCH 2013



Approval of the Thesis:

**ESTROGEN-INDUCED UPREGULATION AND 3'-UTR SHORTENING OF CDC6**

Submitted by **H. BEGÜM AKMAN** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Biological Sciences Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen  
Dean, **Graduate School of Natural and Applied Sciences**

\_\_\_\_\_

Prof. Dr. Gülay Özcengiz  
Head of Department, **Biological Sciences Dept., METU**

\_\_\_\_\_

Assoc. Prof. Dr. A.Elif Erson-Bensan  
Supervisor, **Biological Sciences Dept., METU**

\_\_\_\_\_

**Examining Committee Members:**

Prof. Dr. İnci Togan  
Biological Sciences Dept., METU

\_\_\_\_\_

Assoc. Prof. Dr. A.Elif Erson-Bensan  
Biological Sciences Dept., METU

\_\_\_\_\_

Assoc. Prof. Dr. Sreeparna Banerjee  
Biological Sciences Dept., METU

\_\_\_\_\_

Assoc. Prof. Dr. Mesut Muyan  
Biological Sciences Dept., METU

\_\_\_\_\_

Assoc. Prof. Dr. Cengiz Yakıcıer  
Medical Biology Dept., Acibadem University

\_\_\_\_\_

**Date:** 25. 03. 2013

**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: H. Begüm Akman

Signature :

## ABSTRACT

### ESTROGEN-INDUCED UPREGULATION AND 3'-UTR SHORTENING OF *CDC6*

Akman, H. Begüm  
PhD, Department of Biology  
Supervisor: Assoc. Prof. Dr. A. Elif Erson-Bensan

March 2013, 86 pages

Alternative polyadenylation (APA) may cause mRNA 3'-UTR (untranslated region) shortening in different physiological conditions and/or disease states. About half of mammalian genes use alternative cleavage and polyadenylation to generate multiple mRNA isoforms differing in their 3'-UTRs. Consequently, 3'-UTR shortening of mRNAs via APA has important consequences for gene expression. Studies showed that, by preferential use of proximal APA sites and switching to shorter 3'-UTRs, proliferating cells can avoid 3'-UTR dependent negative post-transcriptional regulations. Hence, such APA and 3'-UTR shortening events may explain the basis of some of the proto-oncogene activation cases observed in cancer cells. Based on the fact that certain cell types switch to more proximal poly(A) sites to rapidly increase translation rates by escaping from miRNAs, we hypothesized that upon estradiol (E2) treatment in estrogen receptor positive (ER+) breast cancer cells, such proximal poly(A) sites may be preferred to provide a rapid growth pattern. Initial probe based screen of independent expression arrays suggested upregulation and 3'-UTR shortening of an essential regulator of DNA replication, *CDC6* (Cell Division Cycle 6), upon E2 treatment. Further investigations confirmed the E2 and ER dependent upregulation and 3'-UTR shortening of *CDC6*, which lead to increased *CDC6* protein levels and higher BrdU incorporation. Consequently, miRNA binding predictions and dual luciferase assays suggested that 3'-UTR shortening of *CDC6* was a likely mechanism to avoid 3'-UTR dependent negative regulations. In summary, we demonstrated *CDC6* APA induction by the proliferative effect of E2 in ER+ cells and provided new insights into the complex regulation of APA. E2 induced APA is likely to be an important but previously overlooked mechanism of E2 responsive gene expression.

Key words: alternative polyadenylation, 3'-UTR, APA, *CDC6*.

## ÖZ

# ÖSTROJEN ETKİSİYLE *CDC6* İFADESİNİN ARTIŞI VE 3'-UTR'SİNİN KISALMASI

Akman H. Begüm  
PhD, Department of Biology  
Tez Yöneticisi: Doç. Dr. A. Elif Erson-Bensan

Mart 2013, 86 sayfa

Alternatif poliadenilasyon (APA) farklı fizyolojik ve/veya hastalık durumlarında mRNA 3'-UTR'lerinin (translasyon olmayan bölge) kısılmasına neden olabilir. Memeli genlerinin yaklaşık yarısı alternatif kesilme ve poliadenilasyon ile 3'-UTR boyları değişen çeşitli mRNA izoformları oluşturabilirler. Sonuç olarak, mRNA'ların 3'-UTR'lerinin kısılmasının gen ifadesi açısından önemli sonuçları vardır. APA'nın hızlı büyüyen hücrelerde ve kanser hücrelerinde kullanılan bir mekanizma olduğu yakın zamanda anlaşılmıştır. APA ve 3'-UTR kısılması olaylarının kanser hücrelerinde ortaya çıkan birtakım proto-onkogen aktivasyonlarını açıklayabileceği düşünülmektedir. Çalışmamızda, bazı hücrelerin translasyon hızlarını arttırmak amacıyla proksimal poly(A) sitelerini tercih ettiklerine dayanarak, östrojen reseptör pozitif (ER+) meme kanseri hücre hatlarına östrojene (E2) bağlı olarak hızlı büyümeye olanak sağlayacak proksimal poly(A) sitelerinin tercih edilebileceğini düşündük. Bağımsız ifade mikroçiplerinde yapılan öncül prob-tabanlı analizler DNA replikasyonunda önemli bir regülatör olan *CDC6*'nın (Cell Division Cycle 6) E2 muamelesi sonrası ifadesinde artış ve 3'-UTR'sinde kısılma olduğuna işaret etmiştir. Takip eden deneylerde, *CDC6*'nın E2 ve ER'ye bağlı ifade artışı ve 3'-UTR kısılması, buna bağlı olarak *CDC6* protein seviyesindeki ve BrdU eklenmesindeki artış doğrulanmıştır. Daha sonra yapılan miRNA bağlanma tahminleri ve dual-lusiferaz sonuçları *CDC6*'nın 3'-UTR'sinin kısılmasının 3'-UTR'a bağlı negatif regülasyonlardan kaçınmak amacıyla olduğuna işaret etmiştir. Sonuç olarak, E2'nin ER+ hücrelerde bölünmeyi tetikleyen etkisinin *CDC6* APA'sını da etkilediğini göstermiş, ve de APA'nın komplike regülasyonuna yeni bir bakış açısı sağlanmıştır. E2 ile tetiklenen APA, E2'ye duyarlı gen ifade mekanizmaları açısından daha önceden gözden kaçan ancak çok önemli bir mekanizma olması muhtemeldir.

Anahtar kelimeler: alternatif poliadenilasyon, 3'-UTR, APA, *CDC6*

## ACKNOWLEDGEMENTS

I would like to thank my supervisor Assoc. Prof. Dr. A. Elif Erson-Bensan for her encouragement and support throughout my study.

I would like to thank all my thesis committee members, Prof. Dr. İnci Togan, Assoc. Prof. Dr. Mesut Muyan, Assoc. Prof. Dr. Sreeparna Banerjee, Assoc. Prof. Dr. Cengiz Yakıcıer.

I would like to express my sincere gratitude to Assoc. Prof. Dr. Tolga Can for the bioinformatics support. I am deeply grateful to Assoc. Prof. Dr. Mesut Muyan for his invaluable suggestions and inspiration. I would like to thank Assoc. Prof. Dr. Uygur Tazebay and Assoc. Prof. Dr. Sreeparna Banerjee for sharing resources.

I thank all current and previous lab members.

I would like to express my gratitude to my family.

Finally, I would like to thank Taner for his endless patience.

I would like to thank TUBITAK for supporting me by National Scholarship Programme for PhD Students (2007-2012).

## TABLE OF CONTENTS

ABSTRACT.....	iv
ÖZ.....	v
ACKNOWLEDGEMENTS.....	vi
TABLE OF CONTENTS.....	vii
LIST OF TABLES.....	ix
LIST OF FIGURES.....	x
LIST OF ABBREVIATIONS.....	xii
CHAPTERS	
1. INTRODUCTION.....	1
1.1    Alternative Polyadenylation.....	1
1.1.1    mRNA 3'-end formation.....	1
1.1.1.1    Polyadenylation Mechanism.....	1
1.1.1.2    Types of Polyadenylation.....	2
1.1.2    APA Regulating Molecular Mechanisms.....	3
1.1.2.1    Trans-acting Regulator Proteins.....	4
1.1.2.2    Cis-acting RNA Elements.....	5
1.1.2.3    Chromatin and Epigenetic Regulation of APA.....	5
1.1.3    Genome-wide Search for APA.....	6
1.1.4    Examples of APA.....	7
1.2    microRNAs and Cancer.....	10
1.3    Estrogen Receptor Signaling.....	11
1.4    Aim of the Study.....	12
2. MATERIALS AND METHODS.....	13
2.1    Probe Level Screen for APA Transcripts.....	13
2.2    Cell Lines and Treatments.....	13
2.3    Expression Analysis.....	14
2.3.1    RNA Isolation.....	14
2.3.2    DNA contamination, Quantification and Integrity Assessment.....	14
2.3.3    cDNA synthesis.....	15



2.3.4	<i>TFF1</i> Expression Analysis.....	16
2.3.5	RT-qPCR Analysis .....	16
2.4	Rapid Amplification of cDNA ends .....	17
2.5	Western Blotting.....	18
2.5.1	Nuclear Protein Isolation .....	18
2.5.2	Total Protein Isolation .....	18
2.5.3	Western Blotting .....	19
2.6	BrdU (Bromodeoxyuridine) Incorporation.....	19
2.7	microRNA Target Sites on mRNAs .....	20
2.8	Cloning of 3'-UTR Isoforms.....	20
2.8.1	Dual-luciferase Assay.....	21
3.	RESULTS AND DISCUSSION .....	23
3.1	<i>CDC6</i> 3'-UTR and Probe Level Screen .....	23
3.2	Expression Analysis.....	25
3.2.1	E2 Treatment of ER+ Cells .....	25
3.2.2	RT-qPCR Analysis .....	26
3.3	Rapid Amplification of cDNA Ends (RACE) .....	28
3.4	<i>CDC6</i> Protein Levels Upon E2 Treatment .....	30
3.5	BrdU Incorporation.....	31
3.6	<i>CDC6</i> mRNA-miRNA Interactions.....	34
3.6.1	MicroRNA target sites on <i>CDC6</i> Isoforms .....	34
3.6.2	Cloning of 3'-UTR Isoforms .....	35
3.6.3	Dual-luciferase Assay.....	37
3.7	E2 and ER Dependency of <i>CDC6</i> APA.....	40
3.7.1	ER Dependency of <i>CDC6</i> APA .....	40
3.8	Transcriptional Regulators of APA .....	44
4.	CONCLUSION .....	47
	REFERENCES .....	51
	APPENDICES	
A.	Primers.....	59
B.	MIQE Guidelines Checklist.....	60
C.	DNA Contamination, RNA Quantification and Integrity Assessment.....	63
D.	RT-qPCR Assay Performance Results.....	65
E.	Buffers.....	74

F. Markers.....	76
G. Plasmid Maps.....	79
H. Sequencing Results.....	81
CURRICULUM VITAE.....	83

## LIST OF TABLES

### TABLES

<b>Table 2.1.</b> Reaction Conditions of DNase Treatment.....	15
<b>Table 2.2.</b> Reaction Conditions of Reverse Transcription.....	16
<b>Table 3.1.</b> Raw intensities of 11 probes for <i>CDC6</i> from three gene expression CEL files...	25
<b>Table 3.2.</b> <i>CDC6</i> 3'-UTR binding miRNA predictions.....	35
<b>Table A.1.</b> Primers used in the study.....	59
<b>Table B.2.</b> MIQE Guidelines Checklist.....	60

## TABLE OF FIGURES

Figure 1.1. Types of polyadenylation events. ....	3
Figure 1.2. APA regulating mechanisms. ....	4
Figure 1.3. APA may affect different cellular events. ....	7
Figure 1.4. A model for regulation of APA events in differentiation/proliferation. ....	9
Figure 1.5. Estrogenic Response. ....	12
Figure 3.1. <i>CDC6</i> 3'-UTR and the positions of the probes. ....	24
Figure 3.2. <i>TFF1</i> is an E2 upregulated gene. ....	26
Figure 3.3. Relative quantification of <i>CDC6</i> 3'-UTR short and long isoforms in E2 treated and control cells. ....	27
Figure 3.4. Figure 3.5. 3'-RACE confirmed the existence of the short isoform. ....	29
Figure 3.5. Sequencing of the 3'-RACE product. ....	30
Figure 3.6. E2 treatment caused increase of the <i>CDC6</i> protein levels. ....	30
Figure 3.7. E2 treatment induced increased BrdU incorporation. ....	31
Figure 3.8. Investigation of MDA-MB-231 cells. ....	33
Figure 3.9. Cloning <i>CDC6</i> 3'-UTR isoforms. ....	36
Figure 3.10. Colony PCR Results. ....	37
Figure 3.11. Dual luciferase reporter assay. ....	38
Figure 3.12. Dual luciferase reporter assay after E2 treatment. ....	39
Figure 3.13. ER protein levels detected in cell lines used in the study. ....	40
Figure 3.14. E2 does not induce <i>CDC6</i> 3' UTR isoform increase in ER silenced MCF7 cells. ....	41
Figure 3.15. E2 induced <i>CDC6</i> expression requires ER and de novo protein synthesis. ....	42
Figure 3.16. <i>PCNA</i> expression detected by RT-qPCR after ICI and CHX treatments. ....	44
Figure 3.17. Expression levels of <i>E2Fs</i> and APA factors after E2 treatment. ....	45
Figure C. 1. Lack of DNA contamination in RNA samples was assessed. ....	63
Figure C. 2. RNA concentrations were determined using NanoDrop ND1000. ....	63
Figure C. 3. RNA integrity assessment. ....	64
Figure F. 1. GeneRuler 100 bp DNA Ladder Plus. ....	76
Figure F. 2. RiboRuler RNA Ladder. ....	77
Figure F. 3. PageRuler Pre-stained Protein Ladder. ....	78
Figure G. 1. pMIR-Report Map. ....	79

Figure G. 2. phRL-TK Map .....	80
Figure H. 1. <i>CDC6</i> short 3'-UTR isoform sequencing result. ....	81
Figure H. 2. <i>CDC6</i> long 3'-UTR isoform sequencing result. ....	82

## LIST OF ABBREVIATIONS

A	Adenosine
ANOVA	Analysis of variance
APA	Alternative polyadenylation
ARE	AU-rich elements
bp	Base pairs
BrdU	Bromodeoxyuridine
BSA	Bovine serum albumin
C	Cytosine
CDC6	Cell Division Cycle 6
CEL	CIMFast Event Language
cDNA	Complementary Deoxyribonucleic Acid
Chip	Chromatin immunoprecipitation
CHX	Cycloheximide
Cq	Quantification cycle
CPSF	Cleavage and polyadenylation-specific factor
CSTF	Cleavage stimulation factor
DNA	Deoxyribonucleic Acid
D-pA	Distal poly(A)
DNase I	Deoxyribonuclease I
dNTP	Deoxyribonucleotide triphosphate
ELISA	Enzyme-linked immunosorbent assay
E2	Estrogen
ER	Estrogen receptor
EtOH	Ethanol
EST	Expressed sequence tag
FBS	Fetal bovine serum
g	Centrifuge gravity force
G	Guanine
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GEO	Gene Expression Omnibus
HRP	Horseradish peroxidase
ICI	ICI 182,780
MIQE	Minimum information for publication of quantitative real-time PCR experiments
miRNA	microRNA
MPSS-DGE	Massively parallel signature sequencing digital gene expression
mRNA	messenger RNA
NCBI	National Center for Biotechnology Information
PAGE	Polyacrylamide gel electrophoresis
PBS-T	Phosphate buffered saline- Tween
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
P-pA	Proximal poly(A)
Pri-miRNA	Primary microRNA
Pre-miRNA	Precursor microRNA
Pre-mRNA	Precursor messenger RNA
RACE	Rapid amplification of cDNA ends

RBP	RNA binding protein
RT-qPCR	Real Time Quantitative Polymerase Chain Reaction
RLU	Relative light units
RNA	Ribonucleic acid
RNase	Ribonuclease
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SAGE	Serial analysis of gene expression
SBS-DGE	Sequencing-by-synthesis digital gene expression
SDHA	Succinate dehydrogenase
SDS	Sodium dodecyl sulfate
shRNA	Short hairpin RNA
T	Thymine
TBS-T	Tris buffered saline- Tween
TFF1	Trefoil factor 1
UTR	Untranslated Region
UV	Ultraviolet





# CHAPTER 1

## INTRODUCTION

### 1.1 Alternative Polyadenylation

#### 1.1.1 mRNA 3'-end formation

Regulation of mRNA processing is accepted to have essential roles in gene expression. mRNA processing involves splicing, capping, editing, and polyadenylation.

Excluding replication-dependent histone transcripts, all eukaryotic mRNAs are polyadenylated through a two-step reaction to produce mature 3'-ends. Polyadenylation starts with the endonucleolytic cleavage of the pre-mRNA at a specific site by polyadenylation factors. Next, an adenosine tail is synthesized by the poly(A) Polymerase (PAP). Poly(A) binding protein (PABP) stimulates PAP to catalyze the addition of adenosine residues and regulates the length of the poly(A) tail [1]. Overall, this process is orchestrated by several protein factors that are directed to specific mRNA sequences to accurately recognize the cleavage site within the pre-mRNA.

Poly(A) tail has impact on mRNA characteristics, such as stability, translation, and transport [2-5]. Addition of poly(A) tail happens almost synchronously with transcription. Hence pre-mRNA to mRNA maturation is a very rapid event [3, 6-8].

##### 1.1.1.1 Polyadenylation Mechanism

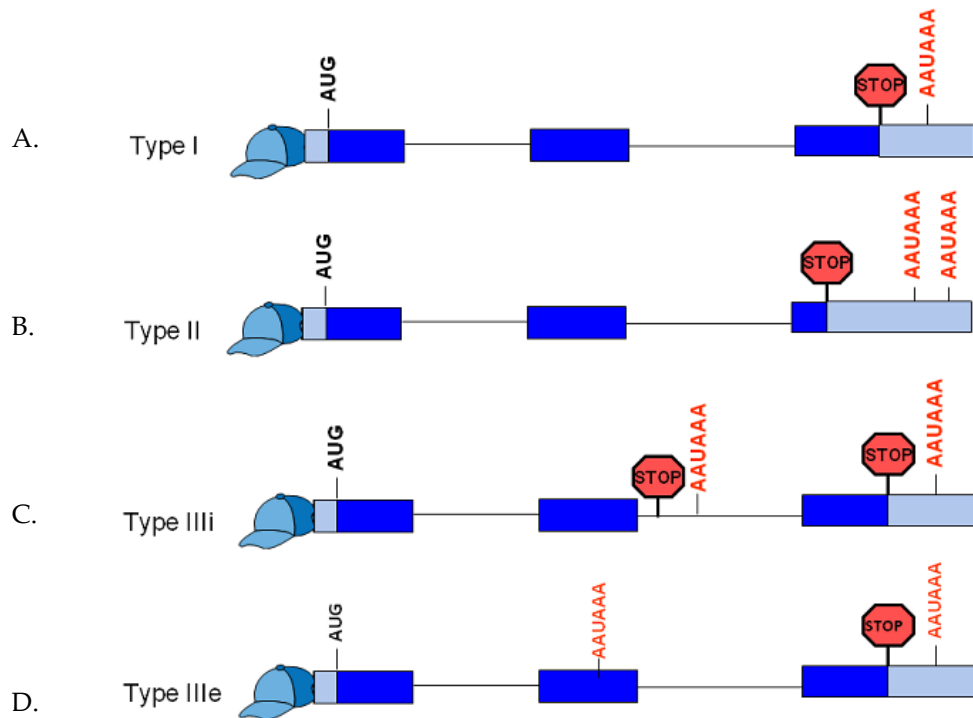
Eukaryotic polyadenylation signals generally contain a consensus sequence (AAUAAA) (or other non-canonical variants), ~10-35 nucleotides upstream of the actual cleavage and polyadenylation site. Recognition of this signal sequence and addition of the poly(A) tail requires a range of multi subunit protein complexes. CPSF (cleavage and polyadenylation specificity factor), CstF (cleavage stimulation factor), CFI (cleavage factor I), and CFII (cleavage factor II) along with accessory factors form the multi subunit complex that is

responsible for signal recognition and cleavage of the mRNA. Next, single subunit poly(A) polymerase (PAP) adds the poly(A) tail [9, 10]. Above mentioned 3'-end processing complex assembles on the pre-mRNA when CPSF and CstF mutually interact with the canonical AAUAAA signal. In addition, ~14-70 nucleotides downstream of the poly(A) signal, U/GU-rich regions are involved in directing polyadenylation. U/GU-rich regions provide a binding site for CstF64 subunit of the CstF protein complex [11].

#### **1.1.1.2 Types of Polyadenylation**

Recently, it has become apparent that alternative polyadenylation (APA) is a widespread mechanism used to provide further control on gene expression. More than 50% of human genes harbor multiple poly(A) signal sites, leading to multiple cleavage sites [12]. Hence, alternative sites of polyadenylation may create a 3'-UTR diversity of transcripts.

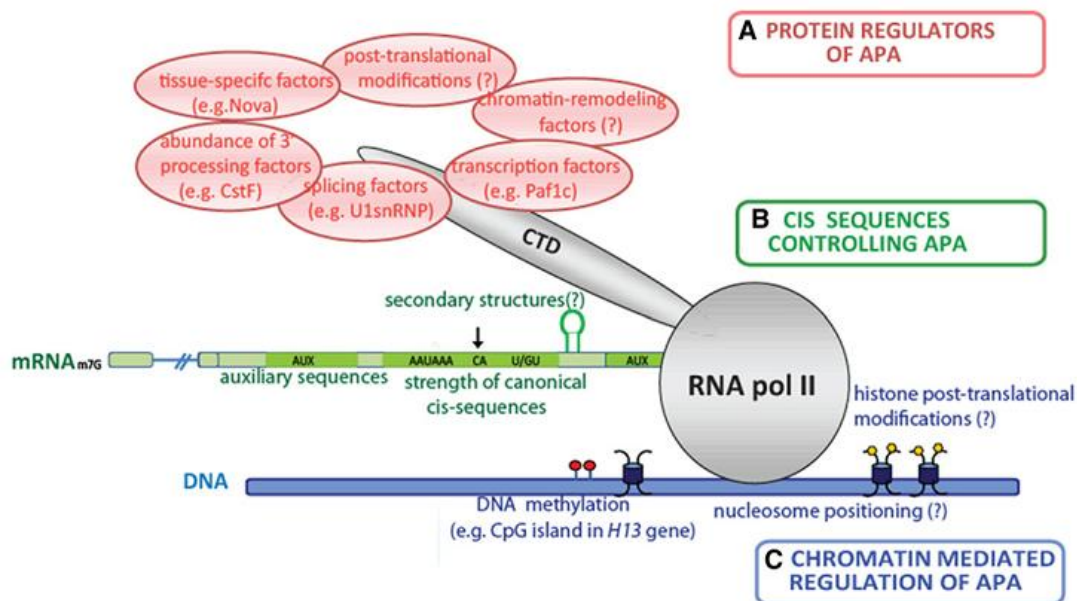
Such polyadenylation events can be classified into three possible groups (Figure 1.1). In the first group there is only one polyadenylation signal in the 3'-UTR, and this signal site leads to only one mRNA isoform (Figure 1.1 A). In the second group, multiple polyadenylation signals are present in the 3'-UTR and these signal sites give rise to different poly(A) tail positions in mature mRNA 3'-UTR isoforms (Figure 1.1 B). Lastly, in the third group, alternative splicing is coupled to alternative polyadenylation. Type III polyadenylation is further categorized in two as type IIIi (intronic APA signal) (Figure 1.1 C) and type IIIe (exonic APA signal) (Figure 1.1 D). Hence, different protein products may or may not be produced depending on the location of the stop codon. Type II and type III groups are both considered as possible APA events.



**Figure 1.1. Types of polyadenylation events.** **A.** mRNAs with only one polyadenylation signal classified as Type I. **B.** mRNAs having more than one polyadenylation signal in the 3' most exon classified as type II polyadenylation. **C.** Type III polyadenylation event is coupled to alternative splicing. In type IIIi type mRNAs have one or more polyadenylation signals in upstream introns **D.** In type IIIe mRNAs have one or more polyadenylation signals in exons. Untranslated regions (light blue boxes), coding regions (darker blue boxes), introns (lines). Figure is adapted from [13].

### 1.1.2 APA Regulating Molecular Mechanisms

The choice of using one polyadenylation site over another is determined by a combination of various features; 1) trans-acting regulator proteins, 2) *cis*-acting RNA elements, and 3) chromatin and epigenetic effects. All of these regulatory mechanisms are summarized in Figure 1.2 and explained in detail in the following sections.



**Figure 1.2. APA regulating mechanisms.** A) Variations in the quantity or activity of trans-acting factors such as core 3'-processing proteins and tissue specific RNA-binding proteins, B) combinatorial effects of strength of binding sites for core 3'-processing factors, auxiliary sequences, motifs directing the interaction of protein components with the mRNA, and possibly RNA secondary structures, C) influence of chromatin such as nucleosome positioning around the polyadenylation site, DNA methylation, and posttranslational histone modifications. Figure is taken from [14].

### 1.1.2.1 Trans-acting Regulator Proteins

Alterations in the levels or activity of core polyadenylation factors can change APA patterns. The choice of alternative poly(A) sites can be regulated by differential expression of general polyadenylation factors. For instance, in B cell differentiation, upregulation of Cstf64 leads to a switch from distal to proximal poly(A) site selection for IgM heavy chain, because CstF64 has a greater affinity for the distal GU-rich downstream element compared to proximal site [15]. This switch leads to conversion of IgM heavy chain from membrane-bound to secreted form, which is required for B cell activation.

Another 3'-processing factor influencing APA site choice is CFI (Cleavage Factor I). When CFI-25 levels were reduced by siRNA, poly(A) site selection for several genes shifted to distal polyadenylation sites. These results indicated that CFI may be recruited to the distal poly(A) site, possibly by sequence-specificity [16].

Besides 3'-processing factors, RNA binding proteins have been found to be associated with APA control. For instance, Nova2 is a tissue-specific factor which was originally characterized as a splicing factor [17]. Later on, it was shown that Nova2 also controlled

APA by enhancing distal poly(A) site usage. The position of Nova2 binding seems to be determining the result of poly(A) selection as it has an inhibitory effect for the formation of 3'-processing complex [18].

#### **1.1.2.2 *Cis*-acting RNA Elements**

Exact poly(A) cleavage site is dictated by particular RNA sequences in the pre-mRNA where various components of the 3'-processing complex bind. These sequences are defined as "core" polyadenylation elements, though there are other auxiliary downstream and upstream elements.

In the choice of APA sites, the consensus sequence AAUAAA is used more frequently than other variants. However, usage of other proximal hexamer variants is not rare [12, 19].

Deep sequencing analysis of 15 human tissues revealed a set of heptanucleotides showing high conservation in the region between 2 possible APA sites [20]. These heptanucleotides match to a consensus binding motif for FOX1/FOX2 tissue-specific splicing factors which suggested further roles for these proteins in terms of APA.

In a recent study, RNA sequencing analysis determined three new sequence patterns close to human poly(A) sites: a TTTTTTTT motif is positioned ~21 nt upstream of the poly(A) site, an AAWAAA motif (W denoting A or T) is positioned upstream of the poly(A) site, and a palindromic CCAGSCTGG sequence (S denoting C or G) is found downstream of the poly(A) site [21]. These new patterns are present in intragenic and intergenic poly(A) sites, however they don't co-exist with the canonical AATAAA signal.

Lastly, secondary structures and stem-loop motifs in 3'-UTRs have been shown to regulate stability of mRNA [22]. It is proposed that such structures could enhance or inhibit possible 3'-processing proteins to modulate APA [14].

#### **1.1.2.3 Chromatin and Epigenetic Regulation of APA**

Recent evidence suggests that 3'-end processing events might also be modulated by chromatin and histone modifications. An important association between epigenetic modifications and APA has been made. Strong nucleosome depletion around polyadenylation sites and nucleosome enrichment downstream of polyadenylation sites have been observed [23]. When genes possess multiple polyadenylation sites, higher downstream nucleosome affinity was associated with higher polyadenylation site usage.

Another study revealed the effect of RNA polymerase II modifications and chromatin architecture on 3'-end formation [24]. It appears the histone modifications at/near the polyadenylation sites are important.

### 1.1.3 Genome-wide Search for APA

To understand the regulation of gene expression, an accurate map of different APA sites and usage in diverse cell types is essential. Microarray-based approaches, deep sequencing, serial analysis of gene expression (SAGE), chromatin immunoprecipitation (Chip)-Seq, and RNA-seq (RNA sequencing) together with bioinformatics algorithms provide tools to explore APA site usage in different cell types and organisms.

Before genome-wide studies, several genes regulated by APA have already been identified. First examples of APA reported in 1980 were for *IgM* (Immunoglobulin M) and *DHFR* (Dihydrofolate reductase) genes [25-28]. Four APA isoforms were identified for *DHFR* which vary in 3'-UTR lengths with same protein product [28]. APA isoforms of *IgM* found to be encoding two distinct proteins; secreted or membrane-bound forms of *IgM* [25-27]. Later on, in 1997 various other genes have been implicated to go through APA such as Androgen Receptor, Collagenase 3, DNA Polymerase  $\beta$  [29]. Single gene studies also revealed Cyclin D1 [30], *Amy-1A* [31] and *NF-ATc* [32] to be regulated by APA.

First genome-wide bioinformatics study that focused on APA, searched expressed sequence tag (EST) data to screen for APA sites [33, 34]. Such global characterization of APA through bioinformatics analyses of the ESTs revealed more than half of the human genes and over 30% of the mouse genes to possess APA sites [12]. With wide usage of microarray-based technologies, genes with known APA sites were identified. In 2008, first study showed coordinated APA changes during T-cell activation [35]. Following studies showed neuronal activity dependent polyadenylation site selection by using microarray-based approaches [36], and APA during mouse embryonic development [37].

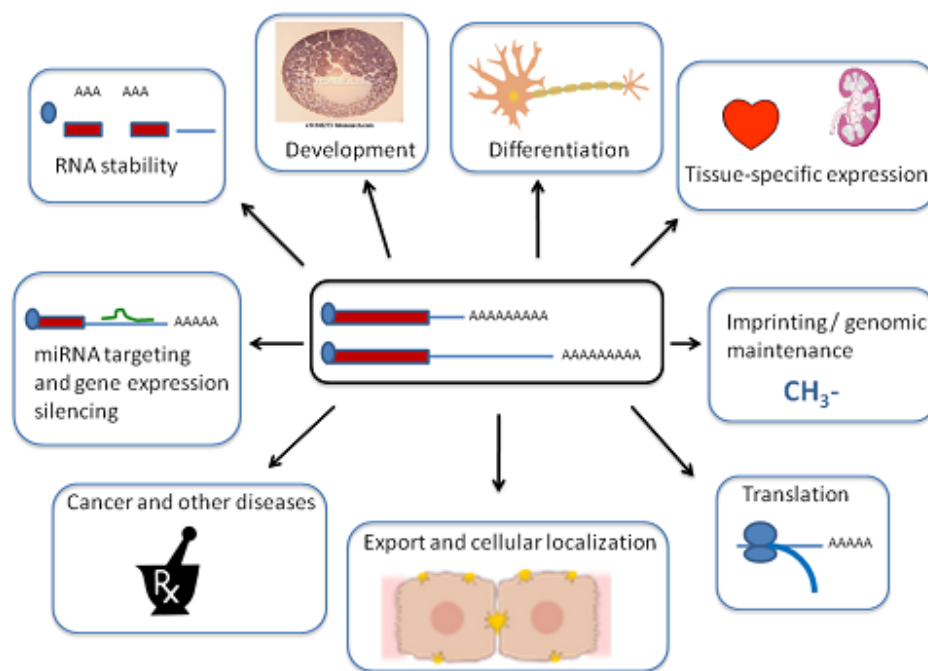
Microarray-based analyses have some limitations that only genes with known APA sites can be analyzed and quantification of mRNAs with more than two APA isoforms can be troublesome. Microarray data availability could be another drawback for less-studied species or physiological states.

Nevertheless, high-throughput sequencing-based techniques significantly increased the accumulation of data in terms of APA. Genome-wide APA events have been analyzed by using sequencing-based approaches; RNA polyadenylation profiles of yeast [21], *Caenorhabditis elegans* [19, 38, 39], *Drosophila melanogaster* [40], *Arabidopsis thaliana* [41, 42] have been investigated.

Sequencing methods may also have some disadvantages compared to EST collections [43]. Massively parallel signature sequencing digital gene expression (MPSS-DGE) and the Illumina sequencing-by-synthesis digital gene expression (SBS-DGE) platforms are used for revealing the role of APA in transcriptome dynamics. However, these platforms do not provide information on the precise locations of the poly(A) sites, which makes it difficult to determine the polyadenylation signals. Also, it is possible to detect the neighboring poly(A) sites by the same signatures, hence full complexity of APA can be masked. Moreover, MPSS-DGE and SBS-DGE data sets usually have short sequences which make them prone to sequencing errors.

### 1.1.4 Examples of APA

In recent years it has become evident that APA has a key role during gene expression regulation. Regulation of biological events like differentiation, transformation, proliferation, and developmental programs by APA has gained attention (Figure 1.3). The choice for a polyadenylation site determines the length of the 3'-UTR of an mRNA. Therefore, presence or absence of regulatory sequences in the 3'-UTR may eventually affect gene expression levels. Differences in 3'-UTR lengths can promote different stabilities and translational competencies. Longer 3'-UTR's harbor more *cis* regulatory elements which may be objected to RNA-binding proteins involved in mRNA stability, localization, and translation. Moreover, number of miRNA binding sites and AU-rich content is higher in longer 3'-UTR's [35, 37, 44]. Eventually, translation of an mRNA may be determined by the choice of polyadenylation signal.



**Figure 1.3. APA may affect different cellular events.** Various organismal and cellular events which have been shown to be affected by APA are summarized. Figure is taken from [13].

3'-UTR shortening events due to APA have been shown under various circumstances such as T cell proliferation [35], spermatocyte development [45], cancer cells [46], and ischemic response in myocardium [47].

In a pioneering study, a probe-level alternative transcript analysis was developed to assess APA during T cell activation [35]. APA association with T lymphocytes in a few genes was already shown [29, 32]. However global analysis showed increased expression of shorter 3'-UTR isoforms during proliferative states. A method for probe-level alternative transcript analysis (PLATA) was developed to analyze variations in probe hybridizations on Affymetrix Mouse Exon 1.0 ST microarrays. By using PLATA, transcripts in resting primary murine T cells expressing CD4<sup>+</sup> was compared with cells stimulated through TCR (T cell antigen receptor) for 6 or 48 hours. APA was found in both splicing-dependent and splicing-independent forms, however splicing-dependent APA events were found to be higher at both early (6h) and late (48h) stages while splicing-independent APA events were observed only at late stages of activation (48h). Analysis of microarray data from activated human T lymphocytes by PLATA showed decreased expression of longer 3'-UTR isoforms, revealing the pattern is conserved between mouse and human. After analysis of a panel of human and mouse tissues and cell lines, they found cell lines tend to have more proximal APA events. This phenomenon was explained by associating increased proliferation rates of cell lines with decreased expression of longer 3'-UTR isoforms. To confirm these findings at protein level, proximal and distal sized 3'-UTR's were cloned into luciferase reporter constructs. When luciferase activity was examined, full-length 3'-UTR yielded lower luciferase activity in stimulated T cells, indicating decreased translation of protein product. These findings together with miRNA predictions indicated that mRNA stability and/or translation was affected by post-transcriptional regulators located in distal 3'-UTR. These findings revealed global gene expression can be regulated by alternative polyadenylation, depending on the proliferative state of the cells.

After associating proliferating T cells with shorter 3'-UTR's, another study revealed Cyclin D1 having truncated 3'-UTR in mantle cell lymphoma (MCL) as a result of point mutations leading to premature poly(A) signals. Shorter 3'-UTR caused increased mRNA stability and associated with increased proliferation rates [48-50].

Another study revealed oncogenic activation through 3'-UTR shortening in cancer cells [46]. During normal cell proliferation and differentiation proto-oncogenes have regulatory roles; however in cancer cells these genes can be overactivated leading to deregulation of above mentioned processes. This activation could result from a gain of function mutation and/or overexpression. Within this perspective, when non-transformed and transformed cancer cell lines were compared, shorter 3'-UTRs were observed in cancer cell lines. 23 genes were analyzed by Northern blotting throughout the study, most of these genes were using proximal poly(A) sites in cancer cell lines. Following reporter assays, it was revealed that shorter mRNA isoforms produced more protein compared to longer isoforms. These results suggested that cancer-related shortening of 3'-UTR's could activate oncogenes.

On the contrary, 3'-UTR length increase was also shown in mouse neuronal cells [51], during embryonic development of mouse [37], in pluripotent stem cells [52], and in MDA-MB-231 breast cancer cells [53]. It appears 3'-UTR length variation control is highly organized upon different physiological states.

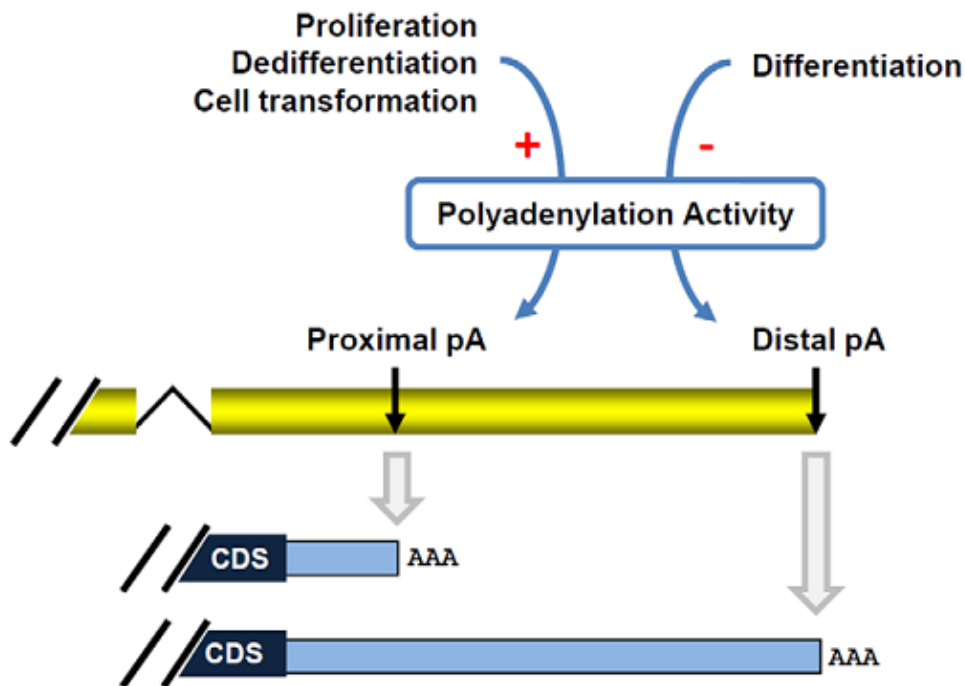
In pluripotent cells, for example, 3'-UTR lengths generally increase [52]. During generation of iPS (induced pluripotent cells) from different cell types, 3'-UTRs were shown to be reprogrammed to lengthen in size. This study revealed the dynamic nature of 3'-UTR



regulation in development, and showed that APA is an integral mechanism of cell reprogramming.

Another example of 3'-UTR lengthening event was revealed during mouse embryonic development [37]. In this study, after analyzing EST, SAGE, and microarray data results, a global lengthening of 3'-UTRs was found. As the number of *cis* regulatory elements increase with 3'-UTR lengthening, post-transcriptional regulation control via *cis* elements are predict to have important regulatory roles during embryonic development.

Taken together the findings of 3'-UTR shortening and lengthening events, a regulatory model was proposed (Figure 1.4). Based on this model, polyadenylation machinery is regulated dynamically in proliferation/differentiation. Hence, 3'-UTR length is controlled by the proliferative or differentiation states through APA.



**Figure 1.4. A model for regulation of APA events in differentiation/proliferation.** In proliferation, cell transformation, and dedifferentiation APA leads to usage of proximal poly(A) sites, while during differentiation APA leads to usage of distal poly(A) sites. Figure is taken from [52]. CDS: Coding sequence, pA: poly(A) site, AAA: poly(A) tail.

## 1.2 microRNAs and Cancer

3'-UTR isoform generation mostly has an impact on microRNA (miRNA) dependent gene expression control since 3'-UTRs dock miRNA binding sites and/or other regulatory sequences like AU-rich elements (AREs) [54, 55]. MiRNAs are ~22 nucleotides long, endogenous, small non-coding single-stranded RNAs, found in both animals and plants. miRNAs function as negative regulators of gene expression at posttranscriptional level. 30% of human protein-coding genes are estimated to be regulated by miRNAs [56]. miRNAs control the expression of genes involved in several biological processes including proliferation, differentiation, apoptosis, and development [56-58]. Most miRNA genes are located in intergenic regions but they can also be found within exonic or intronic regions in either sense or antisense orientation. Besides mRNA degradation and destabilization, the most common mechanism is translational repression as a result of miRNA binding to the 3'-UTR of an mRNA.

miRNAs can be organized as individual genes or localized as clusters representing miRNAs families, which are commonly related in sequence and function. miRNAs are mostly transcribed by RNA polymerase II (RNA pol II) from their own promoter or from promoter of the host gene in which they reside. RNA pol II synthesized large miRNA precursors called primary-miRNAs (pri-miRNAs) [59], which contain both 5'-cap structure (7MGpppG) as well as 3'-end poly(A) tail [60]. Clustered miRNAs might be transcribed from a single transcription unit as polycistronic primary-miRNA.

miRNA biogenesis can be summarized in two central processing steps. First, pri-miRNAs are processed in the nucleus by RNase III Drosha, associated to a double stranded RNA-binding protein DGCR8 (DiGeorge syndrome critical region gene 8; aka as Pasha in flies) known as the microprocessor complex, that generates ~70 nucleotides precursor miRNA products, which locally fold into stable secondary stem-loop structures [61].

Ran-GTP-dependent transporter Exportin 5 recognizes the short stem plus a ~2 nt 3' overhang of the originated precursor molecules, and mediates the translocation to the cytoplasm [62]. In the cytoplasm RNase III enzyme Dicer associated to TRBP (TAR RNA-binding protein) or protein activator of the interferon induced protein kinase (also known as PRKRA), and Argonaute (AGO1-4) perform "dicing" which is the second cropping process. By cleaving the miRNA precursor hairpin, a transitory miRNA/miRNA\* duplex is generated. This duplex contains the mature miRNA guide, generally selected according to thermodynamic properties, and the complementary passenger strand which is usually subjected to degradation.

Cancer is a disease involving multi-step changes in the genome [63]. Although current studies of the cancer genome mostly have focused on protein-coding genes, alterations of functional non-coding sequences in cancer is another field of interest to understand the cancer genome [64-66]. In human cancers, miRNAs can act as either oncogenes [67, 68] or tumor suppressor genes [49, 69, 70]. Over a decade, studies showed that expression of miRNAs is widely deregulated in cancer [64, 65]. While the number of miRNAs (~1000) is much smaller than that of protein-coding genes (~22000), expression signature of miRNAs

reflect the developmental lineage and tissue origin of human cancer more accurately [71, 72].

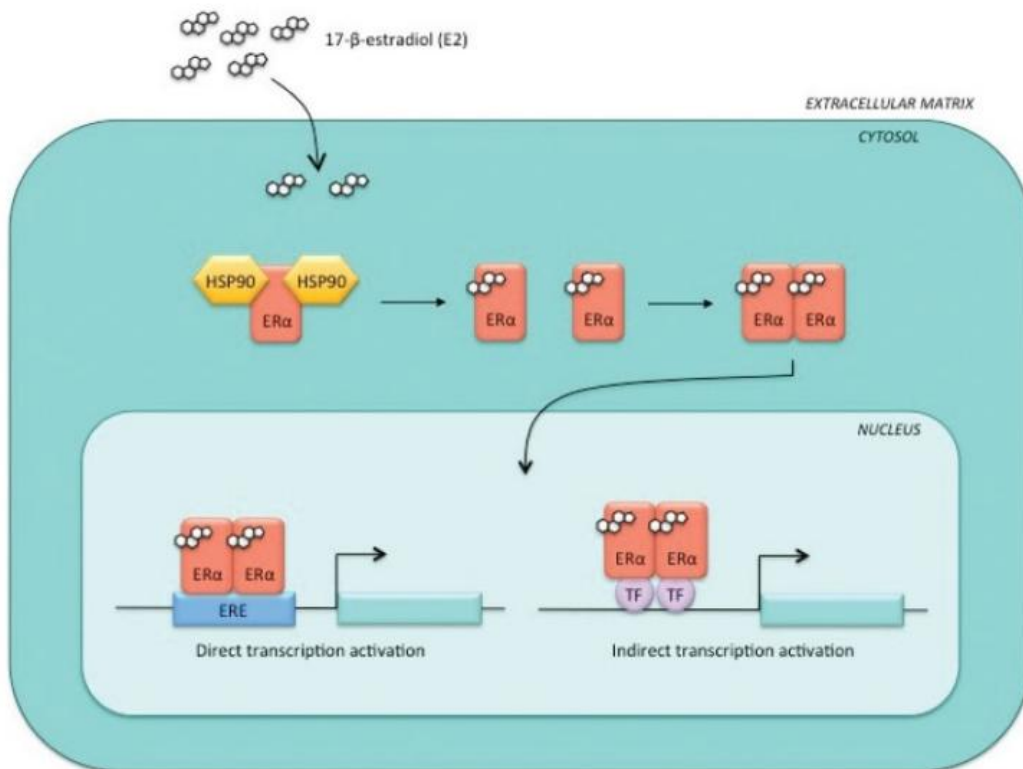
Nevertheless, the fundamental mechanisms of miRNA deregulation in human cancer mostly unknown. Several mechanisms have been proposed to explain miRNA deregulation in human cancers. First, transcriptional deregulation of miRNAs have been reported [67, 73]. DNA copy number abnormalities are also involved in miRNA deregulation [68, 69, 74]. Also, epigenetic alterations have been shown to effect miRNA expression in cancers [75, 76]. Another contributor to miRNA deregulation was found to be mutations [77]. Lastly, the key proteins playing critical roles in miRNA biogenesis pathway have been shown to be deregulated [78, 79] or may be dysfunctional [80]. Thus, transcriptional deregulation, DNA copy number abnormalities, epigenetic alterations, mutations and miRNA biogenesis machinery flaws contribute to miRNA deregulation in human cancer, either alone or more probably together.

### 1.3 Estrogen Receptor Signaling

Estrogen receptor (ER) is a member of the nuclear hormone receptor family. Two subtypes of ER have been identified as ER $\alpha$  and ER $\beta$  which have common functional domains [81], but are differentially expressed in various tissues and have unique functions [82, 83].

The regulation of gene expression by ER $\alpha$  is a multifactorial process which involves genomic and non-genomic actions. These actions may converge at certain response elements located in the promoters or target genes. 17- $\beta$ -estradiol (E2) activates ER to act as a transcription factor [84]. When E2 diffuses through cell membrane, it can bind to ER $\alpha$  which then dimerizes (Figure 1.5). These ER $\alpha$  homodimers translocate into the nucleus and bind directly to the estrogen response elements (ERE) of ER target genes to initiate gene transcription [85]. ER $\alpha$  might also form a complex with other transcription factors such as SP-1 and AP-1, so that transcription of genes without an ERE can be activated [86]. Yet another way of transcriptional activation is possible through non-genomic mechanisms of action; through modulation of signaling cascades [87, 88]. These three mechanisms of ER action enable a broader range of genes to be regulated.

ER $\alpha$  mediates the proliferative action of E2 in breast cancer cells. E2 plays an important role in the progression of breast cancer as most of the human breast cancers emerge as E2 dependent and express ER $\alpha$ . Endocrine therapies by Tamoxifen [89], a selective ER modulator, and/or aromatase inhibitors, prevent peripheral estrogen synthesis, have been revealed to increase disease-free survival [90]. However, initial or acquired resistance to endocrine therapies often arises with metastatic tumors [91]. Hence, better understanding of the action mechanisms of E2 and ER $\alpha$  signaling in breast tumorigenesis through identification of direct and indirect estrogen target genes may provide new avenues for effective therapies.



**Figure 1.5. Estrogenic Response.** Upon the binding of E2 to ER $\alpha$ , it dissociates from heat shock proteins (HSP90) and forms a homodimer. After translocation to the nucleus, ER $\alpha$  homodimers can bind to ERE and initiate transcription. ER $\alpha$  homodimers may also initiate transcription through interactions with other transcription factors. Figure is taken from [92].

#### 1.4 Aim of the Study

3'-UTR shortening of mRNAs through APA has important consequences for gene expression. Proliferating cells avoid miRNA-mediated repression by using proximal APA sites and switching to shorter 3'-UTRs. We think, 3'-UTR shortening events may enlighten the basis of some of the proto-oncogene activation cases observed in cancer cells. Therefore, in this study, we investigated whether estradiol (E2), a potent proliferation signal for breast cells; induces APA and 3'-UTR shortening to activate proto-oncogenes in estrogen receptor positive (ER+) breast cancers.

## CHAPTER 2

### MATERIALS AND METHODS

#### 2.1 Probe Level Screen for APA Transcripts

NCBI Gene Expression Omnibus (GEO) [93] is a functional genomics data repository. Three experiment sets GSE11324 [94], GSE8597 [95], and GSE11791 [96] were selected from GEO. All 3 data sets were for E2 treated MCF7 cells and the arrays were Affymetrix Human Genome U133 Plus 2.0 Arrays (HGU133Plus2, NCBI GEO Accession number: GPL570). CEL files for the corresponding experiments were downloaded to use the non-normalized raw intensities of the probes.

Probe level screen algorithm APADetect was developed by Dr. T. Can (Computer Engineering Department, Middle East Technical University, Ankara). Affymetrix chip annotation files were used to map Unigene identifiers to probe set identifiers. Data sets were analyzed for probe level differences based on the positions of reported [97] polyA sites according to the UCSC Genome Browser Database [98] to detect alternative polyadenylation (APA) dependent events in the 3'-UTRs. Proximal and distal probe sets were recorded based on the positions of multiple poly(A) sites that were reported for 10550 genes. After analysis, 4444 polyA sites were found which split 3067 probe sets into two subsets on the HG U133 Plus 2.0 chip. For each such case, the fold difference was computed between the proximal and distal probe sets of E2 treated and control datasets. All GEO series were analyzed individually and their results were combined to screen for transcripts that consistently had differently expressed split probe sets.

#### 2.2 Cell Lines and Treatments

MCF7, T47D, MDA-MB-231 cell lines were a kind gift from Dr. U. H. Tazebay (Bilkent University, Ankara). MCF7, MDA-MB-231 cells were grown in DMEM with Earle's salts and 10% FBS, T47D cells were grown in RPMI, with 10% FBS and 0.1% non-essential amino acids. Media contained 1% Penicillin/Streptomycin (P/S). MDA-66 cell lines were a kind gift

from Dr. U. H. Tazebay with permission from Dr. F. Gannon [99, 100]. MDA-66 cells were maintained in MDA-MB-231 medium with 0.4 mg/mL Hygromycin (Roche Applied Sciences). For MCF7 ER $\alpha$  silencing studies, MCF7\_EV (Empty pSR vector transfected), MCF7\_CO (Control shRNA transfected) [101], and MCF7\_shER $\alpha$  (ER shRNA transfected clone) [99] cells were used. Stable cell lines were maintained with the addition of 0.25 mg/mL Geneticin (Roche Applied Sciences). Cell lines were grown as monolayers and were incubated at 37°C with 95% humidified air and 5% CO<sub>2</sub>.

For hormone induction experiments, cells were grown in phenol red-free medium supplemented with 10% charcoal stripped FBS, 1% P/S and 1% L-glutamine for 48 hours. MCF7 cells were either treated with 10 nM 17  $\beta$ -Estradiol (E2) (Sigma-Aldrich) or ethanol (vehicle control) for 3 and 12 hours, T47D cells were treated with 10 nM E2 (Sigma-Aldrich) or ethanol for 12 and 18 hours [102-104]. MDA-MB-231 cells were treated with 10 nM E2 (Sigma-Aldrich) or ethanol (vehicle control) for 3, 12 and 18 hours [102-104]. 10  $\mu$ g/mL Cycloheximide (Sigma) [95] and 1  $\mu$ M ICI 182,780 (Tocris Biosciences) [105] pre-treatments were done 1 hour before E2 treatment.

## 2.3 Expression Analysis

MIQE Guidelines [106] and checklist (Appendix B) were followed throughout the RNA isolation, quantification, cDNA synthesis and RT-qPCR analysis.

### 2.3.1 RNA Isolation

Total RNA was isolated using High Pure RNA Isolation Kit (Roche Applied Science). RNA isolations were performed according to the manufacturer's instructions.

### 2.3.2 DNA contamination, Quantification and Integrity Assessment

For DNase treatment, total RNA was incubated with DNase I (Roche Applied Science) at 37°C for 1 hour followed by phenol-chloroform extraction (reaction conditions given in Table 2.1).

After DNase treatment, lack of DNA contamination was confirmed by PCR using *GAPDH* (Glyceraldehyde 3-phosphate dehydrogenase) specific primers. *GAPDH\_F*: 5'-GGGAGCCAAAAGGGTCATCA-3' and *GAPDH\_R*: 5'-TTTCTAGACGGCAGGTCAGGT-3' (product size: 409 bp). Following conditions were used for the PCR reactions: incubation at 94°C for 10 minutes, 40 cycles of 94°C for 30 seconds, 56°C for 30 seconds, and 72°C for 30 seconds, and final extension at 72°C for 5 minutes. MCF7 cDNA was used as a positive control

**Table 2.1.** Reaction Conditions of DNase Treatment

<b>Components</b>	<b>Amount</b>
Total RNA	10 µg
10X Incubation Buffer (400 mM Tris-HCl, 100 mM NaCl, 60 mM MgCl <sub>2</sub> , 10 mM CaCl <sub>2</sub> , pH 7.9)	10 µL
DNase I recombinant, RNase-free (10 units/µl) Roche Applied Sciences Cat. No. 04 716 728 001	1 µL
RiboLock™ RNase Inhibitor (20u/µL) Fermentas #EO0381 *One unit of RiboLock™ RNase Inhibitor inhibits the activity of 5 ng RNase A by 50%	1 µL
RNase-free Water	To 100 µL
DNase treatment was done at 37° C for 60 min. Reaction was terminated by phenol/chloroform extraction.	

After confirming absence of DNA, quantification of RNA samples was done using NanoDrop ND1000 (Thermo Scientific) spectrophotometrically (Appendix C). Purity of the RNA samples was determined by A260/A280 and A260/A230 ratios. Higher or lower A260/A280 ratios indicates whether the sample is contaminated by proteins or a reagent such as phenol, likewise A260/A230 ratio could be an indicator of a problem with the extraction procedure. An exemplificatory result was shown in Appendix C showing a proper RNA sample, with A260/A280 ratio 2.05 and A260/A230 ratio 2.13. All RNA samples used in this study were fitting the criteria in terms of purity.

### 2.3.3 cDNA synthesis

cDNAs (20 µl) were synthesized using the RevertAid First Strand cDNA Synthesis Kit (Fermentas) using 2 µg total RNA and oligo\_dT primers (reaction conditions given in Table 2.2).

**Table 2.2.** Reaction Conditions of Reverse Transcription

Components	Amount
Total RNA	2 µg
Oligo(dT) <sub>18</sub> primer (100 µM 0.5 µg/µl (15 A260 u/ml) )	1 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
10 mM dNTP mix	2 µL
RiboLock™ RNase Inhibitor (20u/µL) *One unit of RiboLock™ RNase Inhibitor inhibits the activity of 5 ng RNase A by 50%	1 µL
RevertAid™ M-MuLV Reverse Transcriptase (200u/µL) *One unit of RevertAid™ M-MuLV RT incorporates 1 nmol of dTMP into a polynucleotide fraction (adsorbed on DE-81) in 10 min at 37°C	1 µL
Nuclease-free water	to 20 µL
cDNA was synthesized at 42° C for 60 minutes. Reaction was terminated by keeping at 70° C for 5 minutes. cDNA samples were stored at -20° C.	

#### 2.3.4 TFF1 Expression Analysis

*TFF1* (trefoil factor 1, aka *pS2*) is a known E2 upregulated gene [99]. *TFF1* expression was determined by RT-PCR as a positive control for the E2 treatment. Following primers were used; TFF1\_F: 5'-CCATGGAGAACAAGGTGATCTGC-3' and TFF1\_R: 5'-GTCAATCTGTGTTGTGAGCCGAG-3' (product size: 399 bp). Incubation at 94°C for 10 minutes, 30 cycles of 94°C for 30 seconds, 56°C for 30 seconds, and 72°C for 30 seconds, and final extension at 72°C for 5 minutes.

#### 2.3.5 RT-qPCR Analysis

For RT-qPCR, SYBR® Green Mastermix (Roche Applied Science) was used using the Rotor Gene 6000 (Corbett, Qiagen) cyclers. 20 µL reactions were performed with 300 nM of specific primer pairs. The *CDC6* (NM\_001254) short and long 3' UTR isoforms were amplified using the following primer sets; *CDC6* 3' UTR- short (product size size: 185 bp) *CDC6\_F*:5'-TTCAGCTGGCATTAGAGAGC-3', *CDC6\_R1*: 5'-AAGGGTCTACCTGGTCACTTTT-3', *CDC6* 3' UTR-long (product size size: 349 bp) *CDC6\_F*:5'-TTCAGCTGGCATTAGAGAGC-



3', *CDC6\_R2*: 5'-CGCCTCAAAAACAACAACAA-3'. The fold change for the isoforms was normalized against the reference gene; *SDHA* (NM\_004168) [107] amplified with the following primer sets: *SDHA* (product size: 86 bp) *SDHA\_F*: 5'-TGGGAACAAGAGGGCATCTG-3', and *SDHA\_R*: 5'-CCACCACTGCATCAATTCATG-3'. Reference gene, *SDHA*, expression is consistent across breast cancer cell lines [107] and did not change in response to E2 treatment. For all three reactions, following conditions were used: incubation at 94°C for 10 minutes, 40 cycles of 94°C for 15 seconds, 56°C for 30 seconds, and 72°C for 30 seconds.

For the relative quantification, the reaction efficiency incorporated  $\Delta\Delta C_q$  formula was used [108]. Three independent biological replicates with 3 technical replicas per experiment were used for each PCR. One-way ANOVA with Tukey's multiple comparison post test was performed using GraphPad Prism (California, USA).

*PCNA* (proliferating cell nuclear antigen) expression detected by RT-qPCR after ICI 182, 789 (ICI) and cycloheximide (CHX) treatments. *PCNA* (NM\_002592) was amplified using the following primer set; *PCNA\_F*: 5'-TGCAGATGTACCCCTTGTTG-3', *PCNA\_R*: 5'-GCTGGCATCTTAGAAGCAGTT-3'. Following conditions were used: incubation at 94°C for 10 minutes, 40 cycles of 94°C for 15 seconds, 56°C for 30 seconds, and 72°C for 30 seconds.

*E2F1*, *E2F2* and 3'-UTR processing gene transcripts *CSTF2* (cleavage stimulation factor, 3' pre-RNA, subunit 2), *CSTF3* (cleavage stimulation factor, 3' pre-RNA, subunit 3) and *CPSF2* (cleavage and polyadenylation specific factor 2) primer sequences were taken from the literature [109, 110]. *E2F1\_F*: 5'-CCATCCAGGAAAAGGTGTGA-3', *E2F1\_R*: 5'-TTCCTGGAGCTGCTGAGC-3' (product size 94 bp), *E2F2\_F*: 5'-ACTGGAAGTGCCCGACAGGA-3' , *E2F2\_R*: 5'-GGTGGAGGTAGAGGGGAGAGG-3' (product size 142 bp), *CSTF2\_F*: 5'-ATCCTTGCCTGCGAATGTCC-3', *CSTF2\_R*: 5'-GGGTGGTCCTCGGCTCTC-3' (product size 178 bp), *CSTF3\_F*: 5'-TGAAGTGGATAGAAAACCAGAATACCC-3' *CSTF3\_R*: 5'-TGAAACAGATAGGAGGAGGGAGA-3' (product size 171 bp), *CPSF2\_F*: 5'-CTTTGGAACCTTGCCACCT-3' , *CPSF2\_R*: 5'-TGCGGACTGCTACTTGATTGTTG-3' (product size 159 bp). Following conditions were used: incubation at 94°C for 10 minutes, 40 cycles of 94°C for 15 seconds, 56°C for 30 seconds, and 72°C for 30 seconds.

## 2.4 Rapid Amplification of cDNA ends

RACE specific cDNA synthesis was performed using the 3'-RACE Kit (Roche) with 2  $\mu$ g total RNA (DNase treated) from E2 treated MCF7 cells using the oligo dT-anchor primer (5'-GACCACGCGTATCGATGTCGACTTTTTTTTTTTTTTTTTV-3'). For PCR, a reverse primer for the anchor sequence was used (*Anchor\_R*: 5'-GACCACGCGTATCGATGTCGAC-3'). Gene Specific Primers (GSP's) *3'RACE\_1*: 5'-GCTCTTGAAGCCAGGGGCATTTTA-3', and *3'RACE\_2*: 5'-CCACCCGAAAGTATTCAGCTGGCATTTA-3' were designed and used as forward primers.

First round 3'-RACE PCR was done using the 3'RACE\_1 and Anchor-R primers with the following PCR conditions; 94°C for 10 minutes, 20 cycles of 94°C for 30 seconds, 64°C for 30 seconds, 72°C for 30 seconds. Nested 3'RACE PCR was performed using the 3'RACE\_2 and Anchor-R primers with the 1/10 diluted PCR product as template with the following PCR conditions; incubation at 94°C for 10 minutes, 35 cycles of 94°C for 30 seconds, 64°C for 30 seconds, 72°C for 30 seconds. Following gel extraction of correct size bands with High Pure Gel Extraction Kit (Roche), another round of 3'-RACE PCR was performed with 2 ng of PCR product (gel purified) as the template using 3'RACE\_2 and Anchor-R primers with the following PCR conditions; incubation at 94°C for 10 minutes, 35 cycles of 94°C for 30 seconds, 64°C for 30 seconds, 72°C for 30 seconds.

## **2.5 Western Blotting**

### **2.5.1 Nuclear Protein Isolation**

Nuclear protein extraction procedure [111] was used to investigate CDC6 protein levels. Cells were washed with PBS (phosphate buffered saline) twice. Then, cells were collected via a scraper into 1 mL ice-cold PBS. After 5 minute centrifugation (1500 rpm) supernatant was discarded and cells were resuspended in 500 uL extraction Buffer A (Appendix E) and kept on ice for 30 minutes. After incubation, tubes were vortexed for 30 seconds, harshly. 10% NP40 was added in tubes to a final concentration of 0.15%. After waiting 5 minutes on ice, tubes were vortexed again for 5 seconds. Nuclei were pelleted by centrifuging at 12 000g for 1 minute. At this point, supernatant which contains cytoplasmic proteins were taken into a fresh tube. Nuclei were washed with 500 uL Buffer A three times. After washing and removing the buffer by centrifugation, 100 uL Buffer B (Appendix E) was added on nuclei. Lysates were kept on ice for 30 minutes, and then centrifuged (12 000 g) for 10 minutes. Nuclear proteins in supernatant were taken into fresh tubes and stored at -80° C. BCA Protein Assay Kit (Pierce) was used for lysate quantification.

### **2.5.2 Total Protein Isolation**

Total protein extraction was done from cell lines by RIPA buffer (Appendix E). A tablet cocktail of protease inhibitors (Roche) dissolved in 10 mL RIPA was used. Cells were washed with cold PBS twice. RIPA was added on monolayer cells (for T75 flasks 500-1000 uL), swirled on shaker for 10 minutes. Cells were collected via scraper into a microcentrifuge tube. Then lysates were centrifuged at 14000g for 20 min. Supernatants were transferred to a new tube and stored at -80° C.

### 2.5.3 Western Blotting

Nuclear extracts (50 µg) were denatured in 6X Laemmli buffer (Appendix E) at 100°C for 5 min and were separated on a 8% polyacrylamide gel and transferred onto a nitrocellulose membrane (Bio-Rad). The membranes were blocked in 5% non-fat milk in TBS-T (Tris Buffer Saline- Tween) (Appendix E) and were incubated overnight with the monoclonal anti-CDC6 rabbit antibody (1:1000 dilution; Cell Signaling); followed by a 1 hour incubation with the HRP-conjugated secondary anti-rabbit antibody (1:2000 dilution, Cell Signaling). Proteins were visualized using an enhanced chemiluminescence kit (ECL Plus; Pierce) according to the manufacturer's instructions. For nuclear protein loading control, Histone H3 (FL-136) antibody, a gift from Dr. S. Banerjee (METU, Ankara) (Santa Cruz Biotechnology), was used. 5% non-fat milk in PBS-T (Phosphate Buffer Saline- Tween) (Appendix E) was used for blocking. Primary antibody was 1:500 diluted and secondary anti-rabbit (Santa Cruz Biotechnology) antibody was 1:2000 diluted.

Total protein extracts (50 µg) were denatured in 6X Laemmli buffer at 100°C for 5 min and were separated on a 8% polyacrylamide gel and transferred onto a nitrocellulose membrane (Bio-Rad). The membranes were blocked in 5% non-fat milk in TBS-T, and were incubated overnight with the monoclonal anti- ER $\alpha$  antibody (F-10) (1:200 dilution; sc-8002, Santa Cruz Biotechnology) followed by a 1 hour incubation with the HRP-conjugated secondary anti-mouse antibody (1:2000 dilution Santa Cruz Biotechnology).  $\beta$ -actin antibody (sc-47778, Santa Cruz Biotechnology) was used for protein loading control. For blocking 5% BSA in TBS-T was used. Primary antibody was 1:1000 diluted and secondary anti-mouse antibody was 1:2000 diluted.

## 2.6 BrdU (Bromodeoxyuridine) Incorporation

For BrdU incorporation assay,  $1 \times 10^4$  MCF7, T47D and MDA-66 cells were seeded and grown in black 96-well plates (Canberra Packard) in 100 µl of DMEM-10% FBS or RPMI-10% FBS, respectively. A day later, media were changed to 10% charcoal stripped FBS-medium and cells were incubated for 48 hours. BrdU incorporation together with E2 treatment was pursued for 0, 3, 12, 24 hours for MCF7; 0, 12, 18, and 24 hours for T47D; 0, 3, 12, 18 hours for MDA-66. BrdU incorporation was detected according to the protocol from Cell Proliferation ELISA, BrdU assay kit (Roche Applied Science) using the Turner Biosystems Luminometer. Readings are presented as relative light units per second (RLU/sec) at respective time points. Two independent assays were performed with 5 replicates per experiment.

## 2.7 microRNA Target Sites on mRNAs

3 different miRNA-mRNA interaction prediction tools were used; TargetScan (<http://www.targetscan.org/>), PicTar (<http://pictar.mdc-berlin.de/>) and FindTar3 (<http://bio.sz.tsinghua.edu.cn/>).

## 2.8 Cloning of 3'-UTR Isoforms

Short (341 bp) and long (915 bp) 3'-UTR fragments of *CDC6* were PCR amplified with cloning primers using a high fidelity polymerase (Fermentas *Pfu* polymerase) and cloned into pMIR-Report vector. SacI and HindIII recognition site harboring cloning primers were as follows; (restriction sites are in bold letters and recognition aiding sites are underlined).

CDC6\_pMIR\_F: 5'- **CGAGCTCG**ATTCTTCTTACACCCAC-3'

CDC6\_pMIR\_R1: 5'- **CCCAAGCTTGGG**AAAATACCCACTCATGTTTGAG-3'

CDC6\_pMIR\_R2: 5'- **CCCAAGCTTGGG**AACTTGAAAATAAATATATTC-3'

Following PCR conditions were used for the short isoform: incubation at 95°C for 10 minutes, 35 cycles of 95°C for 30 seconds, 60°C for 30 seconds, and 72°C for 1 minute, and final extension at 72°C for 10 minutes. Following PCR conditions were used for the long isoform: incubation at 95°C for 10 minutes, 35 cycles of 95°C for 30 seconds, 60-62-64°C for 30 seconds, and 72°C for 2 minute, and final extension at 72°C for 10 minutes.

After PCR amplification of short and long isoforms, PCR products were loaded on 1% Agarose gel, DNA bands were visualized and documented under UV light. Size-confirmed DNA fragments were cut from the agarose gel and isolated from agarose gels by using Agarose Gel Extraction Kit (Roche) according to the manufacturers' instructions. Isolated DNA fragments were quantified using NanoDrop. Purified PCR products and pMIR-Report (Ambion) empty vector were double digested with SacI and HindIII enzymes overnight. After running on 1% Agarose gel and purification from gel DNA's were quantified by NanoDrop again.

To ligate PCR products into empty vector, purified vector (100 ng) and PCR products were ligated at 16°C for 16h using T4 DNA ligase (Roche). After ligation, 2 uL of the ligation reaction was directly transformed into competent TOP10 *E.coli* cells.

The colonies obtained from bacterial transformation were confirmed by PCR. pMIR\_F: 5'-AGGCGATTAAGTTGGGTA-3' and pMIR\_R:5'-GGAAAGTCCAAATTGCTC-3' were vector (pMIR-Report) specific primers. Following conditions were used: incubation at 94°C for 10 minutes, 35 cycles of 94°C for 30 seconds, 56°C for 30 seconds, and 72°C for 30 seconds, and final extension at 72°C for 10 minutes. Based on colony PCR results, 2 colonies (one for short and one for long isoforms) were sent to sequencing.

### 2.8.1 Dual-luciferase Assay

After sequence confirmation, MCF7 cells were co-transfected with pMIR-Report (Firefly luciferase) (375 ng) and phRL-TK (Renilla Luciferase) (125 ng) in 24-well plates using 0.75  $\mu$ L Fugene-HD (Roche). 24 hours after transfection, cells were collected and lysed. Dual-Luciferase® Reporter Assay System (Promega) was used according to the manufacturer's instructions. Dual-luciferase activities were measured using the Modulus Microplate Luminometer (Turner Biosystems). The luminescence intensity of Firefly luciferase (pMIR) was normalized to that of Renilla luciferase (phRL-TK). Experiments were repeated 3 independent times with 4 replicates per experiment. For E2 treatments, MCF7 cells were kept in phenol red-free MEM supplemented with 10% charcoal stripped FBS, 1% P/S and 1% L-glutamine for 48 hours. MCF7 cells were then co-transfected with pMIR and phRL-TK constructs as described above. 12 hours after transfection, cells were treated with 10 nM E2 for 12 hours. E2 responsive TFF1-ERE in pGL3 [99] was used as an E2 treatment control. Experiments were repeated 2 independent times with 3 replicates per experiment.



## CHAPTER 3

### RESULTS AND DISCUSSION

#### 3.1 CDC6 3'-UTR and Probe Level Screen

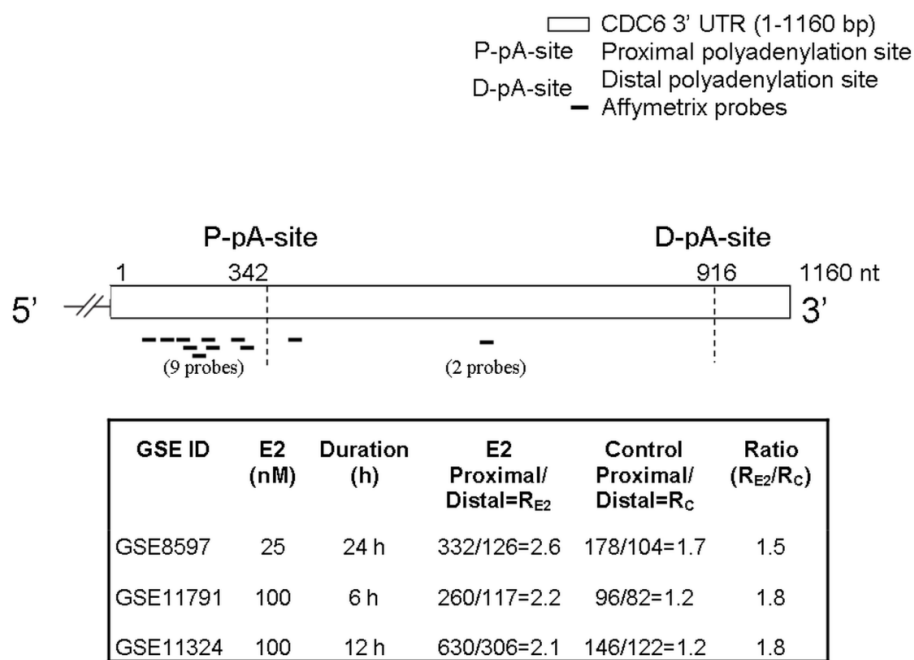
APA and 3'-UTR shortening occurs in highly proliferative cells [35]. E2 is known to increase proliferation rate of breast cancer cells [112, 113]. Therefore, to investigate if E2 induces APA and 3'-UTR shortening in ER positive breast cancer cells, initially an *in silico* approach was used. To identify transcripts regulated by APA, APADetect program (<http://bioserver.ceng.metu.edu.tr/APADetect/>) developed by Dr. Tolga Can was used. APADetect developed for this purpose was unique in the sense that probe level differences of microarray data can be analyzed together with the integrated poly(A) sites.

In the analysis, transcript specific individual probes are grouped into proximal and distal sets based on the positions of the reported 3'-UTR poly(A) sites [97]. APADetect investigates the difference between mean intensities of probes matching to distal and proximal 3'-UTR's. To do this, raw intensities of each probe from CEL file (stores the results of the intensity calculations from microarray) are used to calculate average intensities of distal and proximal matching probes.

To investigate if E2 induced APA, we used APADetect and analyzed already available microarray data sets for E2 treated MCF7 breast cancer cells along with controls (Ethanol treated vehicle control). For this purpose, GSE8597, GSE11791, and GSE11324 gene expression datasets were selected from Gene Expression Omnibus (GEO) database [93] and CEL files were incorporated into APADetect. These three microarray data were obtained after treating MCF7 cells treated with different E2 concentrations (25 or 100 nM) and for different time points (6, 12, and 24 hours) (Figure 3.1). The average proximal to distal probe signal intensity values in E2 treated cells were compared to that of ethanol treated (control) cells. Using this approach, proximal/distal probe set ratio differences in response to E2 were detected.

According to APADetect, three independent gene expression datasets GSE8597, GSE11791, and GSE11324 produced a common APA candidate; *CDC6* (Cell Division Cycle 6), regulator of the DNA pre-replication complex.

*CDC6* 3'-UTR harbors 2 APA sites; as can be seen in Figure 3.1. Proximal poly(A) (P-pA) site is located at the 342<sup>nd</sup> bp and distal poly(A) (D-pA) site is located at the 916<sup>th</sup> bp of the 3'-UTR. Affymetrix probe set for *CDC6* transcript consists of 11 probes. When these probes were aligned to the *CDC6* 3'-UTR, 9 of them mapped to proximal 3'-UTR and 2 probes were mapped on distal 3'-UTR, after the APA site. Raw probe intensities and means of these distal and proximal probe intensities of *CDC6* for 3 independent data sets were listed in Table 3.1. All three arrays (GSE8597, GSE11791, and GSE11324) were performed in three replicates for E2 and EtOH treatments, hence experimental triplicates were analyzed individually through APADetect and mean values and standard deviations were calculated for raw intensities.



**Figure 3.1. *CDC6* 3'-UTR and the positions of the probes.** *CDC6* 3'-UTR (1160 nucleotides) harbor two reported polyadenylation sites: the proximal poly(A) (P-pA) site and distal poly(A) (D-pA) site. Nine of the probes matched to the upstream site of P-pA-site and two of the probes matched to the region between P-pA and D-pA-sites. GSE8597, GSE11791, and GSE11324 were three independent gene expression data sets which were analyzed at the probe level. All three sets were analysis of MCF7 cells treated with E2 (treated) and EtOH (vehicle control). The mean of nine proximal and two distal probe set signals were collected. Ratio of proximal to the distal signals in E2-treated ( $R_{E2}$ ) cells increased at least 1.5X fold compared with the ratios of the controls ( $R_C$ ).



**Table 3.1.** Raw intensities of 11 probes for *CDC6* from three gene expression CEL files.

		1	2	3	4	5	6	7	8	9	10	11
GSE8597	E2	307.25 ±36.57	759.0 ±56.85	330.5 ±51.58	168.25 ±27.37	163.5 ±13.88	407.5 ±44.21	283.25 ±14.77	302.75 ±57.08	269.25 ±73.56	167.5 ±16.99	84.25 ±5.76
	Mean= 332										Mean= 126	
	EtOH	175.50 ±27.54	324.75 ±49.28	197.25 ±50.93	104.0 ±23.01	104.25 ±11.28	229.5 ±55.83	149.75 ±27.54	158.7 ±37.06	160.25 ±73.61	113.25 ±22.28	95.5 ±22.01
	Mean= 178										Mean= 104	
GSE11791	E2	353.33 ±95.77	484.0 ±166.23	265.66 ±72.92	163.0 ±37.85	103.0 ±19.13	368.33 ±101.75	184.33 ±32.29	205.66 ±55.05	217.0 ±27.43	154.66 ±45.76	78.33 ±10.96
	Mean= 260										Mean= 117	
	EtOH	95.66 ±21.79	132.33 ±31.58	116.0 ±49.52	80.66 ±9.67	65.0 ±6.38	127.33 ±29.41	83.0 ±8.64	74.66 ±7.93	91.66 ±13.42	77.0 ±20.46	86.66 ±11.03
	Mean= 96										Mean= 82	
GSE11324	E2	720.0 ±72.68	1234.33 ±34.18	693.66 ±93.66	270.33 ±42.76	367.0 ±42.93	660.33 ±104.89	550.66 ±33.69	799.33 ±158.47	370.33 ±23.47	355.33 ±39.85	257.0 ±20.93
	Mean= 630										Mean= 306	
	EtOH	166.66 ±23.70	237.66 ±26.40	150.0 ±14.99	83.0 ±2.16	107.33 ±6.55	148.66 ±17.21	132.0 ±17.91	164.33 ±18.62	126.33 ±11.26	123.66 ±7.54	119.66 ±6.94
	Mean= 146										Mean= 122	

After calculating the mean of distal and proximal matching probes for E2 and control treated samples (Table 3.1),  $R_{E2}$  and  $R_c$  ratios were calculated as shown in Figure 3.1. Proximal and distal probe means for E2 treated samples ( $R_{E2}$ ) were divided to that of control samples ( $R_c$ ).  $R_{E2}/R_c$  ratio of the three datasets for *CDC6* was above 1.5.

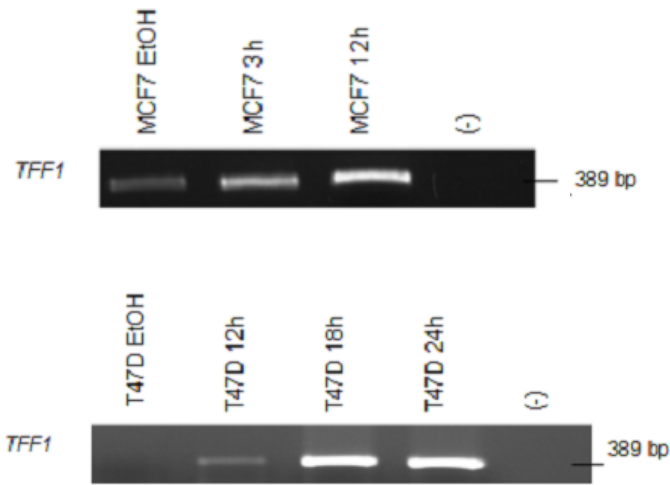
Even with different concentrations of E2, *CDC6* came out as a candidate whose 3'-UTR was predicted to be shortened. Next, *CDC6* was experimentally investigated for possible APA and generation of a shorter 3'-UTR isoform in response to E2 to confirm the *in silico* results.

## 3.2 Expression Analysis

### 3.2.1 E2 Treatment of ER+ Cells

To investigate expression levels and confirm APA events in *CDC6*, ER+ breast cancer cell lines; MCF7 and T47D were treated with 10 nM E2 after growing cells in serum-deprived medium for 48 hours. Although array results were from MCF7, T47D cells were also included in the experimental set-up so as to avoid any possible cell-specific response. Based on previous studies and doubling times of cell lines, MCF7 cells were treated for 3 and 12 hours, while T47D cells were treated for 12 and 18 hours [114-116].

*TFF1* is a known E2 target gene [99]. The effect of E2 treatment was confirmed with the upregulated expression of *TFF1* in E2 treated cells (Figure 3.2). In ER positive MCF7, and T47D cells time-dependent increase of *TFF1* levels was observed.



**Figure 3.2. *TFF1* is an E2 upregulated gene.** MCF7 and T47D cells were treated with ethanol or 10 nM E2 for indicated time points E2 treatment resulted with the upregulation of *TFF1* expression detected by RT-PCR.

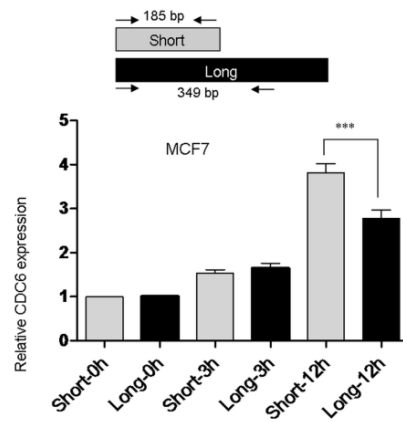
Given that the E2 treatment worked in MCF7 and T47D cells, we proceeded with RT-qPCR quantification of *CDC6* short and long 3'-UTRs.

### 3.2.2 RT-qPCR Analysis

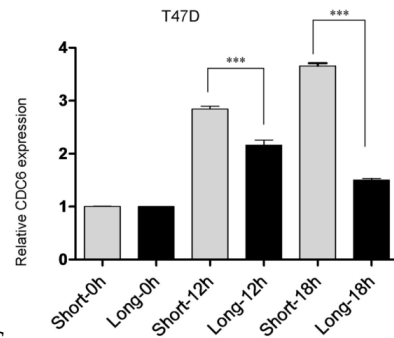
Following MIQE guidelines [117], RT-qPCR conditions were optimized to stabilize reaction efficiency of each primer couple to  $\sim 1$ . Additionally, each RT-qPCR reaction melt analysis was evaluated for primer dimerization or other non-specific product formation. RT-qPCR assay performance results are shown in Appendix D.

Differentiation of expression of short isoform from long isoform is not possible as the primers designed for short isoform also anneals to the long isoform. When RT-qPCR primers were designed to amplify short and long isoforms, same forward primer and two different reverse primers were chosen (Figure 3.3). During RT-qPCR experiments, reaction efficiencies of each experiment were evaluated cautiously for short and long PCR products, and efficiency differences of the short and long isoforms were incorporated into the  $\Delta\Delta Cq$  formula to diminish inter-experimental variations. Hence, relative expression of short product (short+long transcript) to long product (long isoform) were detected by RT-qPCR. As a negative control, MDA-MB-231 breast cancer cells (ER-) were also treated with E2 along with ER+ cell lines.

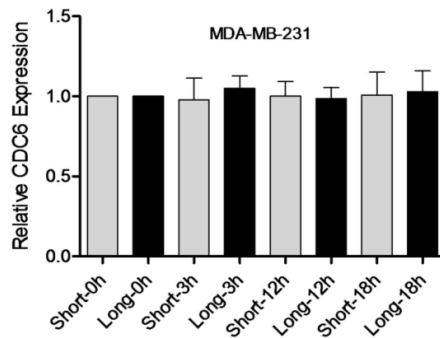
A.



B.



C.



**Figure 3.3. Relative quantification of *CDC6* 3'-UTR short and long isoforms in E2 treated and control cells.** A. MCF7 cells were treated with 10 nM E2 for 3 hours and 12 hours. B. T47D cells were treated with 10 nM E2 for 12 and 18 hours. C. MDA-MB-231 cells were treated with 10 nM E2 for 3, 12 and 18 hours. The fold change for the isoforms was normalized against the reference gene; *SDHA*. Quantification was done using the reaction efficiency correction and  $\Delta\Delta Cq$  method [108]. The baseline for the short and long isoforms in untreated samples was set to 1. \*\*\* indicates significant difference between short and long isoforms' expression,  $p < 0.001$  (One way ANOVA followed by Tukey's multiple comparison test).

In conformity with *in silico* analysis which was based on MCF7 microarray results, short *CDC6* 3'-UTR isoform expression was found to increase approximately 38% more than the long 3'-UTR isoform in response to E2 in MCF7 cells (Figure 3.3 A). After 3 hours E2 treatment, an increase in *CDC6* expression (both short and long) was detected, however there was no significant increase in the short 3'-UTR transcript. After 12 hours E2 treatment, increase of *CDC6* short transcript expression was around 3.5 fold compared to EtOH treated MCF7 cells. Overall, 12h E2 treatment caused both increased *CDC6* expression and 3'-UTR shortening in MCF7 cells.

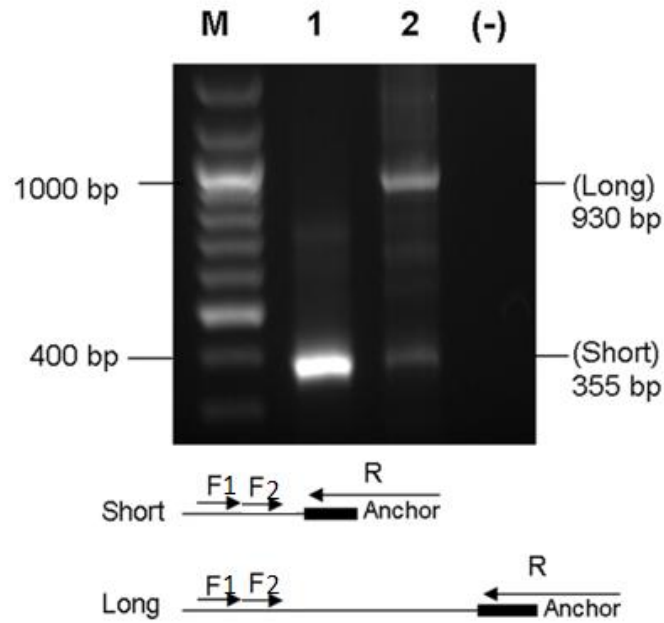
A robust increase of the short 3'-UTR isoform of *CDC6* was detected also in E2 treated T47D cells (Figure 3.3 B). When T47D cells were treated with E2 for 12 and 18 hours, increased expression of *CDC6* together with 3'-UTR shortening was observed at both time-points.

In contrast to the upregulation and 3' shortening of *CDC6* in response to E2 in ER+ cells, there was neither an upregulation nor 3'-UTR shortening of *CDC6* in ER- MDA-MB-231 cells when treated with E2 (Figure 3.3 C). ER negative MDA-MB-231 cells growth rate is known to be unaffected by E2 [102-104].

### 3.3 Rapid Amplification of cDNA Ends (RACE)

Since the short 3'-UTR primers also amplified the long 3'-UTR isoform, a further validation was needed to prove the existence of the short 3'-UTR. To accomplish that, 3'-RACE was performed. 3'-RACE-ready cDNA was synthesized by using an oligo dT-anchor primer as described in Materials and Method section.

Using the first round PCR product as a template, nested 3'-RACE PCR with a new forward and Anchor-R primers was performed. As can be seen from Hata! Başvuru kaynağı bulunamadı., lane 2 is the product of the second round of nested PCR. At the second round, expected *CDC6* 3'-UTR isoforms were detected at 355 bp and 930 bp for the short and long isoforms, respectively.



**Figure 3.4. Figure 3.5. 3'-RACE confirmed the existence of the short isoform.** cDNA was synthesized using E2 treated MCF7 total RNA with anchor oligo-dT using the 3'-RACE Kit (Roche Applied Sciences). CDC6 gene specific forward (F) and anchor specific reverse (R) primers amplified the expected 355 bp (lane 1, 2) and 930 bp (lane 2) products of the short and the long isoforms after nested 3'-RACE. 355 bp PCR product in lane 2 corresponding to the short isoform was gel purified and used as a PCR template (lane 1) with the CDC6 forward and R anchor primers. M: DNA ladder, (-): No template reaction.

To further amplify the 355 bp product to be sequenced, correct size band was gel extracted and used as a template for a third round of PCR. 3'RACE\_2 and Anchor-R primers were used in PCR, and the PCR product can be seen in lane 1 of Hata! Başvuru kaynağı bulunamadı.. Finally, this product was purified and sent to sequencing for verification.

Sequencing the PCR product for the short 3'-UTR verified the presence of the proximal poly(A) site and the poly(A) tail, (Figure 3.5) which further confirmed the existence of the short transcript.

```

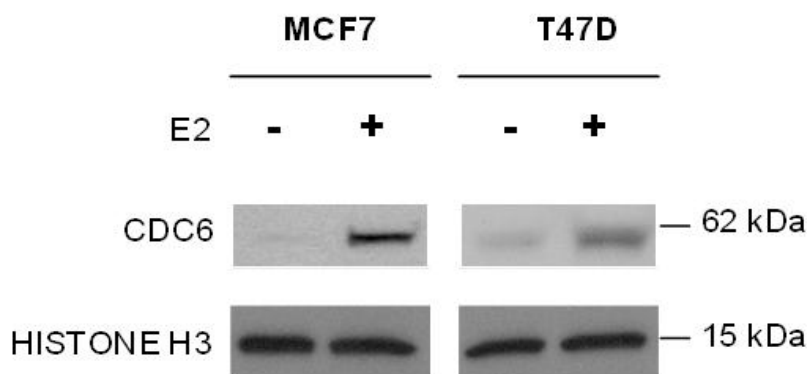
CDC6 1959 AGTGCTTTACACATTTCGGGCCTGAAAACAAATATGACCTTTTTTACTTGAAGCCAATGAA
CDC6 2019 TTTTAATCTATAGATTCTTTAATATTAGCACAGAATAATATCTTTGGGTCTTACTATTTT
CDC6 2079 TACCCATAAAAGTGACCAGGTAGACCCTTTTTAATTACATTCACTACTTCTACCACTTGT
CDC6 2139 GTATCTCTAGCCAATGTGCTTGCAAGTGTACAGATCTGTGTAGAGGAATGTGTGTATATT
CDC6 2199 TACCTCTTCGTTTGCTCAAACATGAGTGGGTATTTTTTTGTNNAAAAAAAAAAAAAAAAAN
                                         P-pA-Site

```

**Figure 3.5. Sequencing of the 3'-RACE product** (size: 355 bp) revealed the existence of the proximal poly(A) tail.

### 3.4 CDC6 Protein Levels Upon E2 Treatment

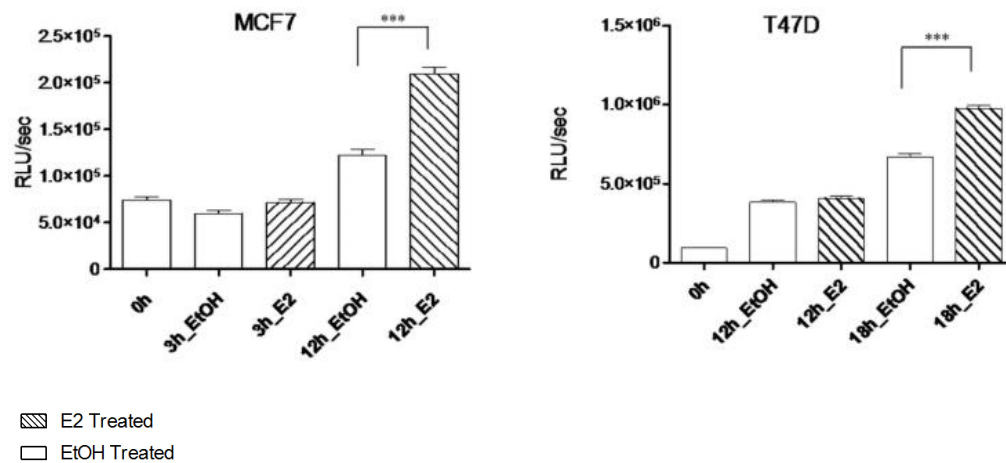
To inspect the correlation between transcript and protein levels, protein levels of CDC6 were investigated by Western blotting after E2 treatment. Antibody against CDC6 protein was used to detect nuclear CDC6 levels, while Histone H3 was used as a nuclear protein loading control. Western blot analysis showed that CDC6 protein levels increased in response to E2 treatment in both MCF7 and T47D cell lines (Figure 3.6). 3'-UTR shortening of *CDC6* in response to E2 indeed correlated with a rapid increase in CDC6 protein levels.



**Figure 3.6. E2 treatment caused increase of the CDC6 protein levels.** MCF7 and T47D nuclear lysates were collected followed by 10 nM E2 treatment for 12 and 18 hours, respectively. Histone H3 antibody was used as a nuclear protein loading control.

### 3.5 BrdU Incorporation

Given the role of CDC6 as the initiator of pre-replication complex formation [118, 119], next we investigated whether increased CDC6 protein levels indeed correlated with increased DNA replication. Asynchronous MCF7 and T47D cells were either treated with E2 or ethanol along with BrdU. The incorporation of BrdU into DNA in E2 treated cells compared to controls significantly increased after 12 hours for MCF7 and 18 hours for T47D (Figure 3.7). This time points also correlated with the CDC6 3'-UTR short isoform and protein level increase.



**Figure 3.7. E2 treatment induced increased BrdU incorporation.** Asynchronous MCF7 and T47D cells were grown in phenol red-free media supplemented with 10% dextran-coated-charcoal stripped FBS for 48 hours. Then, cells were treated with BrdU (10 nM) and E2 (10 nM) for the indicated time points along with the ethanol treated controls (EtOH). Incorporated BrdU was detected using the Turner Biosystems Luminometer. Values are given as relative light units (RLU) per sec at respective time points. \*\*\* indicates statistical significance,  $p < 0.001$  (One way ANOVA followed by Tukey's multiple comparison test).

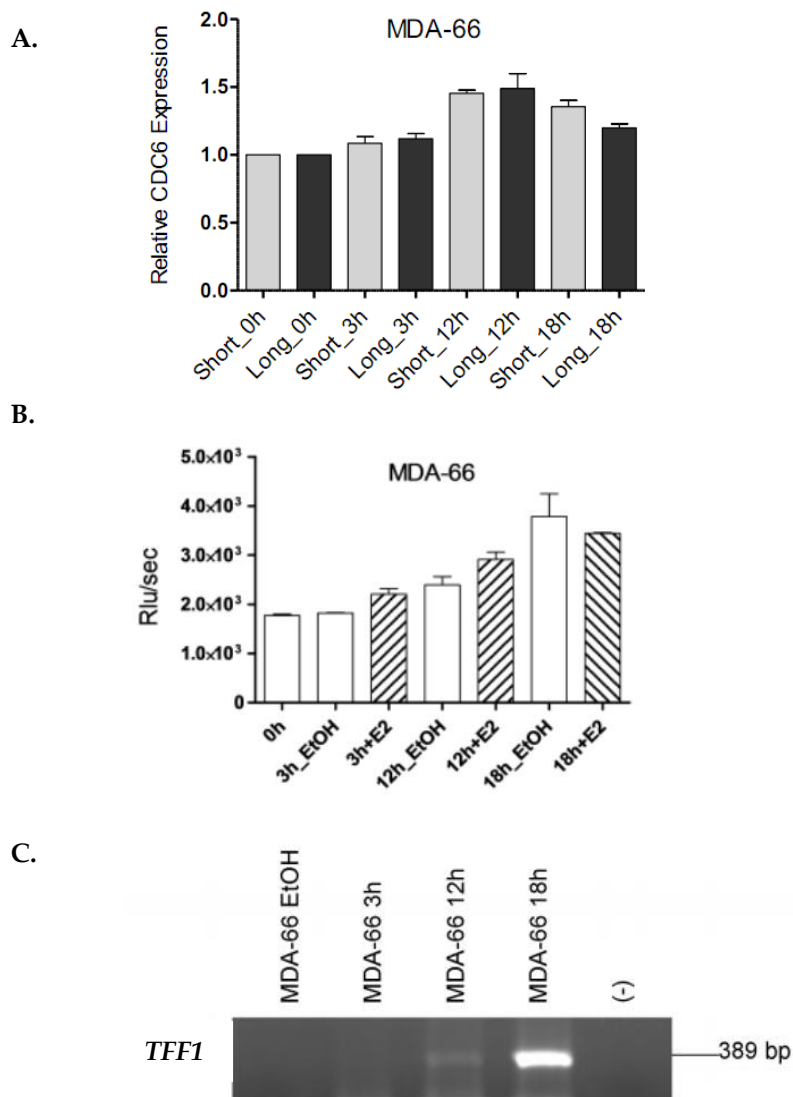
Hence, as a known key regulator of DNA replication, increased CDC6 protein levels were found to be highly correlative with increased BrdU incorporation in both MCF7 and T47D cells.

To this point, 3'-UTR shortening of CDC6 upon E2 exposure was shown by RT-qPCR and 3'-RACE. Furthermore, upregulation of the short isoform correlated with increased CDC6 protein levels and BrdU incorporation which reflected increased S phase entry. The next question was whether this increase in CDC6 was coupled to cell proliferation. E2-mediated response in ER-positive cell lines is associated with the expression of a vast number of

genes, some of which are involved in proliferation. It is therefore difficult to assess whether E2-mediated 3'-UTR shortening of *CDC6* is related to cell proliferation. Apart from the effect of E2 in general, to circumvent this issue, ER $\alpha$  transfected MDA-MB-231 cells (MDA-66) [100] was used. Previous studies showed that, in contrast to ER-positive MCF7 or T47D cells, re-expression of ER $\alpha$  endogenously or exogenously suppresses the growth of MDA-MB-231 cells in a time-dependent manner [120-123]. Consistent with these studies, short-term E2 treatment (up to 18 h) did not cause an increase in *CDC6* 3'-UTR short isoform expression compared to long isoform (Figure 3.8 A) or BrdU incorporation (Figure 3.8 B), while E2 responsive *TFF1* transcript levels increased (Figure 3.8 C).

These findings suggest that E2-mediated 3'-UTR shortening of *CDC6* could be coupled to cellular proliferation.





**Figure 3.8. Investigation of MDA-MB-231 cells.** A. Relative quantification of *CDC6* 3'-UTR short and long isoforms in MDA-66 cells (ER transfected MDA-MB-231) [100]. Cells were treated with E2 as described in the Methods section. E2 is known not to have a growth proliferative effect on ER transfected MDA-MB-231 cells [120-123]. B. E2 treatment did not induce increased BrdU incorporation. Asynchronous MDA-66 cells were grown in phenol red-free media supplemented with 10% dextran-coated-charcoal stripped FBS for 48 hours. Then, cells were treated with BrdU (10 nM) and E2 (10 nM) for the indicated time points along with the ethanol treated controls (EtOH). Incorporated BrdU was detected using the Turner Biosystems Luminometer. Values are given as relative light units (RLU) per sec at respective time points. C. *TFF1* is an E2 upregulated gene. MDA-66 cells were treated with ethanol or 10 nM E2 for indicated time points and RNA was isolated. Following DNase treatment, cDNA was synthesized for RT-PCR analysis. E2 treatment resulted with the upregulation of *TFF1* expression.

## 3.6 *CDC6* mRNA-miRNA Interactions

### 3.6.1 MicroRNA target sites on *CDC6* Isoforms

To further understand whether the E2 induced increase of the *CDC6* 3' UTR short isoform was a mechanism to evade miRNAs, that would allow for a more robust increase in protein levels, evasion of miRNA binding was investigated. First, we used three separate miRNA target prediction tools (TargetScan, PicTar and FindTar3) to detect whether there were any *CDC6* 3' UTR long isoform specific miRNAs.

TargetScan predicts targets of miRNAs by searching for the presence of conserved 8mer and 7mer sites that match the seed region of each miRNA [124]. In mammals, predictions are ranked based on the predicted efficacy of targeting as calculated using the context plus scores of the sites [125, 126]. PicTar is another algorithm for miRNA target identification that computes free energies of RNA:RNA duplexes together with 7mer pairing [127]. Finally, FindTar3 predicts miRNA-mRNA interaction based on the common criteria, including seed-pairing, free energy of miRNA:mRNA duplex, and proper dynamic score. Yet, FindTar3 concerns about central loop score which is the partial complementarity of miRNA: mRNA duplex to avoid miRNA target induced instability of miRNAs [128].

Three different miRNA target prediction programs predicted more than 60 miRNAs binding to the *CDC6* 3'-UTR, but interestingly, 87% of all predictions were localized to the region between the proximal polyA site (P-pA-Site) and the distal polyA site (D-pA-Site), a region which signifies the long 3'-UTR (Table 3.1). The intersection of all 3 prediction sets gave only 4 common miRNAs (miR-25, miR-541, miR-92a, miR-92b) which were predicted to bind only to the long 3'-UTR. No common miRNA binding site predictions were present for the short 3'-UTR (Table 3.2).

**Table 3.2. CDC6 3'-UTR binding miRNA predictions.** miRNAs predicted to be binding to CDC6 3'-UTR are listed. Proximal and distal miRNA predictions were recorded based on predictions of TargetScan, PicTar and FindTar3. \* shows the 4 common miRNAs which were predicted to bind only to long CDC6 3'-UTR.

	<b>P-pA-Site</b>		<b>D-pA-Site</b>	
<b>CDC6 3'UTR</b>	<b>1</b>	<b>342</b>	<b>916</b>	<b>1160</b>

	<b>Proximal</b>	<b>Distal</b>	<b>Total</b>
PITA	4	12	16
FindTar3	0	30	30
TargetScan	5	16	21
Common miRs *	0	4	4

\* miR-25, miR-541, miR-92a, miR-92b

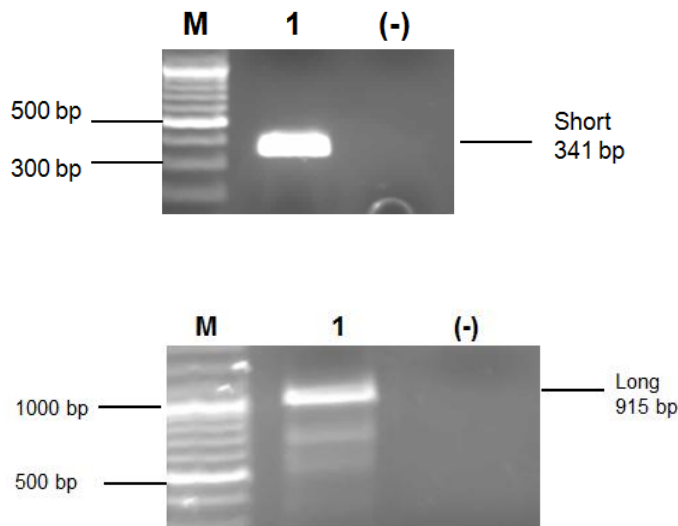
Remarkably, these 4 miRNAs predicted by different bioinformatics tools, were implicated in G1/S entry network and cell proliferation in the literature. Although expressed as part of different clusters, miR-25, miR-92a, and miR-92b have the same seed sequence (AUUGCA) that is *in silico* predicted to bind to CDC6. miR-25 and miR-92b were already shown to target p57, a Cip/Kip family member of Cdk inhibitors and thus may have control over G1/S transition [129, 130]. On the other hand, miR-541 was shown to play roles in differentiation, development and androgen-induced cell proliferation [131, 132]. The possible connection between these miRs, 3' UTR shortening of CDC6 and cell cycle regulation remains to be further investigated.

As results of miRNA binding predictions pointed out that long 3'-UTR isoform could be more susceptible to post-transcriptional miRNA dependent repression than the short isoform, to test this prediction, an *in vitro* reporter assay was used.

### 3.6.2 Cloning of 3'-UTR Isoforms

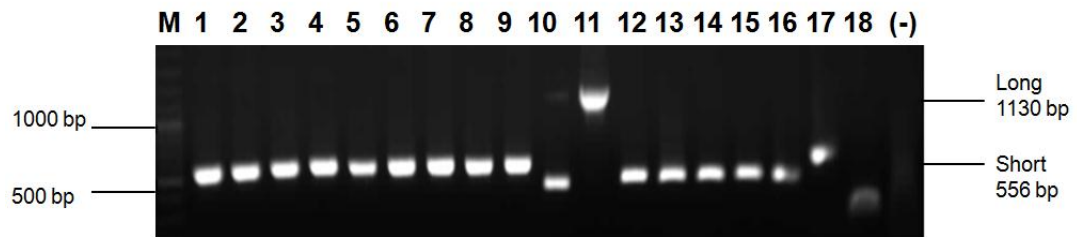
To investigate whether the CDC6 long 3'-UTR actually harbored miRNA binding sites which were not present in the short 3'-UTR, we cloned the short and long 3'-UTR sequences into the 3'-UTR of the firefly luciferase gene of a reporter plasmid pMIR-Report (Appendix G). For this purpose, SacI and HindIII recognition site harboring cloning primers were used. To clone short and long sequences, after PCR amplification, products were loaded on 1% Agarose gel (Figure 3.9). Following electrophoresis, corresponding bands were cut and

purified. Agarose extracted PCR products were double digested and ligated into pMIR-Report.



**Figure 3.9. Cloning CDC6 3'-UTR isoforms.** A. To clone short isoform, CDC6\_pMIR\_F and CDC6\_pMIR\_R1 primers amplified the expected 341 bp (lane 1) product. Following PCR conditions were used for short isoform: incubation at 95°C for 10 minutes, 35 cycles of 95°C for 30 seconds, 60°C for 30 seconds, and 72°C for 1 minute, and final extension at 72°C for 10 minutes. B. To clone long isoform, CDC6\_pMIR\_F and CDC6\_pMIR\_R2 primers amplified the expected 915 bp product (lanes 1). Following PCR conditions used for long isoform, incubation at 95°C for 10 minutes, 35 cycles of 95°C for 30 seconds, 60°C for 30 seconds, and 72°C for 2 minute, and final extension at 72°C for 10 minutes. M: DNA ladder, (-): No template reaction.

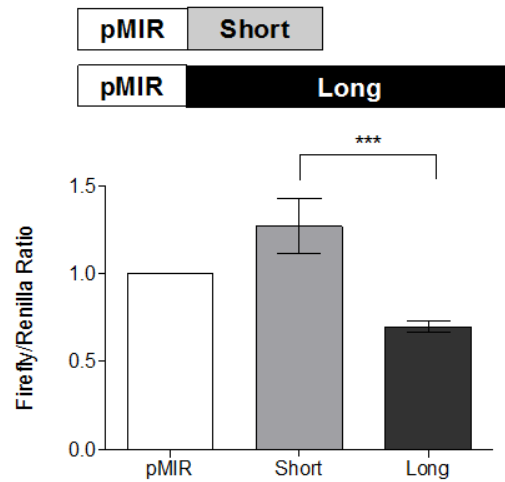
The colonies obtained from bacterial transformation (9 colonies for short and 8 colonies for long isoform) were used for colony PCR. As can be seen from Figure 3.10, all of the 9 colonies used in colony PCR for short isoform gave the expected size 556 bp (Lanes 1 to 9), while only one colony (Lane 11) gave the expected size 1130 bp for the long isoform. Empty pMIR-Report was used as a positive control, which gave the 215 bp band (Lane 18). One colony for short (Lane 2) and one colony for long (Lane 11) were sent to sequencing and confirmed for having the correct DNA sequences (Appendix H).



**Figure 3.10. Colony PCR Results.** 9 colonies for short (Lanes 1 to 9) and 8 colonies for long isoform (Lanes 10 to 17) were used for colony PCR. Primers designed from pMIR-Report were used. pMIR-Report (Lane 18) used as a positive control. Following PCR conditions were used: incubation at 94°C for 10 minutes, 35 cycles of 94°C for 30 seconds, 56°C for 30 seconds, and 72°C for 30 seconds, and final extension at 72°C for 10 minutes. M: DNA ladder, (-): No template reaction.

### 3.6.3 Dual-luciferase Assay

Short and long 3'-UTR constructs in pMIR along with phRL-TK (renilla luciferase), for normalization of transfection, were transfected into MCF7 cells. 24 hours after transfection, cells were lysed and renilla and firefly luciferase readings were taken by using the Modulus Microplate Luminometer. The short 3'-UTR construct had a significantly higher (85%) luciferase activity compared to the long 3'-UTR construct (Figure 3.11). These results suggested that the short 3'-UTR transcript was more efficiently translated which may explain how CDC6 protein levels increased in response to E2.



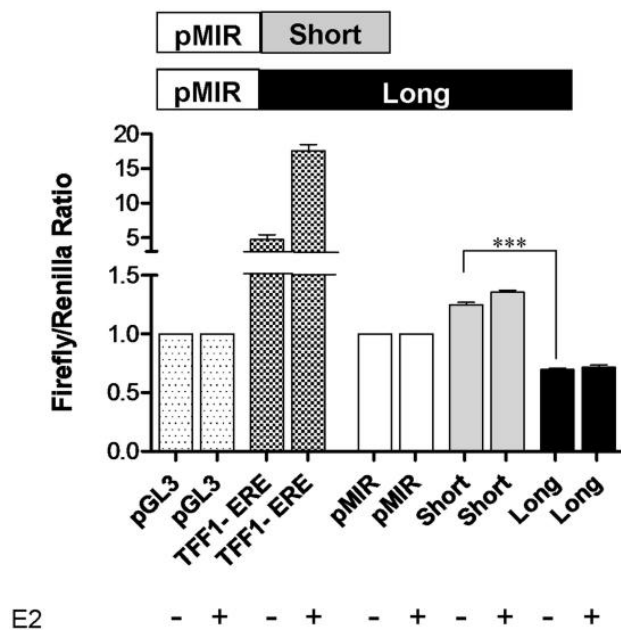
**Figure 3.11. Dual luciferase reporter assay.** *CDC6* short and long 3'UTRs were cloned into the 3' UTR of Firefly luciferase gene in pMIR. pMIR-Short, pMIR-Long and pMIR-empty vectors were co-transfected with phRL-TK into MCF7 cells. Dual-luciferase assay was performed 24 hours after transfection. Firefly/Renilla luciferase read-outs from the constructs were normalized to that of empty pMIR which was set to 1. \*\*\* indicates  $p < 0.05$  (One way ANOVA followed by Tukey's multiple comparison test).

The reason why short isoform was translated more efficiently could be due to several reasons. First, short isoform may be evading miRNAs which are targeting the long isoform. We were able to identify 4 common miRNAs (miR-25, miR-541, miR-92a, miR-92b) targeting only the long isoform by three different miRNA target prediction tools. As miRNAs are negative regulators at post-transcriptional level, these miRNAs may be targeting long *CDC6* isoform and repressing its translation, while short isoform escapes this regulation. Besides, as of yet unknown RNA binding proteins (RBPs) could be another reason for translational repression of long isoform. Sequence-specific RBPs plays important roles in translational regulation by binding to specific *cis*-acting elements located in the 3'-UTR [133, 134].

After that, the mechanism of how *CDC6* 3'-UTR short isoform increased more than the longer isoform in E2 treated cells was investigated. For that, two possibilities were considered; 1) post-transcriptional regulation by E2 regulated miRNAs to explain increased short isoform relative to long isoform in response to E2, or 2) E2 induced the transcriptional upregulation and APA of *CDC6* to generate a shorter 3' isoform.

First, the possibility of E2 regulated miRNAs to explain increased short isoform levels was examined. Therefore, short and long 3'-UTR harboring reporter constructs along with the phRL-TK vector were transfected into ER+ MCF7 cells that were treated with E2 or ethanol. There was no change in the luciferase activity of the short or the long 3'-UTR construct in response to E2 treatment compared to respective untreated short or long isoform constructs (Figure 3.11). While the luciferase activity difference of the short vs the long 3'-UTR constructs was still evident, the difference was E2 independent. The effect of E2 treatment

was verified by using the E2 responsive promoter region of *TFF1* gene as a positive control. Approximately 4 fold upregulation of luciferase expression was detected after E2 treatment (Figure 3.11). Lack of any E2 induced change in luciferase activity for isoforms suggested that E2 did not have a post-transcriptional effect on the isoforms such as binding of E2 up-regulated miRNAs and/or increased degradation of the long isoform in response to E2. Based on the luciferase reporter assay results in E2 treated cells, E2 effect on CDC6 protein increase was more likely to be transcriptional, i.e. through APA and switching to a shorter 3'-UTR, rather than post-transcriptional mechanisms.

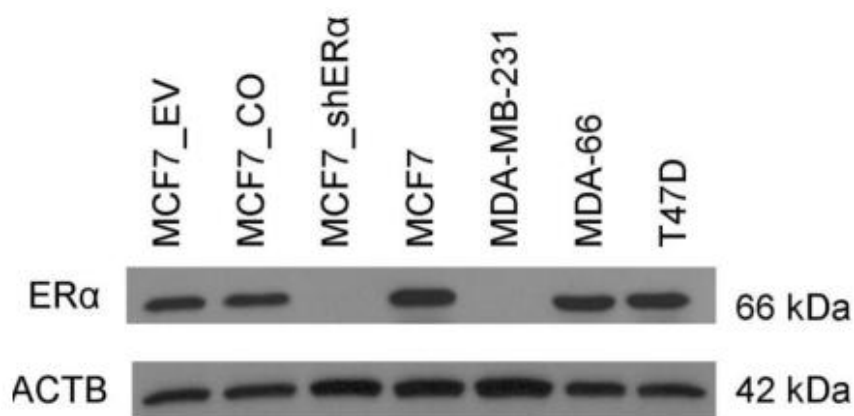


**Figure 3.12. Dual luciferase reporter assay after E2 treatment.** MCF7 cells were kept in phenol red-free medium supplemented with 10% charcoal stripped FBS for 48 hours. 12 hours post-transfection, E2 was applied to a final concentration of 10 nM for 12 hours. E2 responsive *TFF1*-promoter construct was used as a positive control for E2 treatment. pGL3 is the empty vector, *TFF1*-ERE (in pGL3) harbors the promoter and E2 responsive regions of the *TFF1* gene. Dual-luciferase assay was performed 24 hours after transfection. Firefly/Renilla luciferase read-outs from the constructs were normalized to that of empty pMIR which was set to 1. \*\*\* indicates  $p < 0.001$  (One way ANOVA followed by Tukey's multiple comparison test).

### 3.7 E2 and ER Dependency of CDC6 APA

To understand whether increased abundance of the short 3'-UTR isoform compared to the long isoform was indeed due to transcriptional upregulation by E2 and ER, two different approaches were used: 1) Silencing of ER $\alpha$ , 2) ER/E2 antagonists

First, to investigate ER $\alpha$  levels in cell lines, an antibody against ER $\alpha$  was used. As can be seen from Figure 3.13, ER $\alpha$  protein levels of the cell lines used in this study were confirmed. Knock-down of ER in stably transfected MCF7 cells and expression of the protein in ectopically ER expressing MDA-66 cells were confirmed by analyzing ER $\alpha$  protein levels. MDA-66 [99, 100] cells which are ectopically ER expressing MDA-MB-231 cells were used in ER dependency experiments.



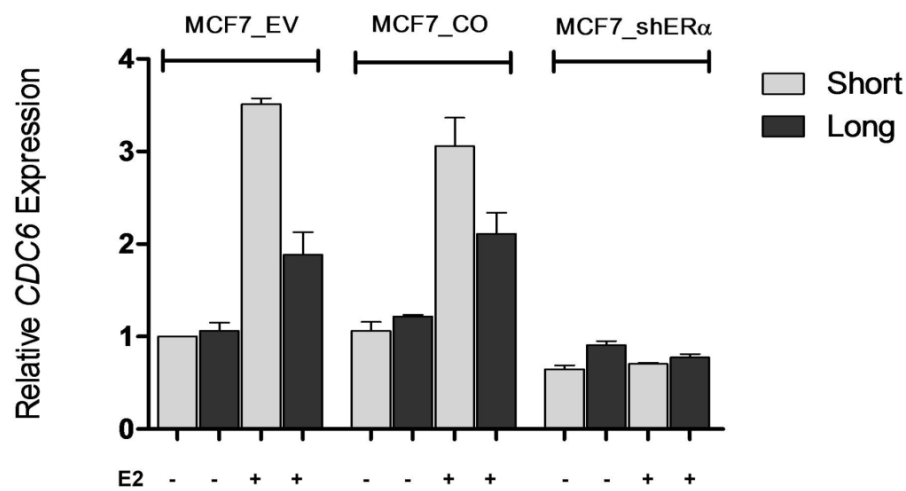
**Figure 3.13. ER protein levels detected in cell lines used in the study.** MCF7\_EV cells were stably transfected with empty vector (pSR), MCF7\_CO cells were transfected with control shRNA [101], and MCF7\_shER $\alpha$  cells were transfected with ER $\alpha$  shRNA [99]. MDA-66 cells are MDA-MB-231 cells that were stably transfected with ER [99, 100]. MCF7, MDA-MB-231 and T47D cells were used throughout the study as ER $^+$  cells. 50  $\mu$ g lysate were loaded on 8% PAGE.  $\beta$ -actin was used as a protein loading control.

#### 3.7.1 ER Dependency of CDC6 APA

ER $\alpha$  is a ligand-dependent transcription factor which mediates estrogen signaling [83]. Estrogen binds to the ER $\alpha$  after traversing the cellular membrane, leading to receptor activation. To do this, ER $\alpha$  interacts with *cis*-regulatory elements of target genes either directly or indirectly. For direct interaction ER $\alpha$  binds to estrogen response elements (EREs). Also ER $\alpha$  interacts indirectly by associating with AP-1 and SP-1 transcription factor complexes and their respective binding sites [135, 136].



To understand whether increased abundance of the short 3'-UTR isoform compared the long isoform was indeed due to transcriptional upregulation by E2 and ER $\alpha$ , we used MCF7 cells that were stably transfected with an shRNA construct to silence ER $\alpha$  [99]. As can be seen from Figure 3.13, in MCF7\_shER $\alpha$  cells ER $\alpha$  was silenced successfully. To test ER dependency of *CDC6* 3'-UTR shortening, these ER $\alpha$  silenced cells together with MCF7\_EV (empty vector transfected) and MCF7\_CO (control shRNA transfected) cells were used. E2 treatment in ER silenced MCF7 cells did not cause an increase in the *CDC6* short isoform whereas MCF7\_ and MCF7\_CO cells had increased *CDC6* short isoform expression in response to E2 (Figure 3.14).



**Figure 3.14. E2 does not induce *CDC6* 3' UTR isoform increase in ER silenced MCF7 cells.** Cells were treated with E2 as described above. Relative quantification of *CDC6* 3' UTR short and long isoforms was determined in MCF7-EV (empty vector transfected), MCF7\_CO (control shRNA transfected) and in MCF7\_shER $\alpha$  cells before and after 12 hours of E2 treatment. The fold change for the isoforms was normalized against the reference gene; SDHA. Quantification was done using the reaction efficiency correction and  $\Delta\Delta C_q$  method [108]. The baseline for the short and long isoforms in untreated samples was set to 1.

To further investigate transcriptional upregulation of *CDC6* 3'-UTR isoforms, an ER specific antagonist ICI 182,780 (ICI) [105] alone or together with E2 were used to treat MCF and T47D cells. Based on the earlier E2 treatment time-points, cells were treated with ICI and E2 for 12 hours. ICI is a high affinity steroidal ER $\alpha$  antagonist which is also known as Fulvestrant (AstraZeneca) used for the treatment of hormone receptor positive breast cancer patients.

When *CDC6* expression was quantified, RT-qPCR results clearly showed increased short 3'-UTR isoform compared to the longer isoform in response to E2 alone, short and long *CDC6* expression decreased in ICI and E2 treated cells, showing the ER specific response in MC7

and T47D cells (Figure 3.15 A and B). Together with ER $\alpha$  silencing results, our findings showed the E2 and ER dependency of *CDC6* APA. When ER $\alpha$  was either silenced by shRNA or blocked by ICI, E2 treatment did not lead to expression increase of *CDC6* or 3'-UTR shortening.

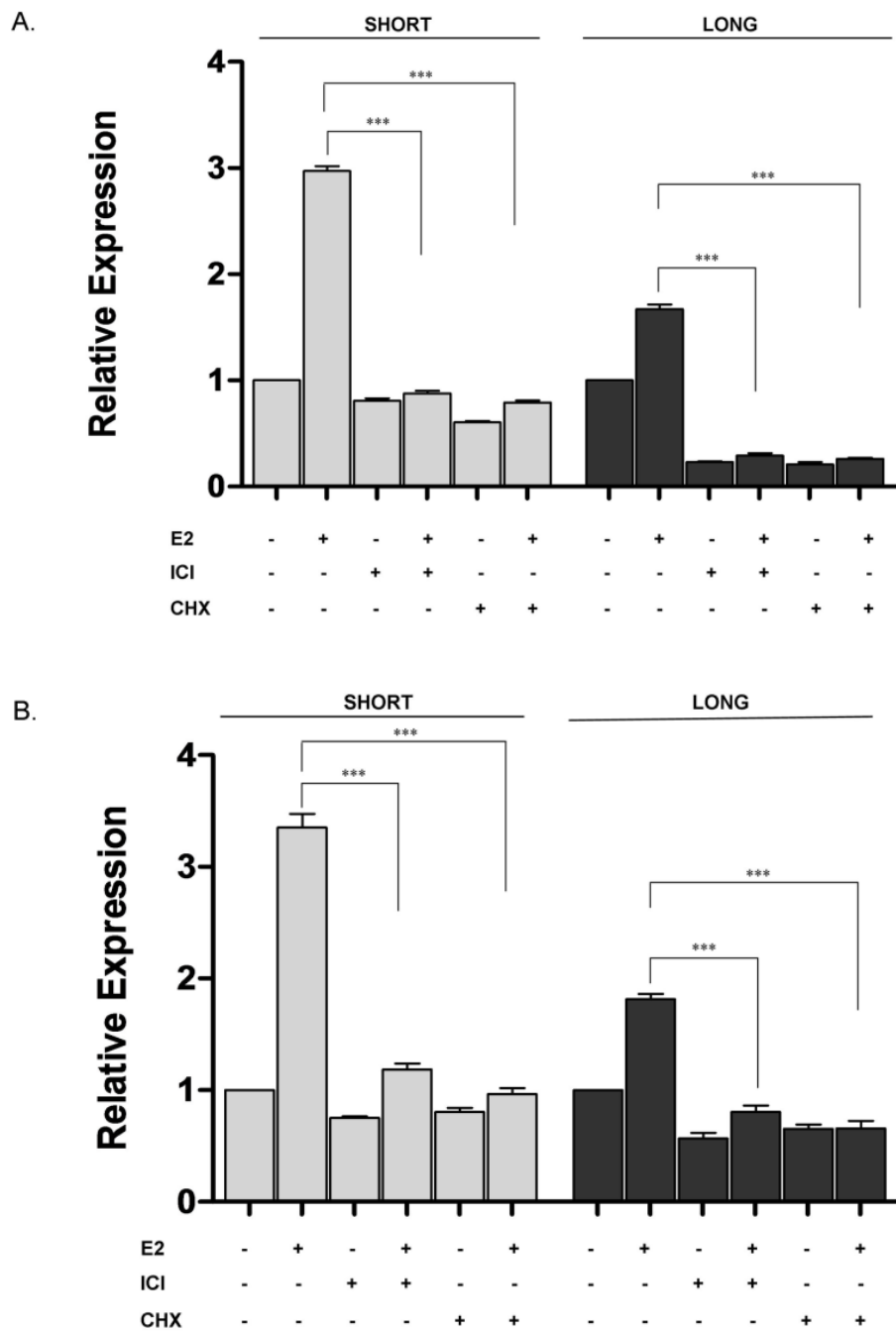


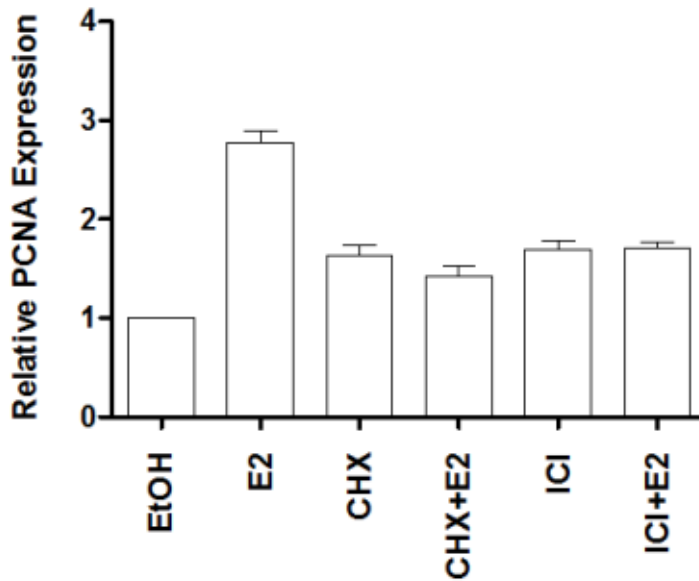
Figure 3.15. E2 induced *CDC6* expression requires ER and de novo protein synthesis. A. MCF7 and B. T47D cells were grown in phenol red-free medium supplemented with 10%

dextran-coated-charcoal stripped FBS, pre-treated with 1  $\mu$ M ICI or with 10  $\mu$ g/mL CHX (Cycloheximide) for 1 hour, then with 10 nM E2 for 12 hours. Relative expression of *CDC6* short (grey bars) and long (black bars) isoforms was determined by RT-qPCR. The baseline for the control treated samples was set to 1. \*\*\* indicates statistical significance ( $p < 0.001$ ).

After that, *CDC6* was further investigated as a direct or indirect target of ER. Assessment of the 5' regulatory region of the *CDC6* by Dragon ERE Finder tool (<http://datam.i2r.a-star.edu.sg/ereV3/>) did not identify any consensus EREs [137]. Furthermore, Con Tra (<http://bioit.dmbr.ugent.be/contrav2/>) transcription factor binding site prediction tool did not reveal any possible AP-1 and SP-1 binding sites [138]. Absence of any ERE consensus sequences or AP-1 and SP-1 binding sites raised the possibility that expression of *CDC6* may require other transcription factors.

To clarify these predictions, translational inhibitor cycloheximide (CHX) was used. CHX inhibits protein biosynthesis by interfering with the translocation step, consequently blocking translational elongation [139]. When cells were pre-treated with CHX and then with E2 for 12 hours, the transcriptional upregulation of *CDC6* was prevented (Figure 3.15 A and B), suggesting *CDC6* to be transcriptionally upregulated possibly by other transcription factors upon E2 exposure. Both short and long isoform transcription was demolished after ICI, and CHX treatments in ER+ cells, confirming the role of E2 induced, ER dependent transcription and APA regulation of *CDC6*.

To confirm effectiveness of the ICI and CHX treatments, *PCNA* (Proliferating Cell Nuclear Antigen) which is a known secondary response gene that is upregulated in the presence of E2 [140] was quantified via RT-qPCR (Figure 3.16). E2-stimulated expression of *PCNA* was lost when MCF7 cells were pre-treated with CHX. Likewise, after blocking the ER $\alpha$  by ICI, expressional upregulation of *PCNA* was inhibited. These data confirmed that ICI and CHX treatments to MCF7 cells were done effectively.



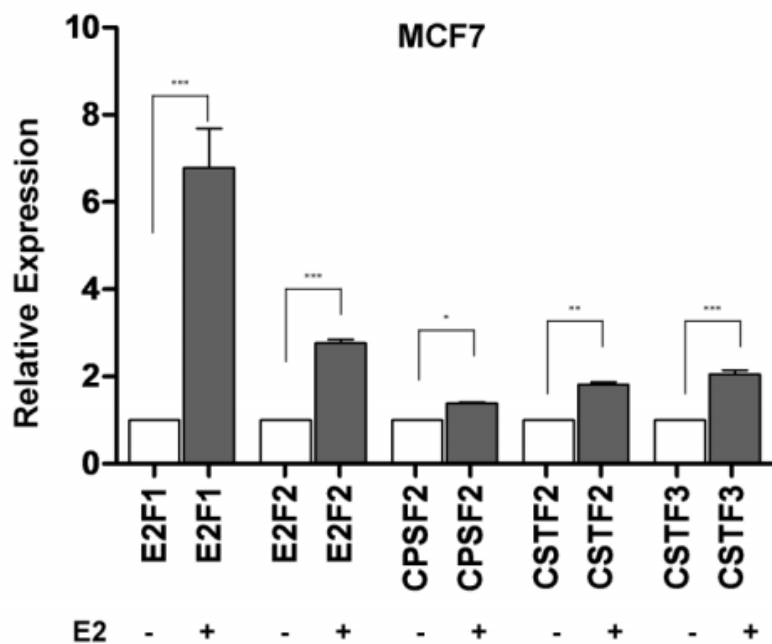
**Figure 3.16. PCNA expression detected by RT-qPCR after ICI and CHX treatments.** MCF7 cells were grown in phenol red-free medium supplemented with 10% dextran-coated-charcoal stripped FBS, pre-treated with 1  $\mu$ M ICI or 10  $\mu$ g/mL CHX (Cycloheximide) for 1 hour, then with 10 nM E2 for 12 hours. The baseline for the control treated samples was set to 1. PCNA (NM\_002592) was amplified using the following primer set; PCNA\_F: 5'-TGCAGATGTACCCCTTGTTG-3', PCNA\_R: 5'- GCTGGCATCTTAGAAGCAGTT-3'. Following conditions were used: incubation at 94°C for 10 minutes, 40 cycles of 94°C for 15 seconds, 56°C for 30 seconds, and 72°C for 30 seconds.

### 3.8 Transcriptional Regulators of APA

APA regulatory mechanisms are yet to be discovered. The mechanism by which proliferation induces APA is unknown. Recently, it has been shown that E2F transcription factors are involved in the enhanced APA in proliferation [110]. It has been long known that E2F family of transcription factors are one of the main regulators of transcriptional network associated with cell cycle progression [141]. E2F1 and E2F2 proteins are located downstream of growth factor signaling cascades, in which they regulate genes required for cell cycle progression by acting as transcriptional activators. Indeed, transcription of *CDC6* is known to be regulated by E2Fs [142, 143].

CSTF2 (cleavage stimulation factor, 3' pre-RNA, subunit 2), CSTF3 (cleavage stimulation factor, 3' pre-RNA, subunit 3) and CPSF2 (cleavage and polyadenylation specific factor 2) are important units in 3' end-processing machinery [9] which also have been shown to be regulated by E2Fs [110]. As a regulator of *CDC6* transcription and APA, we investigated expression levels of *E2F1*, and *E2F2* after treating MCF7 cells with E2. Besides, expression levels of E2F regulated *CSTF2*, *CSTF3*, and *CPSF2* were also examined.

In this study, after 12 h E2 treatment to MCF7 cells, expression levels of *E2F1*, *E2F2*, *CSTF2*, *CSTF3* and *CPSF2* were investigated via RT-qPCR. As can be seen from Figure 3.17, a significant increase in the expression of *E2F* (*E2F1* and *E2F2*) transcription factors and 3' end-processing factors *CSTF2*, *CSTF3* and *CPSF2* were detected. *E2F* transcription factors are not only involved in the regulation of *CDC6* transcription but also known to regulate 3'-UTR processing factors [110, 143]. Thus, E2-mediated increases in *E2F* along with 3' end-processing enzyme transcript levels may be responsible for APA in proliferation.



**Figure 3.17. Expression levels of *E2Fs* and APA factors after E2 treatment.** *E2F1*, *E2F2* and 3' UTR processing gene transcripts (*CSTF2* (cleavage stimulation factor, 3' pre-RNA, subunit 2), *CSTF3* (cleavage stimulation factor, 3' pre-RNA, subunit 3) and *CPSF2* (cleavage and polyadenylation specific factor 2) are upregulated in response to E2 treatment in MCF7 cells. Cells were treated with E2 and cDNA was prepared as described in Methods. The fold change for the transcripts was normalized against the reference gene; *SDHA*. Quantification was done using the reaction efficiency correction and  $\Delta\Delta C_q$  method. The baseline for each transcripts' in untreated samples were set to 1. Experiment was repeated two independent times with 3 replicas. One-way ANOVA with Tukey's multiple comparison post test was performed using GraphPad Prism (California, USA). \* ( $p \leq 0.05$ ), \*\* ( $p \leq 0.01$ ), and \*\*\* ( $p \leq 0.001$ ) indicates statistical significance.

Moreover, *E2F1* and *E2F2* are known to have regulatory roles in G1/S phase entry [144] and they are known to be transcriptional regulators of *CDC6* [142, 143]. Recently increased expression of *E2Fs* and their regulatory role in APA were shown [110]. APA regulatory mechanisms which enable transcripts to have shorter 3'-UTRs during proliferation are not

discovered yet. Still, increased expression of E2Fs and 3'-end processing factors may be the reason of increased expression and 3'-UTR shortening of *CDC6* upon E2 treatment.

## CHAPTER 4

### CONCLUSION

There is growing evidence on the functional consequences of over-active DNA replication machinery contributing to transformation in cells. Recently, over-expression and increased chromatin-association of pre-replication complex proteins, one of which was CDC6, was shown in transformed cells [145]. CDC6 is a proto-oncogenic protein that contributes to oncogenesis through its role in DNA replication initiation and its repressive effect on the INK4/ARF locus [146, 147]. Interestingly, mechanisms to explain increased activation of CDC6 in cancer cells are not well known.

Here, by using an *in silico* and *in vitro* approach, up-regulation and 3'-UTR shortening of CDC6 upon E2 treatment was shown in breast cancer cells. APA dependent 3'-UTR shortening is gaining a well deserved attention due to its potential role in proto-oncogene activation cases where no other causative factor is known. To identify such 3'-UTR shortening cases, an *in silico* approach that allowed us to re-analyze existing gene expression arrays at the probe level based on the positions of the reported polyA sites. Affymetrix Human Genome U133 Plus 2.0 Arrays designed to have 11 probes per transcript to get a better estimation of the expression level of a transcript [148]. Since the microarray cDNA preparation labeling protocol initiates with an oligo-d(T) primer, probes are generally designed to recognize the 3'-ends of a transcript. However, possible APA events are not considered during the microarray design which leads to inaccurate interpretation of data when there are alternatively polyadenylated isoforms [149]. In our study, although limited to array design decisions, we were able to group 3'-UTR matching probes into proximal and distal sets based on the polyA site positions. Using such probe sets, CDC6 proximal and distal probe ratios were found to be differentially increasing in response to E2 in 3 independent datasets, suggesting APA regulation of CDC6. This approach allowed us to distinguish E2 induced differential expression of 3'-UTR isoforms. This observation emphasizes the importance of RNA sequencing and/or development of expression analysis systems that can facilitate detection of APA regulated and alternatively spliced isoforms which may be undetected in conventional gene expression array systems.

However, because such a probe based approach may introduce certain quantitative biases, E2 induced 3'-UTR shortening of CDC6 was further confirmed via *in vitro* assays. In accordance with the *in silico* analysis, E2 did induce upregulation of both isoforms, more so of the shorter isoform, only in ER positive breast cancer cells. Moreover, upregulation of the short isoform correlated with increased CDC6 protein levels and BrdU incorporation which may reflect increased S phase entry. Additionally, higher translation of the shorter isoform compared to the longer one was detected in a luciferase reporter assay system.

The reason of such 3'-UTR shortening of *CDC6* could be due to the need to rapidly upregulate protein levels upon a proliferation signal, in our case E2, by escaping from miRNA dependent negative regulation. Interestingly, the 4 miRNAs predicted by different bioinformatics tools, were implicated in G1/S entry network and cell proliferation in the literature. Although expressed as part of different clusters, miR-25, miR-92a, and miR-92b have the same seed sequence (AUUGCA) that is *in silico* predicted to bind to *CDC6*. miR-25 and miR-92b were shown to target p57, a Cip/Kip family member of Cdk inhibitors and thus may have control over G1/S transition [129, 130]. On the other hand, miR-541 was shown to play roles in differentiation, development and androgen-induced cell proliferation [131, 132]. The possible connection between these miRs, 3'-UTR shortening of *CDC6* and cell cycle regulation remains to be further investigated. Other mechanisms (e.g. Adenylate-uridylylate-rich *cis*-elements (AREs) and/or RNA binding proteins) may also negatively affect the translation rate and/or subcellular localization of the long isoform.

While the upregulation and 3'-UTR shortening of *CDC6* were E2 and ER dependent, we also observed that the expression level of the longer isoform was decreased after the ICI and CHX treatments, independent of E2. The decreased levels of *CDC6* long isoform after inhibition of ER and protein synthesis may indicate the involvement of ER and possibly other transcription factors such as E2F [142, 143] that may induce low level basal expression of the *CDC6* transcript. Interestingly, we also observed increased levels of *E2F1* and *E2F2* in response to E2. *E2F* transcription factors are not only involved in the regulation of G1/S entry proteins but also known to regulate 3'-UTR processing enzymes [110] which, in our study, showed upregulation in response to E2. Recently, 3'-UTR processing genes were found to be upregulated in proliferating cells and E2F was shown to mediate enhanced APA [110]. Together, these findings may pioneer further studies to understand how APA is induced due to proliferative signals such as E2. To our knowledge, this study is the first to show hormone regulated 3'-UTR shortening of a potential oncogene in cancer cells. As thoroughly reviewed, E2 action and miRNA regulation is already an attractive research area [150]. Investigation of such E2 regulated APA events to avoid 3'-UTR dependent negative regulations may further contribute toward elucidating a more comprehensive understanding of diverse responses to E2 and of the mechanisms of rapid proliferation in hormonally responsive cancer cells. Given that there is mounting evidence on the significance of APA as an overlooked mechanism of gene expression control in normal and in cancer cells, further studies are underway to determine proliferation induced APA.

Our future studies will include high-throughput analysis of existing microarray data to investigate the extent of APA and 3'-UTR isoform generation in breast cancers. It also seems crucial to consider and explore APA in specific contexts where it may be controlled by different machinery and lead to different outcomes. For example, in E2 treated cells, APA of *CDC6* seemed to be due to the proliferative effect of E2. Proliferation signals (growth factors, hormones etc) or tissue origin may require and elucidate different APA mechanisms and responses in normal and in heterogeneous patient tumor samples. On the other hand, cancer cells may also favor APA to alter 3'-UTR lengths independent of the proliferative signals. Therefore, combinatorial use of different experimental and model systems is likely to give a more realistic picture of the extent of APA and 3'-UTR shortening events. A genome-wide polyadenylation mapping study based on high-throughput sequencing revealed that most yeast and human transcripts have as of yet uncharacterized and/or non-canonical polyadenylation sites, pointing out to the extent of complexity we are yet to face



[21, 151]. Given the significance of these new findings, we are likely to better understand not only how gene expression is regulated by APA but also how different choices of APA may further be involved in different physiological and disease state.



## REFERENCES

1. Kuhn, U., et al., *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*, in *J Biol Chem*. 2009: p. 22803-14.
2. Wickens, M., P. Anderson, and R.J. Jackson, *Life and death in the cytoplasm: messages from the 3' end*, in *Curr Opin Genet Dev*. 1997: p. 220-32.
3. Ji, X., J. Kong, and S.A. Liebhaber, *An RNA-protein complex links enhanced nuclear 3' processing with cytoplasmic mRNA stabilization*, in *EMBO J*. 2011: p. 2622-33.
4. Jacobson, A. and S.W. Peltz, *Interrelationships of the pathways of mRNA decay and translation in eukaryotic cells*. *Annu Rev Biochem*, 1996. **65**: p. 693-739.
5. Garneau, N.L., J. Wilusz, and C.J. Wilusz, *The highways and byways of mRNA decay*, in *Nat Rev Mol Cell Biol*. 2007: p. 113-26.
6. Proudfoot, N.J., A. Furger, and M.J. Dye, *Integrating mRNA processing with transcription*, in *Cell*. 2002: p. 501-12.
7. Licatalosi, D.D. and R.B. Darnell, *RNA processing and its regulation: global insights into biological networks*, in *Nat Rev Genet*. 2010: p. 75-87.
8. Richard, P. and J.L. Manley, *Transcription termination by nuclear RNA polymerases*, in *Genes Dev*. 2009: p. 1247-69.
9. Shi, Y., et al., *Molecular architecture of the human pre-mRNA 3' processing complex*, in *Mol Cell*. 2009: p. 365-76.
10. Wahle, E., *A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation*, in *Cell*. 1991: United States. p. 759-68.
11. Millevoi, S. and S. Vagner, *Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation*, in *Nucleic Acids Res*. 2010: p. 2757-74.
12. Tian, B., et al., *A large-scale analysis of mRNA polyadenylation of human and mouse genes*, in *Nucleic Acids Res*. 2005: p. 201-12.
13. Lutz, C.S. and A. Moreira, *Alternative mRNA polyadenylation in eukaryotes: an effective regulator of gene expression*. *Wiley Interdiscip Rev RNA*, 2011. **2**(1): p. 22-31.
14. Di Giammartino, D.C., K. Nishida, and J.L. Manley, *Mechanisms and consequences of alternative polyadenylation*, in *Mol Cell*. 2011, 2011 Elsevier Inc: p. 853-66.
15. Takagaki, Y. and J.L. Manley, *Levels of polyadenylation factor CstF-64 control IgM heavy chain mRNA accumulation and other events associated with B cell differentiation*, in *Mol Cell*. 1998: p. 761-71.
16. Kubo, T., et al., *Knock-down of 25 kDa subunit of cleavage factor Im in Hela cells alters alternative polyadenylation within 3'-UTRs*, in *Nucleic Acids Res*. 2006: p. 6264-71.
17. Ule, J., et al., *Nova regulates brain-specific splicing to shape the synapse*, in *Nat Genet*. 2005: p. 844-52.
18. Licatalosi, D.D., et al., *HITS-CLIP yields genome-wide insights into brain alternative RNA processing*, in *Nature*. 2008: p. 464-9.
19. Jan, C.H., et al., *Formation, regulation and evolution of Caenorhabditis elegans 3'UTRs*, in *Nature*. 2011: p. 97-101.
20. Wang, E.T., et al., *Alternative isoform regulation in human tissue transcriptomes*, in *Nature*. 2008: p. 470-6.

21. Ozsolak, F., et al., *Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation*, in *Cell*. 2010, 2010 Elsevier Inc: United States. p. 1018-29.
22. Erlitzki, R., J.C. Long, and E.C. Theil, *Multiple, conserved iron-responsive elements in the 3'-untranslated region of transferrin receptor mRNA enhance binding of iron regulatory protein 2*, in *J Biol Chem*. 2002: United States. p. 42579-87.
23. Spies, N., et al., *Biased chromatin signatures around polyadenylation sites and exons*. *Mol Cell*, 2009. **36**(2): p. 245-54.
24. Lian, Z., et al., *A genomic analysis of RNA polymerase II modification and chromatin architecture*. *Genome Res*, 2008. **18**(8): p. 1224-37.
25. Alt, F.W., et al., *Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is*. *Cell*, 1980. **20**(2): p. 293-301.
26. Early, P., et al., *Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA*. *Cell*, 1980. **20**(2): p. 313-9.
27. Rogers, J., et al., *Two mRNAs with different 3' ends encode membrane-bound and secreted forms of*. *Cell*, 1980. **20**(2): p. 303-12.
28. Setzer, D.R., et al., *Size heterogeneity in the 3' end of dihydrofolate reductase messenger RNAs in*. *Cell*, 1980. **22**(2 Pt 2): p. 361-70.
29. Edwards-Gilbert, G., K.L. Veraldi, and C. Milcarek, *Alternative poly(A) site selection in complex transcription units: means to an*. *Nucleic Acids Res*, 1997. **25**(13): p. 2547-61.
30. Yarden, A., D. Salomon, and B. Geiger, *Zebrafish cyclin D1 is differentially expressed during early embryogenesis*. *Biochim Biophys Acta*, 1995. **1264**(3): p. 257-60.
31. Tosi, M., et al., *Multiple polyadenylation sites in a mouse alpha-amylase gene*. *Nucleic Acids Res*, 1981. **9**(10): p. 2313-23.
32. Chuvpilo, S., et al., *Alternative polyadenylation events contribute to the induction of NF-ATc in*. *Immunity*, 1999. **10**(2): p. 261-9.
33. Gautheret, D., et al., *Alternate polyadenylation in human mRNAs: a large-scale analysis by EST*. *Genome Res*, 1998. **8**(5): p. 524-30.
34. Beaudoin, E. and D. Gautheret, *Identification of alternate polyadenylation sites and analysis of their tissue*. *Genome Res*, 2001. **11**(9): p. 1520-6.
35. Sandberg, R., et al., *Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites*. *Science*, 2008. **320**(5883): p. 1643-7.
36. Flavell, S.W., et al., *Genome-wide analysis of MEF2 transcriptional program reveals synaptic target*. *Neuron*, 2008. **60**(6): p. 1022-38.
37. Ji, Z., et al., *Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development*. *Proc Natl Acad Sci U S A*, 2009. **106**(17): p. 7028-33.
38. Mangone, M., et al., *The landscape of C. elegans 3'UTRs*, in *Science*. 2010: p. 432-5.
39. Haenni, S., et al., *Analysis of C. elegans intestinal gene expression and polyadenylation by*. *Nucleic Acids Res*, 2012. **40**(13): p. 6304-18.
40. Smibert, P., et al., *Global patterns of tissue-specific alternative polyadenylation in Drosophila*. *Cell Rep*, 2012. **1**(3): p. 277-89.
41. Sherstnev, A., et al., *Direct sequencing of Arabidopsis thaliana RNA reveals patterns of cleavage and*. *Nat Struct Mol Biol*, 2012. **19**(8): p. 845-52.

42. Wu, X., et al., *Genome-wide landscape of polyadenylation in Arabidopsis provides evidence for*. Proc Natl Acad Sci U S A, 2011. **108**(30): p. 12533-8.
43. Shen, Y., et al., *Transcriptome dynamics through alternative polyadenylation in developmental and environmental responses in plants revealed by deep sequencing*, in *Genome Res*. 2011: United States. p. 1478-86.
44. Legendre, M., et al., *Differential repression of alternative transcripts: a screen for miRNA targets*. PLoS Comput Biol, 2006. **2**(5): p. e43.
45. Singh, P., et al., *Global Changes in Processing of mRNA 3' Untranslated Regions Characterize Clinically Distinct Cancer Subtypes*. Cancer Res, 2009.
46. Mayr, C. and D.P. Bartel, *Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells*. Cell, 2009. **138**(4): p. 673-84.
47. Tranter, M., et al., *Coordinated post-transcriptional regulation of Hsp70.3 gene expression by*. J Biol Chem, 2011. **286**(34): p. 29828-37.
48. Wiestner, A., et al., *Point mutations and genomic deletions in CCND1 create stable truncated cyclin D1*. Blood, 2007. **109**(11): p. 4599-606.
49. Mayr, C., M.T. Hemann, and D.P. Bartel, *Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation*. Science, 2007. **315**(5818): p. 1576-9.
50. Lee, Y.S. and A. Dutta, *The tumor suppressor microRNA let-7 represses the HMGA2 oncogene*. Genes Dev, 2007. **21**(9): p. 1025-30.
51. Ghosh, T., et al., *MicroRNA-mediated up-regulation of an alternatively polyadenylated variant of the mouse cytoplasmic {beta}-actin gene*. Nucleic Acids Res, 2008. **36**(19): p. 6318-32.
52. Ji, Z. and B. Tian, *Reprogramming of 3' untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types*. PLoS One, 2009. **4**(12): p. e8419.
53. Fu, Y., et al., *Differential genome-wide profiling of tandem 3' UTRs among human breast cancer and normal cells by high-throughput sequencing*, in *Genome Res*. 2011: p. 741-7.
54. Chen, C.Y. and A.B. Shyu, *AU-rich elements: characterization and importance in mRNA degradation*, in *Trends Biochem Sci*. 1995: England. p. 465-70.
55. Barreau, C., L. Paillard, and H.B. Osborne, *AU-rich elements and associated factors: are there unifying principles?*, in *Nucleic Acids Res*. 2005: p. 7138-50.
56. Filipowicz, W., S.N. Bhattacharyya, and N. Sonenberg, *Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight?*, in *Nat Rev Genet*. 2008: England. p. 102-14.
57. Ambros, V., *The functions of animal microRNAs*, in *Nature*. 2004: p. 350-5.
58. Bartel, D.P., *MicroRNAs: genomics, biogenesis, mechanism, and function*, in *Cell*. 2004: p. 281-97.
59. Lee, Y., et al., *MicroRNA genes are transcribed by RNA polymerase II*, in *EMBO J*. 2004: p. 4051-60.
60. Cai, X., C.H. Hagedorn, and B.R. Cullen, *Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs*, in *RNA*. 2004: p. 1957-66.
61. Lee, Y., et al., *The nuclear RNase III Drosha initiates microRNA processing*, in *Nature*. 2003: p. 415-9.

62. Yi, R., et al., *Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs*, in *Genes Dev.* 2003: p. 3011-6.
63. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation*, in *Cell.* 2011, 2011 Elsevier Inc: p. 646-74.
64. Calin, G.A. and C.M. Croce, *MicroRNA signatures in human cancers*, in *Nat Rev Cancer.* 2006: p. 857-66.
65. Esquela-Kerscher, A. and F.J. Slack, *Oncomirs - microRNAs with a role in cancer*, in *Nat Rev Cancer.* 2006: p. 259-69.
66. Calin, G.A., et al., *Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas*, in *Cancer Cell.* 2007: p. 215-29.
67. O'Donnell, K.A., et al., *c-Myc-regulated microRNAs modulate E2F1 expression*, in *Nature.* 2005: p. 839-43.
68. He, L., et al., *A microRNA polycistron as a potential human oncogene*, in *Nature.* 2005: p. 828-33.
69. Calin, G.A., et al., *Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia*, in *Proc Natl Acad Sci U S A.* 2002: p. 15524-9.
70. Johnson, S.M., et al., *RAS is regulated by the let-7 microRNA family*, in *Cell.* 2005: p. 635-47.
71. Lu, J., et al., *MicroRNA expression profiles classify human cancers*, in *Nature.* 2005: England. p. 834-8.
72. Rosenfeld, N., et al., *MicroRNAs accurately identify cancer tissue origin*, in *Nat Biotechnol.* 2008: p. 462-9.
73. Kulshreshtha, R., et al., *A microRNA signature of hypoxia*, in *Mol Cell Biol.* 2007: p. 1859-67.
74. Selcuklu, S.D., M.C. Yakicier, and A.E. Erson, *An investigation of microRNAs mapping to breast cancer related genomic gain and loss regions.* *Cancer Genet Cytogenet*, 2009. **189**(1): p. 15-23.
75. Zhang, L., et al., *Genomic and epigenetic alterations deregulate microRNA expression in human epithelial ovarian cancer*, in *Proc Natl Acad Sci U S A.* 2008: United States. p. 7004-9.
76. Saito, Y., et al., *Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells*, in *Cancer Cell.* 2006: p. 435-43.
77. Calin, G.A., et al., *A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia*, in *N Engl J Med.* 2005, 2005 Massachusetts Medical Society.: p. 1793-801.
78. Chiosea, S., et al., *Overexpression of Dicer in precursor lesions of lung adenocarcinoma*, in *Cancer Res.* 2007: p. 2345-50.
79. Karube, Y., et al., *Reduced expression of Dicer associated with poor prognosis in lung cancer patients*, in *Cancer Sci.* 2005: p. 111-5.
80. Thomson, J.M., et al., *Extensive post-transcriptional regulation of microRNAs and its implications for cancer*, in *Genes Dev.* 2006: p. 2202-7.
81. Green, S., et al., *Structural and functional domains of the estrogen receptor.* *Cold Spring Harb Symp Quant Biol*, 1986. **51 Pt 2**: p. 751-8.

82. Ascenzi, P., A. Bocedi, and M. Marino, *Structure-function relationship of estrogen receptor alpha and beta: impact on human health*, in *Mol Aspects Med.* 2006:. p. 299-402.
83. Nilsson, S. and J.A. Gustafsson, *Estrogen receptor action*. *Crit Rev Eukaryot Gene Expr*, 2002. **12**(4): p. 237-57.
84. Nilsson, S., et al., *Mechanisms of estrogen action*. *Physiol Rev*, 2001. **81**(4): p. 1535-65.
85. Barone, I., L. Brusco, and S.A. Fuqua, *Estrogen receptor mutations and changes in downstream gene expression and signaling*, in *Clin Cancer Res.* 2010, 2010 Aacr.: p. 2702-8.
86. Landers, J.P. and T.C. Spelsberg, *New concepts in steroid hormone action: transcription factors, proto-oncogenes, and the cascade model for steroid regulation of gene expression*. *Crit Rev Eukaryot Gene Expr*, 1992. **2**(1): p. 19-63.
87. Levin, E.R., *Integration of the extranuclear and nuclear actions of estrogen*, in *Mol Endocrinol.* 2005: p. 1951-9.
88. Bjornstrom, L. and M. Sjoberg, *Mechanisms of estrogen receptor signaling: convergence of genomic and nongenomic actions on target genes*, in *Mol Endocrinol.* 2005: p. 833-42.
89. Lewis-Wambi, J.S. and V.C. Jordan, *Treatment of Postmenopausal Breast Cancer with Selective Estrogen Receptor Modulators (SERMs)*. *Breast Dis*, 2005. **24**: p. 93-105.
90. Leary, A. and M. Dowsett, *Combination therapy with aromatase inhibitors: the next era of breast cancer treatment?*, in *Br J Cancer.* 2006:. p. 661-6.
91. Dorssers, L.C., et al., *Tamoxifen resistance in breast cancer: elucidating mechanisms*. *Drugs*, 2001. **61**(12): p. 1721-33.
92. Moghadam, S.J., A.M. Hanks, and K. Keyomarsi, *Breaking the cycle: An insight into the role of ERalpha in eukaryotic cell cycles*, in *J Carcinog.* 2011: India. p. 25.
93. Barrett, T., et al., *NCBI GEO: archive for functional genomics data sets--10 years on*. *Nucleic Acids Res*, 2011. **39**(Database issue): p. D1005-10.
94. Carroll, J.S., et al., *Genome-wide analysis of estrogen receptor binding sites*. *Nat Genet*, 2006. **38**(11): p. 1289-97.
95. Bourdeau, V., et al., *Mechanisms of primary and secondary estrogen target gene regulation in breast cancer cells*. *Nucleic Acids Res*, 2008. **36**(1): p. 76-93.
96. Musgrove, E.A., et al., *Identification of functional networks of estrogen- and c-Myc-responsive genes and their relationship to response to tamoxifen therapy in breast cancer*. *PLoS One*, 2008. **3**(8): p. e2987.
97. Zhang, H., et al., *PolyA\_DB: a database for mammalian mRNA polyadenylation*. *Nucleic Acids Res*, 2005. **33**(Database issue): p. D116-20.
98. Fujita, P.A., et al., *The UCSC Genome Browser database: update 2011*. *Nucleic Acids Res*, 2011. **39**(Database issue): p. D876-82.
99. Alotaibi, H., et al., *Unliganded estrogen receptor-alpha activates transcription of the mammary gland Na<sup>+</sup>/I<sup>-</sup> symporter gene*, in *Biochem Biophys Res Commun.* 2006: p. 1487-96.
100. Metivier, R., et al., *Transcriptional complexes engaged by apo-estrogen receptor-alpha isoforms have divergent outcomes*, in *EMBO J.* 2004:. p. 3653-66.
101. Akhavantabasi, S., et al., *USP32 is an active, membrane-bound ubiquitin protease overexpressed in breast cancers*. *Mamm Genome*, 2010. **21**(7-8): p. 388-97.

102. Lee, H.O. and Y.Y. Sheen, *Estrogen modulation of human breast cancer cell growth*. Arch Pharm Res, 1997. **20**(6): p. 566-71.
103. Pang, H. and L.E. Faber, *Estrogen and rapamycin effects on cell cycle progression in T47D breast cancer cells*. Breast Cancer Res Treat, 2001. **70**(1): p. 21-6.
104. Dalvai, M. and K. Bystricky, *Cell cycle and anti-estrogen effects synergize to regulate cell proliferation and ER target gene expression*. PLoS One, 2010. **5**(6): p. e11011.
105. Ali, S. and R.C. Coombes, *Endocrine-responsive breast cancer and strategies for combating resistance*. Nat Rev Cancer, 2002. **2**(2): p. 101-12.
106. Bustin, S.A., et al., *The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments*. Clin Chem, 2009. **55**(4): p. 611-22.
107. Gur-Dedeoglu, B., et al., *Identification of endogenous reference genes for qRT-PCR analysis in normal matched breast tumor tissues*. Oncol Res, 2009. **17**(8): p. 353-65.
108. Fleige, S., et al., *Comparison of relative mRNA quantification models and the impact of RNA integrity in quantitative real-time RT-PCR*. Biotechnol Lett, 2006. **28**(19): p. 1601-13.
109. Frieze, S., et al., *CARM1 regulates estrogen-stimulated breast cancer growth through up-regulation of E2F1*, in Cancer Res. 2008: p. 301-6.
110. Elkon, R., et al., *E2F mediates enhanced alternative polyadenylation in proliferation*. Genome Biol, 2012. **13**(7): p. R59.
111. Dignam, J.D., R.M. Lebovitz, and R.G. Roeder, *Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei*. Nucleic Acids Res, 1983. **11**(5): p. 1475-89.
112. Prall, O.W., E.M. Rogan, and R.L. Sutherland, *Estrogen regulation of cell cycle progression in breast cancer cells*. J Steroid Biochem Mol Biol, 1998. **65**(1-6): p. 169-74.
113. Foster, J.S., et al., *Estrogens and cell-cycle regulation in breast cancer*, in Trends Endocrinol Metab. 2001: p. 320-7.
114. Lee, H.O. and Y.Y. Sheen, *Estrogen modulation of human breast cancer cell growth*. Arch Pharm Res, 1997. **20**(6): p. 566-71.
115. Pang, H. and L.E. Faber, *Estrogen and rapamycin effects on cell cycle progression in T47D breast cancer cells*. Breast Cancer Res Treat, 2001. **70**(1): p. 21-6.
116. Dalvai, M. and K. Bystricky, *Cell cycle and anti-estrogen effects synergize to regulate cell proliferation and ER target gene expression*. PLoS One, 2010. **5**(6): p. e11011.
117. Bustin, S.A., et al., *The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments*, in Clin Chem. 2009: p. 611-22.
118. Cocker, J.H., et al., *An essential role for the Cdc6 protein in forming the pre-replicative complexes of budding yeast*. Nature, 1996. **379**(6561): p. 180-2.
119. Donovan, S., et al., *Cdc6p-dependent loading of Mcm proteins onto pre-replicative chromatin in budding yeast*. Proc Natl Acad Sci U S A, 1997. **94**(11): p. 5611-6.
120. Garcia, M., et al., *Activation of estrogen receptor transfected into a receptor-negative breast cancer cell line decreases the metastatic and invasive potential of the cells*. Proc Natl Acad Sci U S A, 1992. **89**(23): p. 11538-42.
121. Jiang, S.Y. and V.C. Jordan, *Growth regulation of estrogen receptor-negative breast cancer cells transfected with complementary DNAs for estrogen receptor*. J Natl Cancer Inst, 1992. **84**(8): p. 580-91.



122. Ferguson, A.T., et al., *Demethylation of the estrogen receptor gene in estrogen receptor-negative breast cancer cells can reactivate estrogen receptor gene expression*. *Cancer Res*, 1995. **55**(11): p. 2279-83.
123. Lazennec, G. and B.S. Katzenellenbogen, *Expression of human estrogen receptor using an efficient adenoviral gene delivery system is able to restore hormone-dependent features to estrogen receptor-negative breast carcinoma cells*, in *Mol Cell Endocrinol*. 1999: Ireland. p. 93-105.
124. Lewis, B.P., C.B. Burge, and D.P. Bartel, *Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets*, in *Cell*. 2005: p. 15-20.
125. Grimson, A., et al., *MicroRNA targeting specificity in mammals: determinants beyond seed pairing*, in *Mol Cell*. 2007: p. 91-105.
126. Garcia, D.M., et al., *Weak seed-pairing stability and high target-site abundance decrease the proficiency of *lscy-6* and other microRNAs*, in *Nat Struct Mol Biol*. 2011: p. 1139-46.
127. Krek, A., et al., *Combinatorial microRNA target predictions*, in *Nat Genet*. 2005: p. 495-500.
128. Ye, W., et al., *The effect of central loops in miRNA:MRE duplexes on the efficiency of miRNA-mediated gene regulation*. *PLoS One*, 2008. **3**(3): p. e1719.
129. van Haften, G. and R. Agami, *Tumorigenicity of the miR-17-92 cluster distilled*. *Genes Dev*, 2010. **24**(1): p. 1-4.
130. Sengupta, S., et al., *MicroRNA 92b controls the G1/S checkpoint gene p57 in human embryonic stem cells*. *Stem Cells*, 2009. **27**(7): p. 1524-8.
131. Zhang, J., et al., *Effects of miR-541 on neurite outgrowth during neuronal differentiation*. *Cell Biochem Funct*, 2011. **29**(4): p. 279-86.
132. Östling, P., et al., *Systematic analysis of microRNAs targeting the androgen receptor in prostate cancer cells*. *Cancer Res*, 2011. **71**(5): p. 1956-67.
133. Gebauer, F., T. Preiss, and M.W. Hentze, *From cis-regulatory elements to complex RNPs and back*. *Cold Spring Harb Perspect Biol*, 2012. **4**(7): p. a012245.
134. Szostak, E. and F. Gebauer, *Translational control by 3'-UTR-binding proteins*, in *Brief Funct Genomics*. 2013: p. 58-65.
135. Klinge, C.M., *Estrogen receptor interaction with estrogen response elements*. *Nucleic Acids Res*, 2001. **29**(14): p. 2905-19.
136. Kushner, P.J., et al., *Estrogen receptor pathways to AP-1*, in *J Steroid Biochem Mol Biol*. 2000: p. 311-7.
137. Bajic, V.B., et al., *Dragon ERE Finder version 2: A tool for accurate detection and analysis of estrogen response elements in vertebrate genomes*. *Nucleic Acids Res*, 2003. **31**(13): p. 3605-7.
138. Broos, S., et al., *ConTra v2: a tool to identify transcription factor binding sites across species, update 2011*, in *Nucleic Acids Res*. 2011: p. W74-8.
139. Obrig, T.G., et al., *The mechanism by which cycloheximide and related glutarimide antibiotics inhibit peptide synthesis on reticulocyte ribosomes*. *J Biol Chem*, 1971. **246**(1): p. 174-81.
140. Wang, C., J. Yu, and C.B. Kallen, *Two estrogen response element sequences near the PCNA gene are not responsible for its estrogen-enhanced expression in MCF7 cells*. *PLoS One*, 2008. **3**(10): p. e3523.

141. DeGregori, J. and D.G. Johnson, *Distinct and Overlapping Roles for E2F Family Members in Transcription, Proliferation and Apoptosis*. *Curr Mol Med*, 2006. **6**(7): p. 739-48.
142. Hateboer, G., et al., *Cell cycle-regulated expression of mammalian CDC6 is dependent on E2F*. *Mol Cell Biol*, 1998. **18**(11): p. 6679-97.
143. Yan, Z., et al., *Cdc6 is regulated by E2F and is essential for DNA replication in mammalian cells*. *Proc Natl Acad Sci U S A*, 1998. **95**(7): p. 3603-8.
144. van den Heuvel, S. and N.J. Dyson, *Conserved functions of the pRB and E2F families*, in *Nat Rev Mol Cell Biol*. 2008:. p. 713-24.
145. Di Paola, D. and M. Zannis-Hadjopoulos, *Comparative analysis of pre-replication complex proteins in transformed and normal cells*. *J Cell Biochem*, 2011.
146. Gonzalez, S., et al., *Oncogenic activity of Cdc6 through repression of the INK4/ARF locus*. *Nature*, 2006. **440**(7084): p. 702-6.
147. Borlado, L.R. and J. Mendez, *CDC6: from DNA replication to cell cycle checkpoints and oncogenesis*. *Carcinogenesis*, 2008. **29**(2): p. 237-43.
148. Auer, H., D.L. Newsom, and K. Kornacker, *Expression Profiling Using Affymetrix GeneChip Microarrays*. *Methods Mol Biol*, 2009. **509**: p. 35-46.
149. D'Mello, V., et al., *Alternative mRNA polyadenylation can potentially affect detection of gene expression by affymetrix genechip arrays*. *Appl Bioinformatics*, 2006. **5**(4): p. 249-53.
150. Klinge, C.M., *miRNAs and estrogen action*. *Trends Endocrinol Metab*, 2012.
151. Nunes, N.M., et al., *A functional human Poly(A) site requires only a potent DSE and an A-rich upstream sequence*. *EMBO J*, 2010. **29**(9): p. 1523-36.
152. Sayers, E.W., et al., *Database resources of the National Center for Biotechnology Information*. *Nucleic Acids Res*, 2009.

## APPENDIX A

### Primers

**Table A. 1. Primers used in the study.**

Primer Name	Primer Sequence (5'→3')	Experiment
GAPDH_F	GGGAGCCAAAAGGGTCATCA	PCR
GAPDH_R	TTTCTAGACGGCAGGTCAGGT	PCR
TFF1_F	CCATGGAGAACAAGGTGATCTGC	RT-PCR
TFF1_R	GTCAATCTGTGTTGTGAGCCGAG	RT-PCR
CDC6_F	TTCAGCTGGCATTAGAGAGC	RT-qPCR
CDC6_R1	AAGGGTCTACCTGGTCACTTTT	RT-qPCR
CDC6_R2	CGCCTCAAAAACAACAACAA	RT-qPCR
SDHA_F	TGGGAACAAGAGGGCATCTG	RT-qPCR
SDHA_R	CCACCACTGCATCAATTCATG	RT-qPCR
PCNA_F	TGCAGATGTACCCCTTGTTG	RT-qPCR
PCNA_R	GCTGGCATCTTAGAAGCAGTT	RT-qPCR
E2F1_F	CCATCCAGGAAAAGGTGTGA	RT-qPCR
E2F1_R	TTCCTGGAGCTGCTGAGC	RT-qPCR
E2F2_F	ACTGGAAGTGCCCGACAGGA	RT-qPCR
E2F2_R	GGTGGAGGTAGAGGGGAGAGG	RT-qPCR
CSTF2_F	ATCCTTGCCTGCGAATGTCC	RT-qPCR
CSTF2_R	GGGTGGTCCCTCGGCTCTC	RT-qPCR
CSTF3_F	TGAAGTGGATAGAAAACCAGAATACCC	RT-qPCR
CSTF3_R	TGGAAACAGATAGGAGGAGGGAGA	RT-qPCR
CPSF2_F	CTTTGGAACCCTTGCCACCT	RT-qPCR
CPSF2_R	TGCGGACTGCTACTTGATTGTTG	RT-qPCR
oligo dT anchor	GACCACGCGTATCGATGTCGACTTTTTTTTTT TTTTTV	3' RACE
Anchor_R	GACCACGCGTATCGATGTCGAC	3' RACE
3'RACE_1	GCTCTTGGAAGCCAGGGGCATTTTA	3' RACE
3'RACE_2	CCACCCGAAAGTATTCAGCTGGCATTTA	3' RACE
CDC6_pMIR_F	CGAGCTCGATTCTTCTTACACCCAC	Cloning
CDC6_pMIR_R1	CCCAAGCTTGGGAAAATACCACTCATGTTT GAG	Cloning
CDC6_pMIR_R2	CCCAAGCTTGGGAACTTGAAAATAAATATA TTC	Cloning
pMIR_F	AGGCGATTAAGTTGGGTA	Colony PCR
pMIR_R	GGAAAGTCCAAATTGCTC	Colony PCR

## APPENDIX B

**Table B.2. MIQE Guidelines Checklist**

ITEM TO CHECK	IMPORTANCE	CHECKLIST
<b>EXPERIMENTAL DESIGN</b>		
Definition of experimental and control groups	E	YES
Number within each group	E	YES
Assay carried out by core lab or investigator's lab?	D	YES
Acknowledgement of authors' contributions	D	NO
<b>SAMPLE</b>		
Description	E	N/A
Volume/mass of sample processed	D	N/A
Microdissection or macrodissection	E	N/A
Processing procedure	E	N/A
If frozen - how and how quickly?	E	N/A
If fixed - with what, how quickly?	E	N/A
Sample storage conditions and duration (especially for FFPE samples)	E	N/A
<b>NUCLEIC ACID EXTRACTION</b>		
Procedure and/or instrumentation	E	YES
Name of kit and details of any modifications	E	YES
Source of additional reagents used	D	YES
Details of DNase or RNase treatment	E	YES
Contamination assessment (DNA or RNA)	E	YES
Nucleic acid quantification	E	YES
Instrument and method	E	YES
Purity (A260/A280)	D	YES
Yield	D	YES
RNA integrity method/instrument	E	YES
RIN/RQI or Cq of 3' and 5' transcripts	E	NO
Electrophoresis traces	D	N/A
Inhibition testing (Cq dilutions, spike or other)	E	YES

**Table B.2. Continued. MIQE Guidelines Checklist**

<b>REVERSE TRANSCRIPTION</b>		
Complete reaction conditions	E	YES
Amount of RNA and reaction volume	E	YES
Priming oligonucleotide and concentration	E	YES
Reverse transcriptase and concentration	E	YES
Temperature and time	E	YES
Manufacturer of reagents and catalogue numbers	D	YES
Cqs with and without RT	D	NO
Storage conditions of cDNA	D	YES
<b>qPCR TARGET INFORMATION</b>		
If multiplex, efficiency and LOD of each assay.	E	N/A
Sequence accession number	E	YES
Location of amplicon	D	YES
Amplicon length	E	YES
<i>In silico</i> specificity screen (BLAST, etc)	E	YES
Pseudogenes, retropseudogenes or other homologs?	D	YES
Sequence alignment	D	YES
Secondary structure analysis of amplicon	D	NO
Location of each primer by exon or intron (if applicable)	E	YES
What splice variants are targeted?	E	N/A
<b>qPCR OLIGONUCLEOTIDES</b>		
Primer sequences	E	YES
RTPrimerDB Identification Number	D	N/A
Probe sequences	D	N/A
Location and identity of any modifications	E	N/A
Manufacturer of oligonucleotides	D	YES
Purification method	D	YES

**Table B.2. Continued. MIQE Guidelines Checklist**

<b>qPCR PROTOCOL</b>		
Complete reaction conditions	E	YES
Reaction volume and amount of cDNA/DNA	E	YES
Primer, (probe), Mg <sup>++</sup> and dNTP concentrations	E	YES
Polymerase identity and concentration	E	YES
Buffer/kit identity and manufacturer	E	YES
Exact chemical constitution of the buffer	D	NO
Additives (SYBR Green I, DMSO, etc.)	E	N/A
Manufacturer of plates/tubes and catalog number	D	YES
Complete thermocycling parameters	E	YES
Reaction setup (manual/robotic)	D	YES
Manufacturer of qPCR instrument	E	YES
<b>qPCR VALIDATION</b>		
Evidence of optimization (from gradients)	D	
Specificity (gel, sequence, melt, or digest)	E	YES
For SYBR Green I, C <sub>q</sub> of the NTC	E	YES
Standard curves with slope and y-intercept	E	YES
PCR efficiency calculated from slope	E	YES
Confidence interval for PCR efficiency or standard error	D	N/A
r <sup>2</sup> of standard curve	E	YES
Linear dynamic range	E	YES
C <sub>q</sub> variation at lower limit	E	YES
Confidence intervals throughout range	D	N/A
Evidence for limit of detection	E	NO
If multiplex, efficiency and LOD of each assay.	E	N/A
<b>DATA ANALYSIS</b>		
qPCR analysis program (source, version)	E	YES
C <sub>q</sub> method determination	E	YES
Outlier identification and disposition	E	N/A
Results of NTCs	E	YES
Justification of number and choice of reference genes	E	YES
Description of normalization method	E	YES
Number and concordance of biological replicates	D	YES
Number and stage (RT or qPCR) of technical replicates	E	YES
Repeatability (intra-assay variation)	E	YES
Reproducibility (inter-assay variation, %CV)	D	NO
Power analysis	D	NO
Statistical methods for result significance	E	YES
Software (source, version)	E	YES
C <sub>q</sub> or raw data submission using RDML	D	N/A

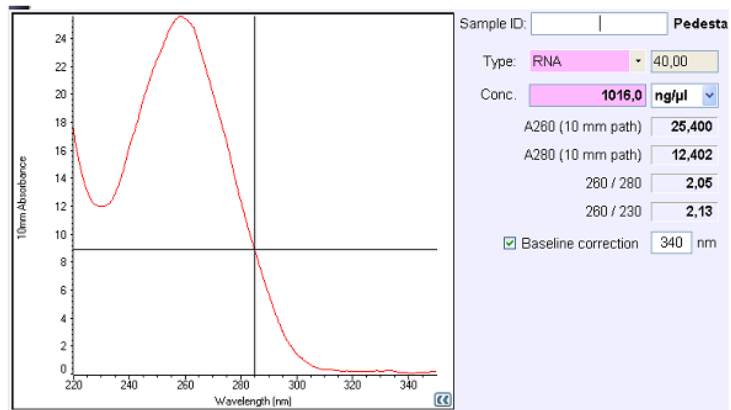
E: Essential information, D: Desirable information, N/A: Not applicable

## APPENDIX C

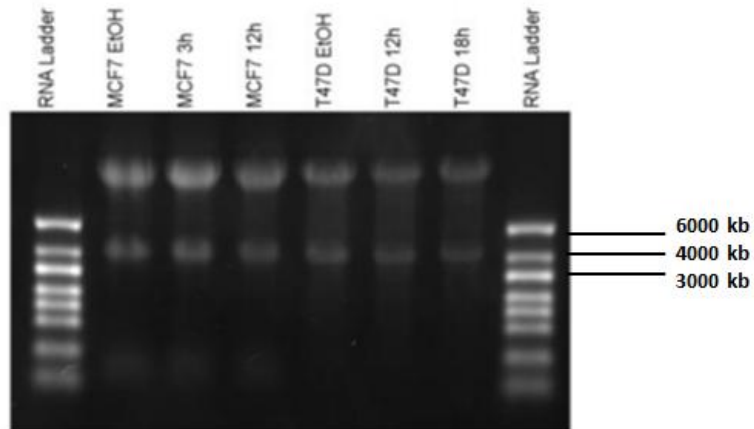
### DNA Contamination, RNA Quantification and Integrity Assessment



**Figure C. 1. Lack of DNA contamination in RNA samples was assessed.** PCR was performed using *GAPDH* specific primers. *GAPDH\_F*: 5'-GGGAGCCAAAAGGGTCATCA-3' and *GAPDH\_R*: 5'-TTTCTAGACGGCAGGTCA GGT-3' (product size: 409 bp). Following conditions were used for the PCR reactions: incubation at 94°C for 10 minutes, 40 cycles of 94°C for 30 seconds, 56°C for 30 seconds, and 72°C for 30 seconds, and final extension at 72°C for 5 minutes. MCF7 cDNA was used as a positive control.



**Figure C. 2. RNA concentrations were determined using NanoDrop ND1000 (Thermo Scientific).** Purity was determined by A260/A280 and A260/A230 ratios.



**Figure C. 3. RNA integrity assessment.** RNA integrities were confirmed by running 5 ug RNA samples on a 0.7 % Formaldehyde RNA Gel (10X MOPS, 37% Formaldehyde, DEPC-water). All the RNA samples from MCF7 and T47D cells were showing intact 28S and 18S rRNA bands. RNA ladder (Thermo Scientific).

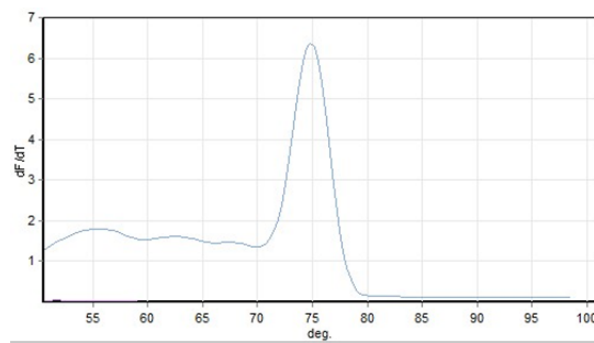


## APPENDIX D

### RT-qPCR Assay Performance Results

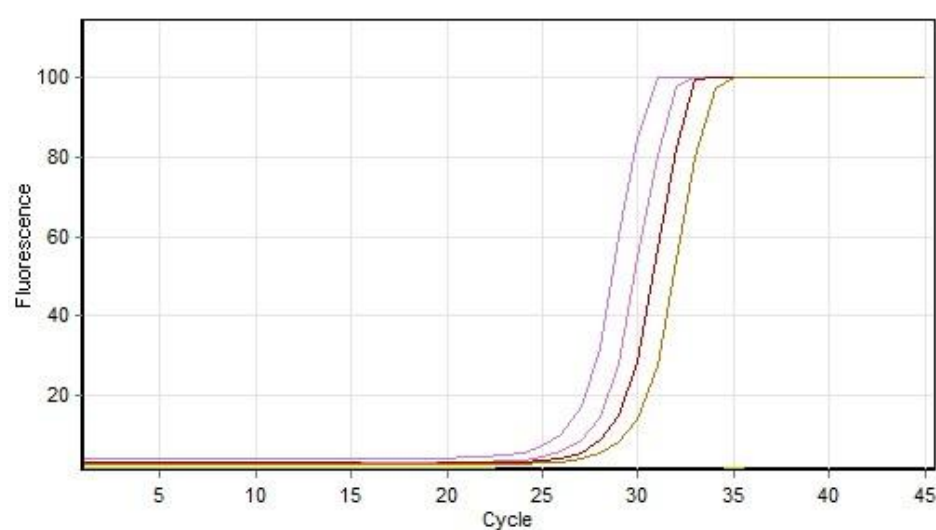
#### CDC6\_Short Primers RT-qPCR Assay Performance Results

Melt Analysis for CDC6\_Short primers

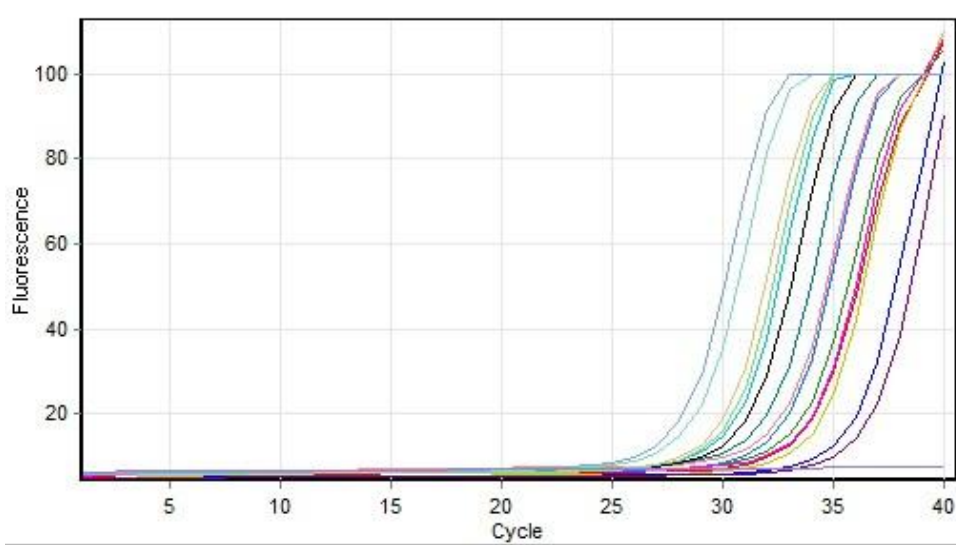


No	Colour	Name	Genotype	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5	Peak 6	Peak 7	Peak 8	Peak 9	Peak 10
1	Blue	CDC6 Short		55,5	62,5	67,5	74,8	87,0	97,0				
2	Purple	CDC6 Short NTC		51,7	57,0	61,7	67,2	72,7	76,0	81,8	87,0	92,5	97,0

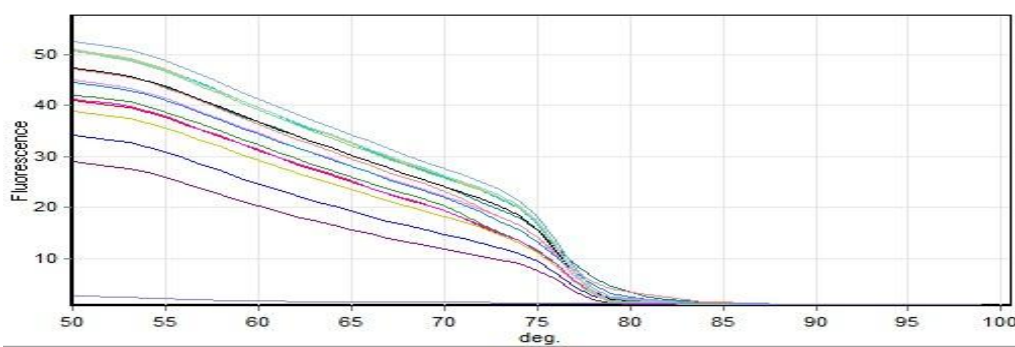
Raw Data For Cycling A.Green








Quantitation data for Cycling A.Green



Standard Curve



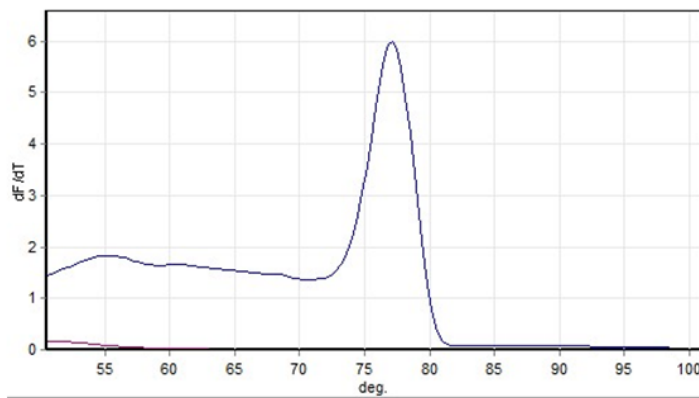
No.	Colour	Name	Type	Ct	Given Conc (copies/ul)	Calc Conc (copies/ul)	% Var
18		Std 1_CDC6 Short	Standard	22,84	100.000	100.501	0,5%
19		Std 2_CDC6 Short	Standard	23,85	50.000	48.956	2,1%
20		Std 3_CDC6 Short	Standard	24,75	25.000	25.689	2,8%
21		Std 4_CDC6 Short	Standard	25,77	12.500	12.362	1,1%
34		NTC_CDC6 Short	NTC				

## Quantitation Information

Threshold	0,1232
Left Threshold	1,000
Standard Curve Imported	No
Standard Curve (1)	conc= 10 <sup>^</sup> (-0,311*CT + 12,108)
Standard Curve (2)	CT = -3,215*log(conc) + 38,925
Reaction efficiency (*)	1,04669 (* = 10 <sup>^</sup> (-1/m) - 1)
M	-3,21489
B	38,92467
R Value	0,99972
R <sup>2</sup> Value	0,99945
Start normalising from cycle	1
Noise Slope Correction	No
No Template Control Threshold	0%
Reaction Efficiency Threshold	Disabled
Normalisation Method	Dynamic Tube Normalisation
Digital Filter	Light
Sample Page	Page 1
Imported Analysis Settings	

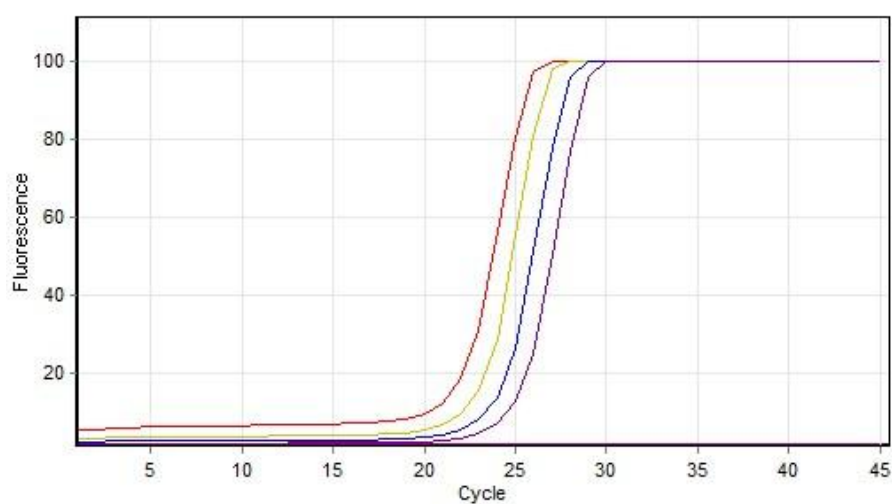
## CDC6\_Long Primers RT-qPCR Assay Performance Results

Melt Analysis for *CDC6*\_Long primers

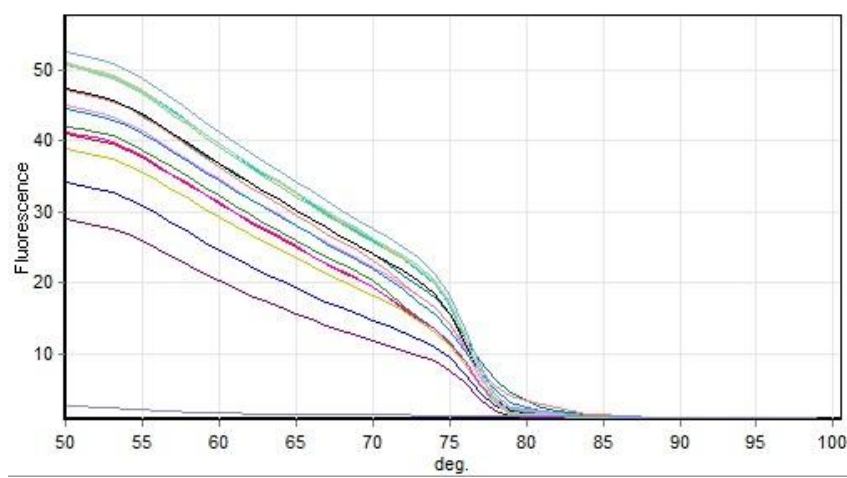


No.	Colour	Name	Genotype	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5	Peak 6
1	Blue	CDC6 Long		55,3	60,5	77,0	83,0	90,3	95,5
2	Magenta	CDC6 Long NTC		68,0	74,8	82,5	87,0	91,2	94,5

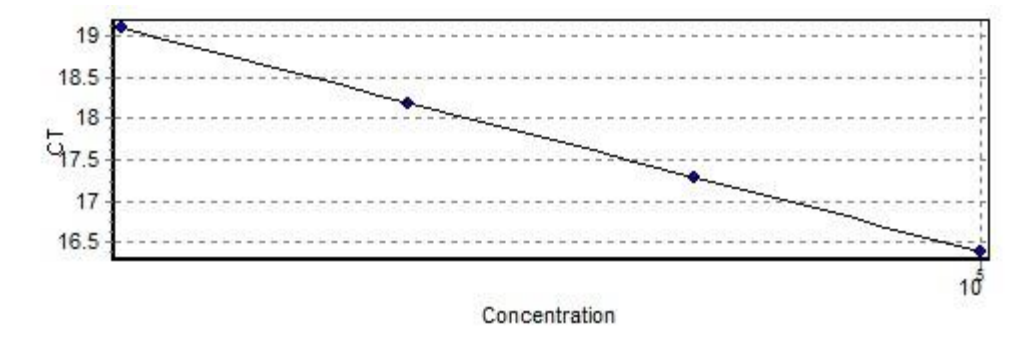
Raw Data For Cycling A.Green



Quantitation data for Cycling A.Green



Standard Curve



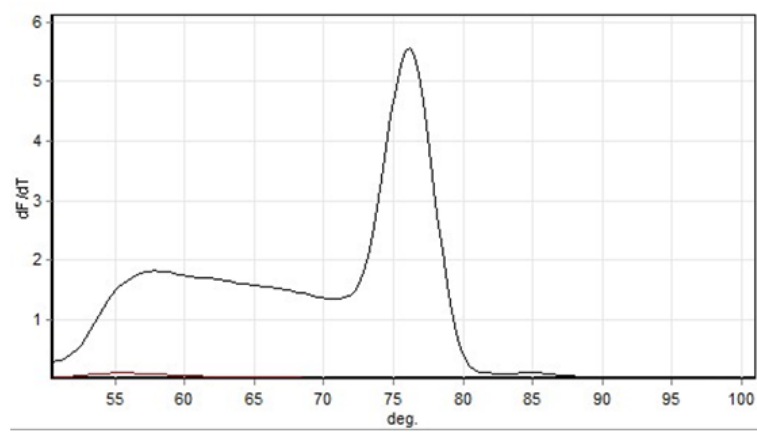
No.	Colour	Name	Type	Ct	Given Conc (copies/ul)	Calc Conc (copies/ul)	% Var
1	Red	Std 1_CDC6 Long	Standard	16,39	100.000	99.959	0,0%
2	Yellow	Std 2_CDC6 Long	Standard	17,29	50.000	49.894	0,2%
3	Blue	Std 3_CDC6 Long	Standard	18,19	25.000	25.137	0,5%
4	Purple	Std 4_CDC6 Long	Standard	19,10	12.500	12.463	0,3%
36	Purple	NTC	NTC				

## Quantitation Information

Threshold	0,0554
Left Threshold	1,000
Standard Curve Imported	No
Standard Curve (1)	$\text{conc} = 10^{(-0,333 \cdot \text{CT} + 10,465)}$
Standard Curve (2)	$\text{CT} = -2,999 \cdot \log(\text{conc}) + 31,384$
Reaction efficiency (*)	1,15494 (* = $10^{(-1/m)} - 1$ )
M	-2,99908
B	31,3844
R Value	0,99999
R^2 Value	0,99998
Start normalising from cycle	1
Noise Slope Correction	No
No Template Control Threshold	0%
Reaction Efficiency Threshold	Disabled
Normalisation Method	Dynamic Tube Normalisation
Digital Filter	Light
Sample Page	Page 1
Imported Analysis Settings	

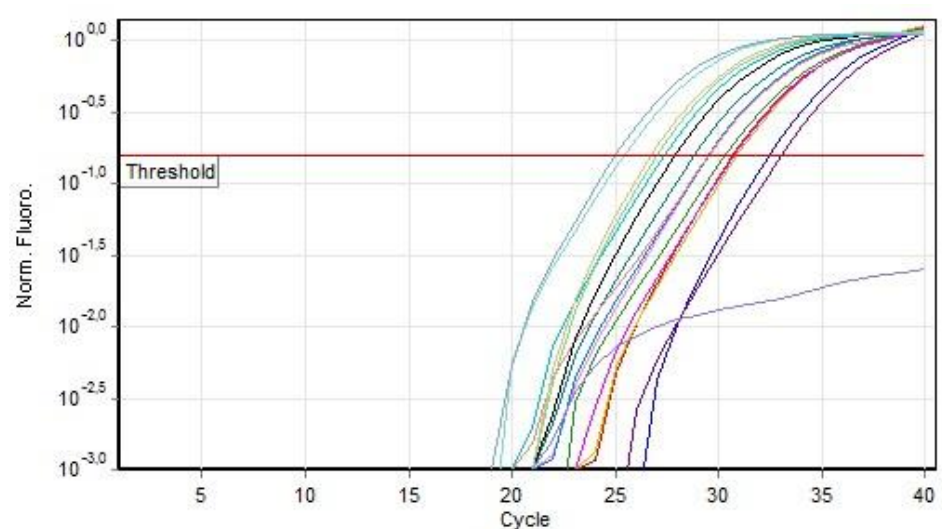
## SDHA Primers RT-qPCR Assay Performance Results

Melt Analysis for *SDHA* primers

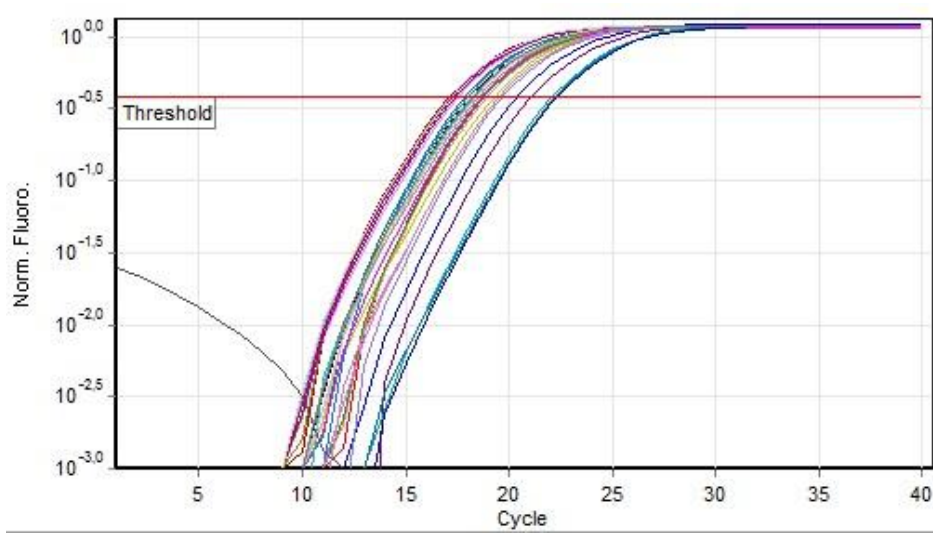


No.	Colour	Name	Genotype	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5	Peak 6	Peak 7
1	■	SDHA		58,0	76,0	85,0				
2	■	SDHANTC		55,3	66,7	75,3	81,5	85,5	90,3	96,5

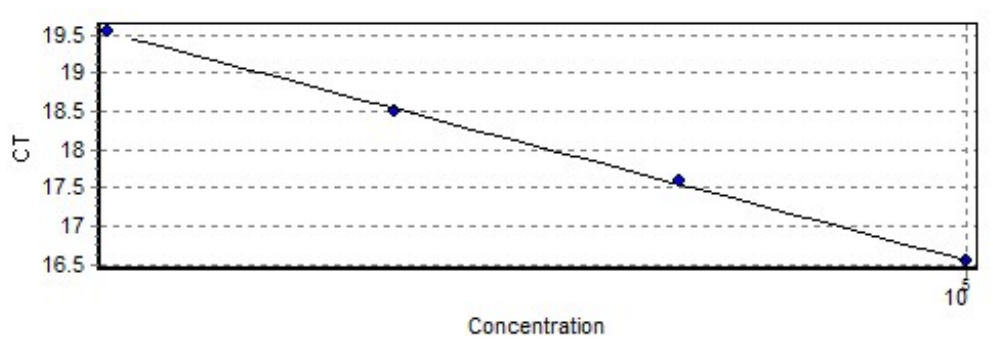
Raw Data For Cycling A.Green



Quantitation data for Cycling A.Green



Standard Curve



No.	Colour	Name	Type	Ct	Given Conc (copies/ul)	Calc Conc (copies/ul)	% Var
1	Red	Std 1	Standard	16,54	100.000	100.703	0,7%
2	Yellow	Std 2	Standard	17,58	50.000	48.899	2,2%
3	Blue	Std 3	Standard	18,50	25.000	25.594	2,4%
4	Purple	Std 4	Standard	19,54	12.500	12.397	0,8%
5	Grey	NTC	NTC				



## Quantitation Information

Threshold	0,111
Left Threshold	1,000
Standard Curve Imported	No
Standard Curve (1)	conc= $10^{(-0,303*CT + 10,019)}$
Standard Curve (2)	CT = $-3,298*\log(\text{conc}) + 33,039$
Reaction efficiency (*)	1,01023 (* = $10^{(-1/m)} - 1$ )
M	-3,29766
B	33,03911
R Value	0,99976
R^2 Value	0,99952
Start normalising from cycle	1
Noise Slope Correction	No
No Template Control Threshold	0%
Reaction Efficiency Threshold	Disabled
Normalisation Method	Dynamic Tube Normalisation
Digital Filter	Light
Sample Page	Page 1
Imported Analysis Settings	

## APPENDIX E

### Buffers

#### Buffer A:

20 mM HEPES

10 mM KCl

0.1 mM EDTA

1 mM DTT

1.5 mM MgCl<sub>2</sub>

0.5 mM NaCl

Protease inhibitor cocktail

#### Buffer B:

20 mM HEPES

10 mM KCl

0.1 mM EDTA

1 mM DTT

1.5 mM MgCl<sub>2</sub>

0.5 mM NaCl

25% glycerol

Protease inhibitor cocktail

**RIPA Buffer:**

50 mM Tris base

150 mM NaCl

0.5 % Sodium deoxycholate

0.1 % SDS

1% Tergitol

**6X Laemmli Buffer:**

12% SDS

30% 2-mercaptoethanol

60% Glycerol

0.012% bromophenol blue

0.375 M Tris

**TBS-T:**

20 mM Tris

137 mM NaCl

0.1% Tween 20

pH: 7.6

APPENDIX F

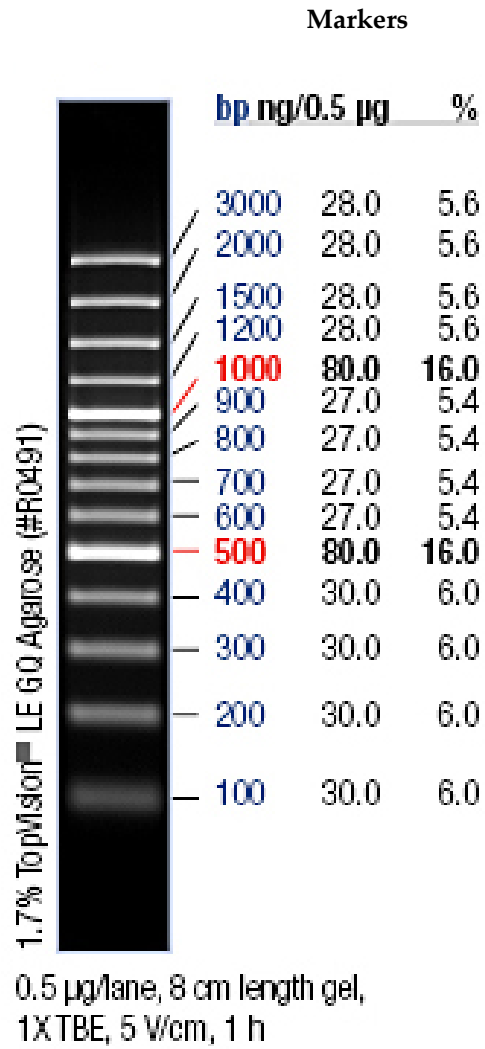
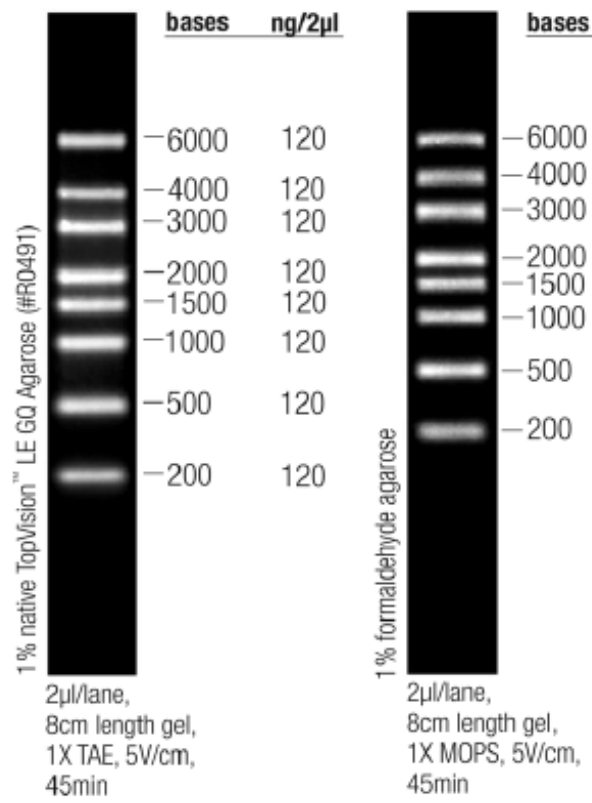
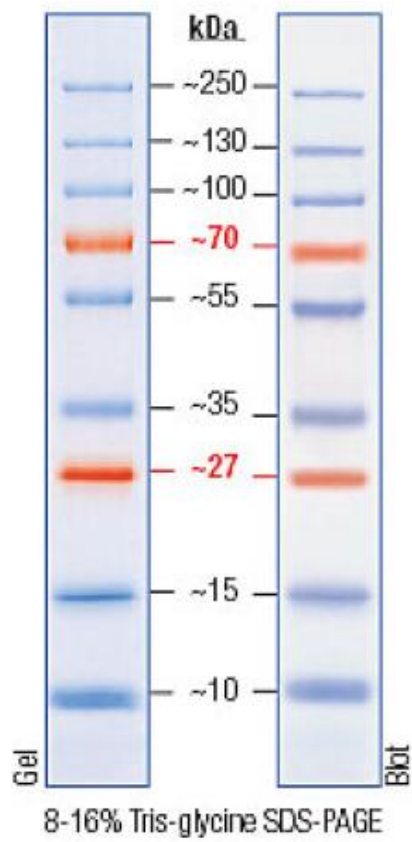


Figure F. 1. GeneRuler 100 bp DNA Ladder Plus.



**Figure F. 2. RiboRuler RNA Ladder.**



**Figure F. 3. PageRuler Pre-stained Protein Ladder.** Fermentas SM1811 Protein ladder was used during Western Blot experiments.

## APPENDIX G

### Plasmid Maps

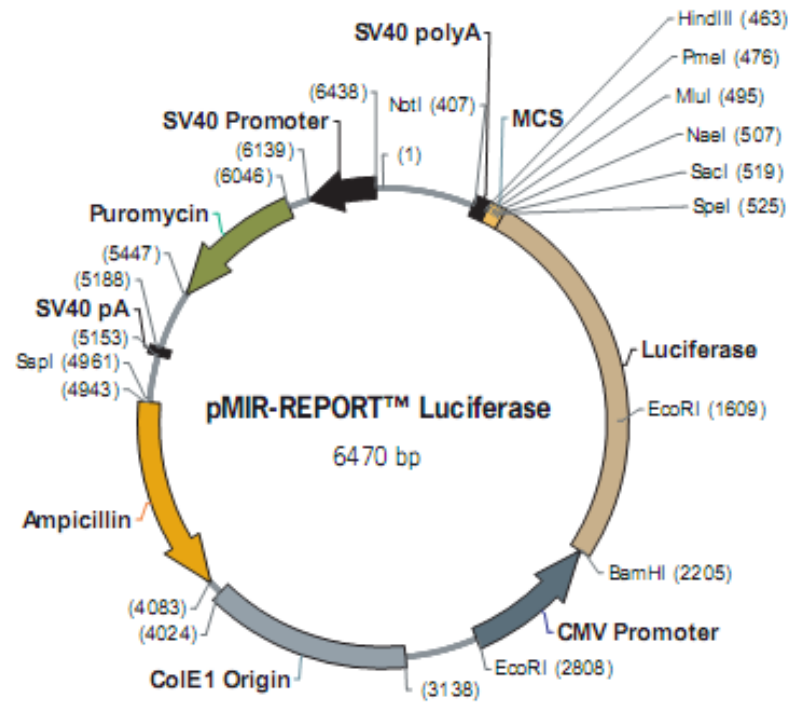


Figure G. 1. pMIR-Report Map

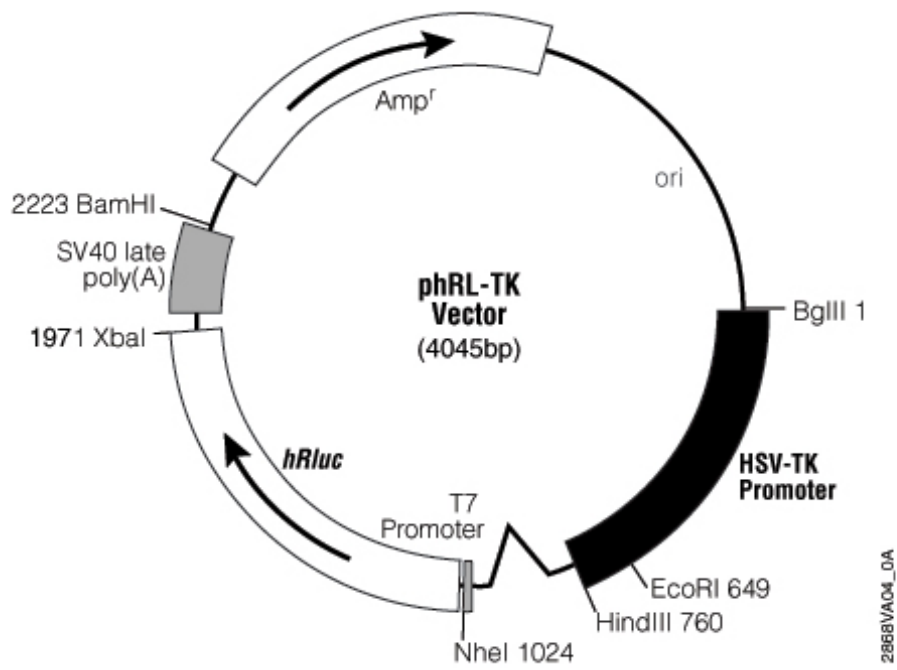


Figure G. 2. phRL-TK Map



## APPENDIX H

### Sequencing Results

```
>ref|NM_001254.3| GM Homo sapiens cell division cycle 6 homolog (S. cerevisiae) (CDC6), mRNA
Length=3053

GENE ID: 990 CDC6 | cell division cycle 6 homolog (S. cerevisiae)
[Homo sapiens] (Over 10 PubMed links)

Score = 630 bits (341), Expect = 4e-178
Identities = 341/341 (100%), Gaps = 0/341 (0%)
Strand=Plus/Minus

Query 115   AAAATACCCACTCATGTTTGAGCAAACGAAGAGGTAAATATACACACATTCCCTCTACACA 174
          |||
Sbjct 2234   AAAATACCCACTCATGTTTGAGCAAACGAAGAGGTAAATATACACACATTCCCTCTACACA 2175

Query 175   GAICTGTACACTTGCAAGCACATTGGCTAGAGATACACAAGTGGTAGAAGTAGTGAATGT 234
          |||
Sbjct 2174   GAICTGTACACTTGCAAGCACATTGGCTAGAGATACACAAGTGGTAGAAGTAGTGAATGT 2115

Query 235   AATTAAAAAGGGTCTACCTGGTCACTTTTATGGGTAAAAATAGTAAGACCCAAAGATATT 294
          |||
Sbjct 2114   AATTAAAAAGGGTCTACCTGGTCACTTTTATGGGTAAAAATAGTAAGACCCAAAGATATT 2055

Query 295   ATCTGTGCTAATATTAAGAATCTATAGATTAATTCATTGGCTTCAAGTAAAAAAGG 354
          |||
Sbjct 2054   ATCTGTGCTAATATTAAGAATCTATAGATTAATTCATTGGCTTCAAGTAAAAAAGG 1995

Query 355   TCATATTTGTTTTCAGGCCCGAATGTGTAAAGCACTAAAAATGAAGACTGTAGCTCTCTAA 414
          |||
Sbjct 1994   TCATATTTGTTTTCAGGCCCGAATGTGTAAAGCACTAAAAATGAAGACTGTAGCTCTCTAA 1935

Query 415   ATGCCAGCTGAATACTTTTCGGGTGGGGTGTAAAGAGAAGAAAT 455
          |||
Sbjct 1934   ATGCCAGCTGAATACTTTTCGGGTGGGGTGTAAAGAGAAGAAAT 1894
```

**Figure H. 1. CDC6 short 3'-UTR isoform sequencing result.** After cloning pMIR construct was sent to Mclab (USA) for sequencing. Sequencing results were aligned through BLAST [152].

GENE ID: 990 CDC6 | cell division cycle 6 homolog (S. cerevisiae)  
 [Homo sapiens] (Over 10 PubMed links)

Score = 1589 bits (860), Expect = 0.0  
 Identities = 878/891 (99%), Gaps = 3/891 (0%)  
 Strand=Plus/Minus

```

Query 113  AACTTGAAAATAAATATATTCACAATCTGTAGAGTTATAGTCTCCTTATGACCCCAACGC 172
          |||
Sbjct 2808  AACTTGAAAATAAATATATTCACAATCTGTAGAGTTATAGTCTCCTTATGACCCCAACGC 2749

Query 173  CCTTGGGAGTGCCCAAGAACTCAACGAGTCTAAAATGCCACAATGTTGTAGTATCTGCA 232
          |||
Sbjct 2748  CCTTGGGAGTGCCCAAGAACTCAACGAGTCTAAAATGCCACAATGTTGTAGTATCTGCA 2689

Query 233  CCATGAACATAGTATACTATACAAAGAAATAAAATCTTTAACCCAGTGTCTCCAAATCTC 292
          |||
Sbjct 2688  CCATGAACATAGTATACTATACAAAGAAATAAAATCTTTAACCCAGTGTCTCCAAATCTC 2629

Query 293  CAITGGGACTTGTCTGCAATTTTCTGGATCAATTTTCATTCTTCTTACCCTAAACTTAAAG 352
          |||
Sbjct 2628  CAITGGGACTTGTCTGCAATTTTCTGGATCAATTTTCATTCTTCTTACCCTAAACTTAAAG 2569

Query 353  TTCTTAAACAAAATACCTTAATGGCTGAGCATGGTGGCTCACGCCTATAATCCCAGCACTT 412
          |||
Sbjct 2568  TTCTTAAACAAAATACCTTAATGGCTGAGCATGGTGGCTCACGCCTATAATCCCAGCACTT 2509

Query 413  AGGGAGGCCAAGGTGGGCAGATCACTTGAGGGTCAGGAGTTCAGACCAGCCTGGCCAAC 472
          |||
Sbjct 2508  AGGGAGGCCAAGGTGGGCAGATCACTTGAGGGTCAGGAGTTCAGACCAGCCTGGCCAAC 2449

Query 473  ATGGTAAAACCCCTGTCTCTACTAAAAATTAAAAAATTAGCTGGGCGCGGTGGTGGGCACC 532
          |||
Sbjct 2448  ATGGTAAAACCCCTGTCTCTACTAAAAATTAAAAAATTAGCTGGGCGCGGTGGTGGGCACC 2389

Query 533  TGTAAATCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATCACTTCAACCTGGGAAGCGGGT 592
          |||
Sbjct 2388  TGTAAATCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATCACTTCAACCTGGGAAGCGGGT 2329

Query 593  GCTGTAGTGAGCAGAGAACCGCCATTGCACTCCAGCCTGGGNAACAGGGTGAGACGCGC 652
          |||
Sbjct 2328  GCTGTAGTGAGCAGAGAACCGCCATTGCACTCCAGCCTGGGNAACAGGGTGAGACGCGC 2269

Query 653  CTCAAAAACAACAACAACAAAAAACAACAACAAAAAATACCCACTCATGTTTGAGCAAA 712
          |||
Sbjct 2268  CTCAAAAACAACAACAACAAAAAACAACAACAAAAAATACCCACTCATGTTTGAGCAAA 2209

Query 713  CGAAGAGGTAATAATACACACATTCTCTACACAGATCTGTACACTTGCAAGCACATTGG 772
          |||
Sbjct 2208  CGAAGAGGTAATAATACACACATTCTCTACACAGATCTGTACACTTGCAAGCACATTGG 2149

Query 773  CTAGAGATACACAAGTGGTAGAAGTAGTGAATGTAATTAATAAGGGTCTACCTGGTCACT 832
          |||
Sbjct 2148  CTAGAGATACACAAGTGGTAGAAGTAGTGAATGTAATTAATAAGGGTCTACCTGGTCACT 2089

Query 833  TTTATGGGTAAAAATAGTAAGACCCAAAGATATTATTCTGTGCTAATATTAAGAATCTA 892
          |||
Sbjct 2088  TTTATGGGTAAAAATAGTAAGACCCAAAGATATTATTCTGTGCTAATATTAAGAATCTA 2029

Query 893  TAGATTAATAATTCATTGNNTCAAGTAAAAAAGGTCATATTGTTTTCNNGNNN-AAATGT 951
          |||
Sbjct 2028  TAGATTAATAATTCATTGGCTTCAAGTAAAAAAGGTCATATTGTTTTCAGGCCCGAATGT 1969

Query 952  GTAAAGCACTAAAATGNAGACTGTAGCTCTCTAAATGNACAGCTGA-TACT 1001
          |||
Sbjct 1968  GTAAAGCACTAAAATGAAGACTGTAGCTCTCTAAATGC-CAGCTGAATACT 1919
  
```

Figure H. 2. CDC6 long 3'-UTR isoform sequencing result. After cloning pMIR construct was sent to Mclab (USA) for sequencing. Sequencing results were aligned through BLAST [152].

## CURRICULUM VITAE

### H. Begüm Akman

Middle East Technical University,  
Department of Biological Sciences,  
Lab: B56  
06800 Ankara/TURKEY  
Tel: 0090 3122105044  
Phone: 0090 5327624938  
Fax: 0090 3122107976  
[begumakman@gmail.com](mailto:begumakman@gmail.com)

#### **Education:**

**PhD** 2007-present

Department of Biological Sciences

Middle East Technical University, Ankara, Turkey

**MSc with honors** 2004-2007

Department of Biological Sciences

Middle East Technical University, Ankara, Turkey

**BSc with honors** 2001-2004

Department of Biological Sciences

Middle East Technical University, Ankara, Turkey

#### **Employment Record**

**Research Assistant** 2005-2012

Department of Biological Sciences

Middle East Technical University, Ankara, Turkey

## Teaching Assistant

2005-2012

Department of Biological Sciences

Middle East Technical University, Ankara, Turkey

- GENE475- Molecular Biology Laboratory
- BIOL310- Biochemistry Laboratory
- GENE367- General and Molecular Biology Laboratory
- GENE405- Animal Cell Culture Techniques Laboratory

Department of Biochemistry

2008-2009

University of College Cork, Cork, Ireland

- BC3010- Bioinformatics Practical Course

## Publications:

1. **Akman BH**, Can T, Erson-Bensan AE. Estrogen-induced upregulation and 3'-UTR shortening of CDC6. *Nucleic Acids Res.* 2012 Sep 12; Vol. 40, pp 10679-10688.
2. Akhavantabasi S\*, **Akman HB\***, Sapmaz A\*, Keller J, Petty EM, Erson AE. USP32 is an active, membrane bound ubiquitin protease overexpressed in breast cancers. *Mamm Genome.* 2010 Aug; 21(7-8): 388-397. \*Equal contribution

## Online Publications:

1. **Akman B**, Erson AE. MIRN10B (microRNA 10b). *Atlas Genet Cytogenet Oncol Haematol.* June 2008  
URL :<http://AtlasGeneticsOncology.org/Genes/MIRN10BID44292ch2q31.html>

## Abstracts and Presentations:

1. **Akman BH**, Can T, Erson-Bensan AE. Estrogen induced upregulation and 3' UTR shortening of *CDC6* by Alternative Polyadenylation, 4th EMBO Meeting, Nice, France, September 22-25, 2012
2. **Akman BH**, Can T, Erson-Bensan AE. Shortening of CDC6 3'UTR by Alternative Polyadenylation Triggered by Estrogen in Breast Cancer Cells (oral presentation), EMBO Young Scientists Forum, Istanbul, Turkey, June 14-16, 2012.

3. **Akman BH**, Erson-Bensan AE. Silencing of a ubiquitin specific protease; *USP32* in cancer cell lines, European Human Genetics Conference, Amsterdam, Holland, May 28-31, 2011
4. **H.B. Akman**, A. Sapmaz , S. Akhavantabasi , A.E. Erson. Characterization of a ubiquitin specific protease in the 17q23 amplicon. The European Association for Cancer Research 21st meeting, June 26-29 2010, Oslo, Norway
5. S. Duygu Selcuklu, Mark T.A Donoghue, **H. Begum Akman**, M. Cengiz Yakicier, A. Elif Erson and Charles Spillane. Expression profiling of miR-21 sensitive MCF-7 transcriptome leads to deregulation of cancer-related genes and pathways, MicroRNA and Cancer Meeting, Keystone Symposia, Keystone, Colorado USA, June 10 - 15, 2009
6. **HB. Akman** and A. E. Erson, "Structural analysis of a 50 kb region on 17q23" HIBIT'09 Ankara, Turkey April 16-17, 2009
7. **Akman B**, Erson AE, Potential Roles of Non-coding RNAs in Regulation of an Oncogene Candidate(oral presentation), ESF Non-Coding RNAs: Computational Challenges and Applications, Antalya, Turkey, April 28-30, 2008.
8. **Akman B**, Erson AE, Structural Analysis of a 50 kb Region on 17q23, EMBO YOUNG SCIENTISTS FORUM, Istanbul, Turkey, Feb 20-22, 2008.

#### **Training Courses:**

- Advanced qPCR Techniques for Publication Success: Following MIQE Guidelines, EMBL Heidelberg, Germany July 11-15, 2011.

#### **Awards and Honors:**

- PhD scholarship from The Scientific and Technological Research Council of Turkey (TUBITAK) (2007-2012)
- Publication award from Middle East Technical University for Nucleic Acids Res. 2012 Sep 12 doi:10.1093/nar/gks855
- Publication award from TUBITAK for Nucleic Acids Res. 2012 Sep 12 doi:10.1093/nar/gks855
- Publication award from Middle East Technical University for Mamm Genome. 2010 Aug; 21(7-8): 388-397
- Publication award from TUBITAK for Mamm Genome. 2010 Aug; 21(7-8): 388-397
- Travel grant for European Human Genetics Conference, Amsterdam, May 28-31, 2011

This thesis represents portions of a published manuscript with permission from Oxford Journals; Akman, B.H., Can, T. and Erson-Bensan, A.E. (2012), *Nucleic Acids Res*, Vol. 40, pp. 10679-10688.