3'UTR ALTERATIONS IN CONNECTION WITH ESTROGEN RECEPTOR-ALPHA IN BREAST CANCER

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

BY

MERVE ÖYKEN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
BIOLOGY

FEBRUARY 2014

Approval of the thesis:

3’UTR ALTERATIONS IN CONNECTION WITH ESTROGEN RECEPTOR-ALPHA IN BREAST CANCER

submitted by **MERVE ÖYKEN** in partial fulfillment of the requirements for the degree of **Master of Sciencein BiologyDepartment, Middle East Technical University**by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural andAppliedSciences**

Prof. Dr. Gülay Özcengiz
Head of Department, **Biology**

Assoc. Prof. Dr. A.Elif Erson-Bensan
Supervisor, **BiologyDept., METU**

ExaminingCommitteeMembers:

Assoc. Prof. Dr. Mesut Muyan
BiologyDept., METU

Assoc. Prof. Dr. A.Elif Erson-Bensan
BiologyDept., METU

Assoc. Prof. Dr. SreeparnaBanerjee
BiologyDept., METU

Assist. Prof. Dr. Mehmet Somel
BiologyDept., METU

Assoc. Prof. Dr. Özlen Konu
MolecularBiologyand Genetics Dept.,Bilkent University

Date: 06.02.2014

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: MerveÖyken

Signature:

ABSTRACT

3'UTR ALTERATIONS IN CONNECTION WITH ESTROGEN RECEPTOR-ALPHA IN BREAST CANCER

Öyken, Merve
MSc., Department of Biology
Supervisor: Assoc. Prof. Dr. A. ElifErson-Bensan

February 2014, 110 pages

Alternative polyadenylation (APA) generates transcript isoforms with different 3'UTR (untranslated region) lengths due to the position of poly(A) tail. This difference may alter the stability and location of mRNA isoforms in cell. APA has been observed in various cellular states. For example, preferential use of distal poly(A) sites and 3'UTR lengthening occurs during development. However, in proliferation, use of poly(A) sites result with 3'UTR shortening for a large group of genes. In cancer cells, APA seems to be deregulated based on few observations. Widespread 3'UTR shortening was reported in vaious cancers. For example, our group revealed 3'UTR shortening of *CDC6*induced by estrogen in breast cancer cells. In this study, we aim to investigate the 3'UTR alterations in estrogen receptor (ER)-positive breast cancer patients and cell models. For this purpose, publicly available microarray datasets were analyzed by a probe-based microarray analysis tool (APADetect). Based on the means of proximal to distal poly(A) site ratios of individual transcripts, the SLR (short-long ratio) was calculated as an indication of short vs. long 3'UTR abundance. Significance Analysis of Microarrays (SAM) determined significant genes that had 3'UTR shortened or lengthened genes. These genes were further analyzed *in silico* to reveal 3'UTR alterations in ER(+) breast cancer patients compared to normal breast samples. Interestingly, we did not detect APA based changes in patients in connection with relapse information. Further studies of APA in an ERα transfected MDA-MB-231 cell line model revealed a complicated picture of APA possibly as a consequence of direct and indirect effects of ERα. Our results suggest involvement of APA mechanisms in ERα action mechanisms. Possible link between ERα regulated transcription and APA remains to be elucidated.

Keywords: Alternative Polyadenylation, 3’UTR, Estrogen Receptor, Breast Cancer

ÖZ

MEME KANSERİNDE ÖSTROJEN RESEPTÖR-ALFA'YA BAĞLI 3'UTR DEĞİŞİMLERİ

Öyken, Merve
YüksekLisans, BiyolojiBölümü
TezYöneticisi: Doç. Dr. A. ElifErson-Bensan

Şubat 2014, 110sayfa

Alternatif poliadenilasyon (APA) farklı 3'UTR (transle olmayan bölge) uzunluklarına sahip farklı transkript izoformlar meydana getirir. Bu farklılıklar mRNA'nın hücre içindeki stabilitesini ve lokasyonunu değiştirebilir. 3'UTR değişimleri birçok hücresel durumda gözlemlenir. Örneğin, gelişim sırasında 3'UTR uzaması görülürken, proliferasyon sırasında kısalma gözlemlenmektedir. Kanser hücrelerinde de düzensiz bir APA söz konusudur. Birçok çalışmada yaygın olarak 3'UTR kısalması rapor edilmiştir. Meme kanseri hücrelerinde östrojen etkisiyle *CDC6* geninde 3'UTR kısalması olduğunu daha önceki çalışmalarımızda göstermiştik. Bu çalışmada ise, östrojen reseptör (ER)-pozitif meme kanseri hastalarında ve hücre modellerindeki 3'UTR değişimlerinin incelenmesi amaçlanmıştır. Bu amaçla, halka açık mikroçip datasetleriprob-tabanlı mikroçip analiz aracı (APADetect) ile analiz edildi. Her bir transkriptin proksimal/distalpoly(A) bölgesi oranlarının ortalamalarına bağlı olarak, kısa vs. uzun 3'UTR varlığının göstergesi olan kısa-uzun oranı (SLR) hesaplandı.

Mikroçiplerin Anlamlılık Analizi (SAM) ile anlamlı olarak 3'UTR'si uzayan ya da kısalan genler belirlendi. Daha sonra, bu genler in siliko analize tabi tutuldu. Bu analizlerin sonucunda ER(+) meme kanseri hastalarında, normal meme örneklerine kıyasla, 3'UTR değişimleri saptandı. Fakat hastalığın nüks ettiği hastalarda, nüks etmeyenlere göre, herhangi bir değişim bulunmadı. ERα'nın direkt ve indirekt etkilerinin sonucu olarak, ERαtransfekte edilmiş MDA-MB-231 hücre hattı modelinde ise daha karmaşık bir APA tablosu meydana geldi. Sonuçlarımız ERα'nın etki mekanizmalarında APA mekanizmalarının da dahil olduğunu önermektedir. ERα'nın düzenlediği transkripsiyon ve APA arasındaki muhtemel bağlantı tam olarak açıklığa kavuşturulamamıştır.

Anahtar Kelimeler: Alternatif Poliadenilasyon, 3'UTR, Östrojen Reseptör, Meme Kanseri

To my family

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest appreciation to my supervisor Assoc. Prof. Dr. A. ElifErson-Bensan for her endless support, motivation, and immense knowledge. Without her inspiring guidance and encouragement, I could not have accomplished this thesis successfully.

I would like to thank all my thesis committeemembers; Assoc. Prof. Dr. MesutMuyan, Assoc. Prof. Dr. Sreeparna Banerjee, Assist. Prof. Dr. Mehmet Somel, and Assoc. Prof. Dr. ÖzlenKonu.

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. MesutMuyan for his invaluable suggestions and inspiration.

I would like to thank to all previous and current members of Erson Lab; BegümAkman, TanerTuncer, NihanÖzdemirler, Tuna Çınkıllı, ÜlküUzun, Shiva Akhavan, AyşegülSapmaz, and CansaranSaygılı for providing a fun filled environment, for their precious friendship, and for their advices.

I would like to thank to Muyan Lab members; GizemGüpür, PelinYaşar, SırmaDamla User, GamzeAyaz, and Hasan Hüseyin Kazan for their support and friendship.

I would like to thank to Banerjee Lab for sharing their resources.

I am deeply thankful to my friends for their encouragement, moral support, and suggestions.Many thanks to M. Şeyma Ceyhan for her friendship, patience, and motivation throughout my undergraduate and graduate studies.

I would like to express my endless gratitude to my family; my mother İncilayÖyken, my father EthemÖyken, and my brother UğurÖyken. I wish to offer my deepest thanks to my cousin Senem Ayşe Haser who was always beside me to motivate me during the happy and tough moments of my life.

I would like to thank TUBITAK for support (112S478).

TABLE OF CONTENTS

CHAPTERS

APPENDICES

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

1.1 Alternative polyadenylation

1.1.1 3’ end processing of pre-mRNA

Following transcription,precursor mRNAs (pre-mRNAs)are processed into mature mRNAs in the nucleus. During maturation, poly(A) tail is added to the 3’ end of almost all mRNAs by template-independent poly(A) polymerases (PAPs)(Eckmann et al., 2011). The length of poly(A) tails differs among species; however, there is an average length of 250-300 Adenines in humans and 70-80 in yeast (Elkon et al., 2013). Poly(A) tail provides transport of the mRNA from the nucleus to cytoplasm, mRNA stability, and efficient translation(Colgan and Manley, 1997). The variation of the length of the poly(A) tail changes some aspects of mRNA characteristics. mRNAs having relatively long tails have higher half-life in cytoplasm than those with shorter tails.An insufficient tail length leads to untranslation or enzymatic degradation of the mRNA (D'Ambrogio et al., 2013).Hence, the length of 3’UTRs of mRNAs is another critical factor determining protein levels. In this regard, polyadenylation mechanisms become important for 3’UTR length regulation.

1.1.2 Mechanism of polyadenylation

Poly(A) tail is added to pre-mRNA in a two-step reaction: First, pre-mRNA is cleaved endonucleolytically, and then synthesis of the tail occurs(Di Giammartino et al., 2011). This reaction requires many multi subunit protein complexes (Figure 1.1). The main proteins for the formation of 3’ end are CPSF (cleavage and polyadenylation specificityfactor), CstF (cleavage stimulation factor), CFIm and CFIIm(cleavage factors Im and IIm), poly(A) polymerase (PAP), and RNA polymerase II (RNAP II)(Shi et al., 2009).

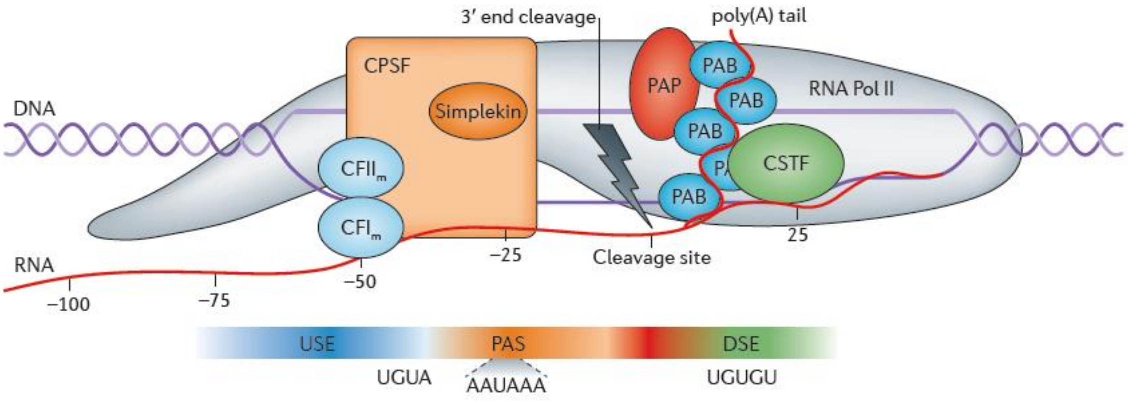


Figure 1.1 Proteins involved in cleavage and polyadenylation processes. Several proteins are involved in cleavage and polyadenylation processes. Cleavage is performed by CPSF, CSTF, CFIm, and CFIIm. PAP and poly(A)-binding proteins (PABs) are required for cleavage and polyadenylation. Poly(A) signals (PAS) represent the cleavage site. Upstream and downstream sequence elements (USE and DSE) enhance cleavage efficiency. Figure is taken from (Elkon et al., 2013).

Poly(A) signals (PAS) are6-nucleotide motifs that specifies the cleavage site of mRNA. They are located in the 15-30 nucleotides upstream of the cleavage site. AAUAAA is the common sequence of PAS; however, ten additional PAS sequences were recognized(Proudfoot, 2011). UGUA motif which is located 40-100 nucleotides upstream of the cleavage site is recognized by CFIm and CFIIm(Brown and Gilmartin, 2003). Upstream sequence elements (USE) and downstream sequence elements (DSE) are U- or GU-rich sequences toimprove cleavage efficiency.Poly(A)-binding proteins (PABs) are needed for effective polyadenylation and the length of the poly(A) tail is controlled by PABs.

1.1.2.1 Types of polyadenylation

50% of human genes contain multiple poly(A) sites (Tian et al., 2005). As a result, alternative polyadenylation(APA) is likely to generate different isoforms with different 3'UTR lengths(Andreassi and Riccio, 2009).

There are four types of APA;1) Tandem 3'UTR APA, 2) Alternative Terminal Exon APA, 3) Intronic APA, and 4) Internal Exon APA (Figure 1.2). In tandem 3'UTR type of APA, there are two polyadenylation sites in the 3'UTR. As a result of APA, the two transcripts share the same coding sequence, but have different 3'UTR lengths. This type is one of the most common types observed. Alternative terminal exon APA, another common type of APA, creates two transcripts that differ in their coding sequences and 3'UTRs. When proximal poly(A) site is in an intron, an internal exon becomes the terminal exon, so part of an intron becomes 3'UTR. This type of APA is called intronic APA.In internal exon APA,proximal poly(A) site is in an exon. Proximalpolyadenylation within the coding region is observed(Elkon et al., 2013).

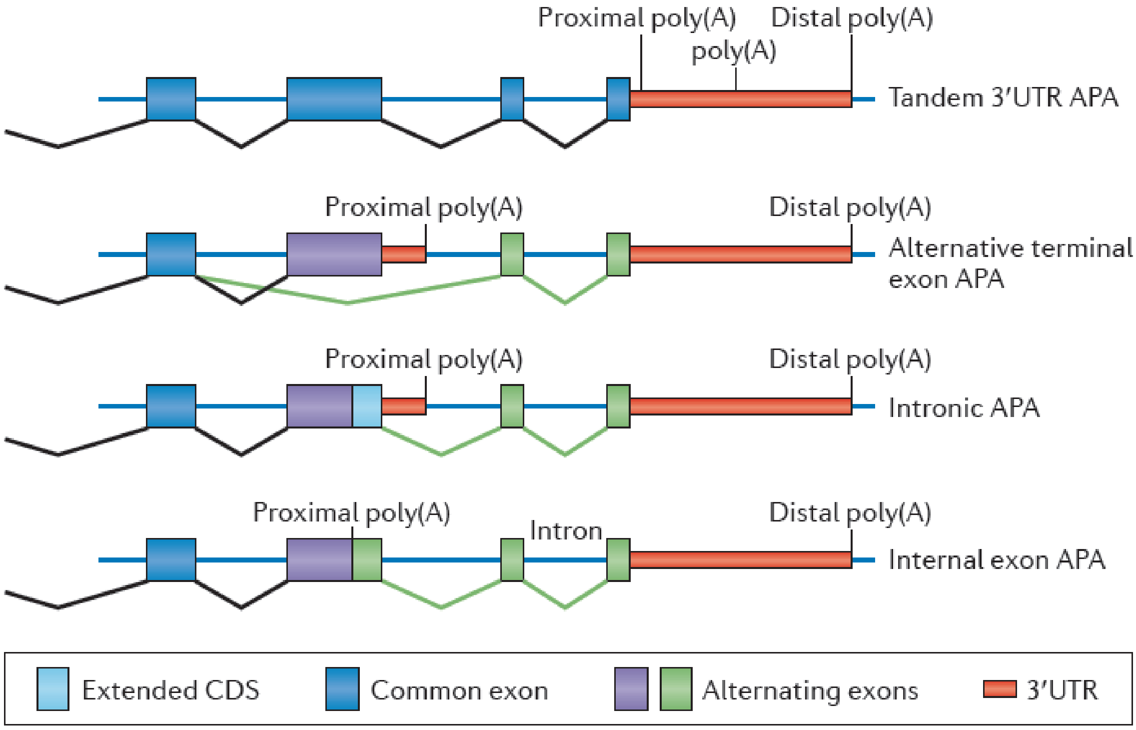


Figure 1.2 Types of alternative polyadenylation. In tandem 3'UTR APA type, the coding sequence does not change as a result of APA, only 3'UTR changes. In the other isoforms, the coding sequence, as well as 3'UTR, differs between the alternatively polyadenylated isoforms. Figure is taken from (Elkon et al., 2013).

1.1.2.2 APA-regulating *cis*-acting elements

The consensus sequence AAUAAA is the most frequently used *cis*-acting element (Tian et al., 2005). However, there are newly discovered several sequences near poly(A) sites (Figure 1.3). A TTTTTTTT motif and an AAWAAA motif (W denoting A or T) are positioned upstream of the poly(A) site. CCAGSCTGG sequence (S denoting C or G) is found downstream of the poly(A) site. They are located in intragenic and intergenic poly(A) sites, but they are not found with the AAUAAA motif (Ozsolak et al., 2010). Moreover, stem-loop motifs and secondary structures in 3'UTRs are likely to have roles in the regulation of mRNA stability. Such structuresmay also improve or suppress the modulation of APA by regulating 3' end processing proteins (Di Giammartino et al., 2011).

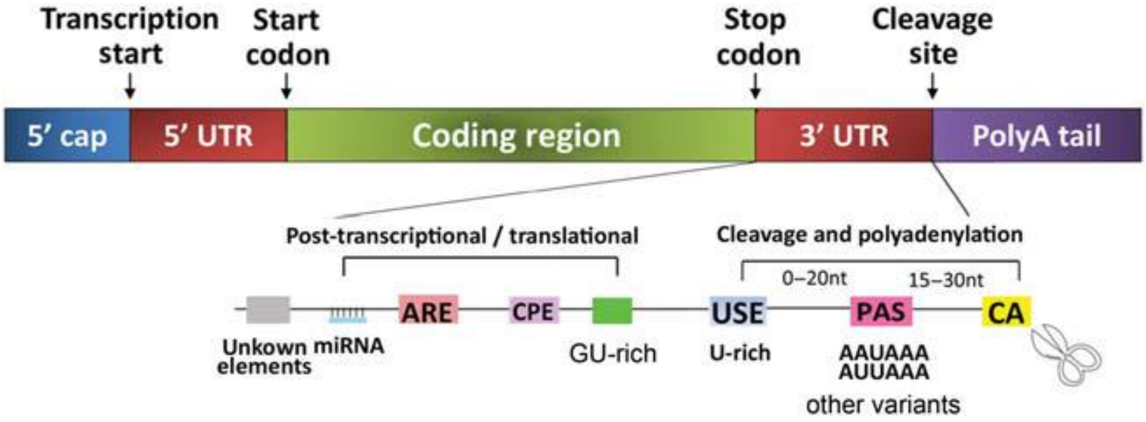


Figure 1.3 The *cis*-acting elements in 3'UTR. *Cis*-acting elements can be divided into two categories. The first category is that miRNA binding sites, AU-rich element (ARE), cytoplasmic polyadenylation element (CPE), and GU-rich elements are classified as post-transcriptional/translational elements. The second category is that cleavage and polyadenylation sites include U-rich element (USE) and poly(A) signal(PAS). Figure is retrieved from (Sun et al., 2012).

1.1.2.3 *Trans*-acting factors regulating APA

Recent studies suggested RNA-binding proteins (RBPs) to be involved in APA regulation. Polyadenylation binding protein nuclear 1 (PABPN1) is one example. It inhibitsAPA at the proximal sites in 3’UTRs (Jenal et al., 2012). Another RBP is neuronal ELAV (embryonic lethal, abnormal vision-Drosophila) protein. It suppresses cleavage at proximal poly(A) sites(Mansfield and Keene, 2012). Another example to inhibitory proteins is polypyrimidine-tract-binding protein 1 (PTB1). It competes with CSTF64 for binding to DSEs(downstream sequence elements) (Castelo-Branco et al., 2004).

Cytoplasmic polyadenylation element binding protein 1 (CPEB1) binds to *cis*-regulatory cytoplasmic polyadenylation element (CPE) in cytoplasm and recruits CPSF and cytoplasmic poly(A) polymerase GLD2 [poly(A) polymerase associated domain containing 4) to the PAS element. In addition to this regulation, when CPEB1 binds to CPE near to proximal poly(A) sites in the nucleus, 3’ end processing proteins are also recruited to this site and generateshorter 3’UTRtranscripts (Bava et al., 2013).

1.1.2.4 Chromatin and APA

Epigenetic and chromatin modifications are also thought to influence APA patterns. There is a depletion of nucleosomes at poly(A) sites; whereas, there is an enrichment at regions downstream of these sites.This is due to the AT rich sequence of poly(A) sites resulting in poor affinity to nucleosomes. Eventually, nucleosome positioning affects APA. When a poly(A) site is highly depleted of nucleosomes and its downstream is highly enriched with nucleosomes, this poly(A) site becomes stronger, meaning; more preferable than other sites (Spies et al., 2009).

DNA methylation is another example to the relation between chromatin and APA. This relation was firstly confirmed in mouse cells.*H13 and Herc3*genes have two retrogenes (*Mcts2* and *Napl1*, respectively) in their introns. On the maternal allele, both genes are methylated, so they are not active; however, on the paternal allele, they are not methylated, so they are active. The host gene uses its intronicpoly(A) site when the transcription of the retrogene occurs on the paternal allele. In contrast, the host gene uses its distal poly(A) site on the maternal copy in which the retrogene is silenced (Cowley et al., 2012; Wood et al., 2007). These observations suggest that the regulation of APA is also controlled by complex chromatin and epigenetic modifications.

1.1.2.5 Interaction between transcription and APA

The coupling between 3’ end processing and transcription has been reported earlier(Buratowski, 2005). The specific relation between APA and transcription is of interest.There are two principles about how APA and transcription is associated.

The first one is that cleavage efficiency of poly(A) tail and use of proximal sites increase with the interplay between mechanisms of transcription and 3’ end processing (Di Giammartino et al., 2011). This claim was supported with RNA polymerase II-associated factor complex (PAF1c). It has roles in both transcription elongation and 3’ end processing machinery(Nagaike et al., 2011). When CDC73 (one of the subunits of PAF1c) was knocked down, one of the targets of CDC73 had a longer 3’UTR. This indicates that PAF1c enhances proximal site cleavage(Rozenblatt-Rosen et al., 2009). In a similar manner, ELL2 (transcription elongation factor) supports loading of CSTF on the mechanism of transcription and increases the use of proximal poly(A) site of IgM(Martincic et al., 2009).

The second principle is the kinetic coupling. Because the proximal poly(A) sites are firstly transcribed, they first meet with the 3' end processing factors; therefore, the promixal sites are more favorable than distal sites (Shi, 2012). It means that longer distance between the proximal and distal poly(A) site results with more use of proximal sites than distal ones. Furthermore, when transcription elongation rate decreases, the use of proximal poly(A) sites increasesin developmental processes rather than proliferation (Pinto et al., 2011) (Figure 1.4).

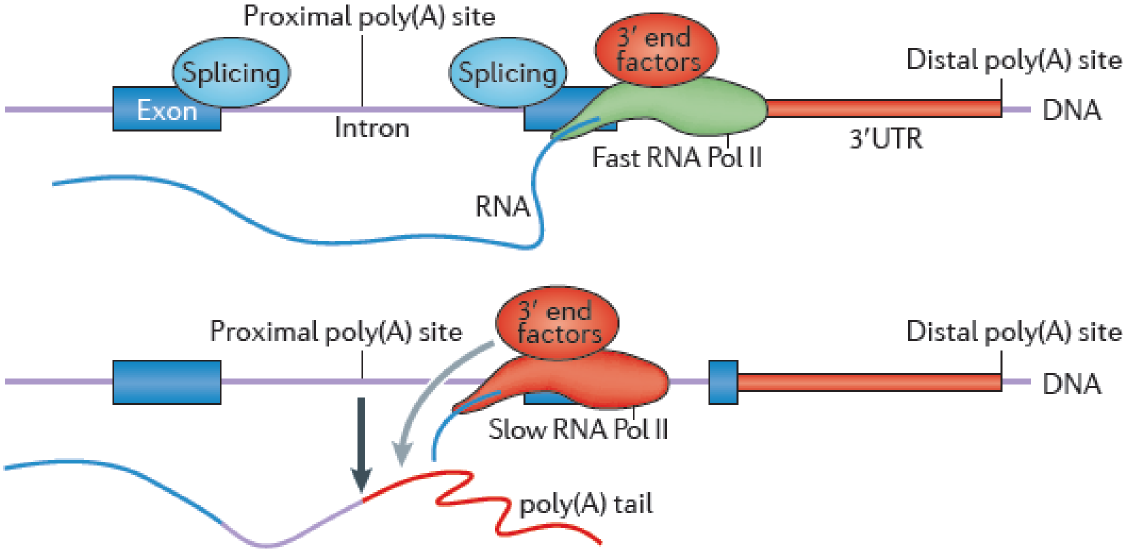


Figure 1.4Kinetic coupling of transcription and APA. The rate of the RNA Pol II may determine the choice of poly(A)site. When it is fast, distal poly(A) site is chosen, whereas when it is slow, proximal poly(A) site is preferred. Figure is adapted from (Elkon et al., 2013).

1.1.2.6 Interaction between splicing and APA

Cleavage efficiency increases when splicing and polyadenylation factors physically interact at the terminal introns of pre-mRNA(Millevoi and Vagner, 2010). Alternative terminal exon and intronic APA events are influenced by this interaction. There are splicing-related regulatory motifs in 3’UTRs, thus both splicing and APA may be regulated by the factors binding to these motifs.

One example to this interaction is NOVA2 (neuron-specific RNA binding protein). It is a splicing factor that binds to 3’UTRs.Transcripts with longer 3’UTRs are produced by the help of NOVA2. This claim was tested with *Nova2*-knockout mouse brains, resulting in shortening of 3’UTR as compared to the wild type (Licatalosi et al., 2008).

The roles of CPE (cytoplasmic polyadenylation element) and CPEB1 (cytoplasmic polyadenylation element binding protein) in polyadenylation mechanismwere mentioned earlier. CPEB1 binds to CPE and recruits CPSF (Bava et al., 2013). Both CPEB1 and CPEare thought to be involved in APA along with alternative splicing. If alternative splice sites have CPEs, splicing machinery may be interrupted by binding of the CPEB1-CPSF to that CPEs (Bava et al., 2013).

Another example is a spliceosome component U1 (small nuclear ribonucleoprotein). When U1 was knocked-down, intronicpoly(A) sites were preferred (Figure 1.5). Therefore, U1 may have a role in protection of pre-mRNAs from early cleavage. In addition, when transcription rate increases, U1 shortage is observed, leading to shorter transcripts(Kaida et al., 2010).

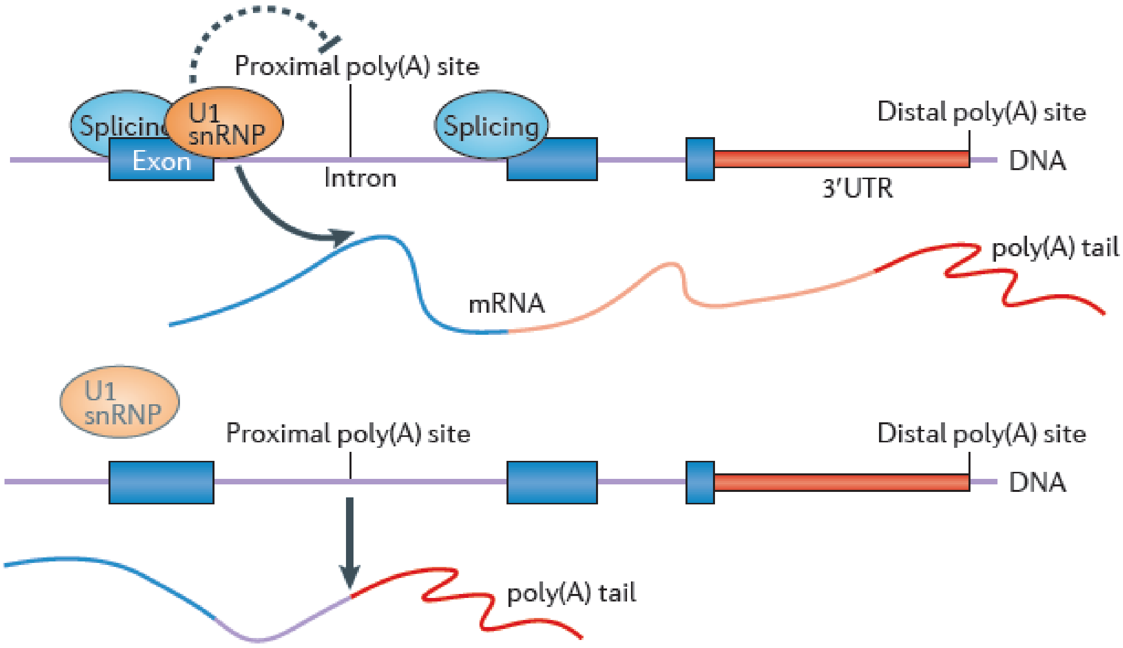


Figure 1.5Interaction between splicing and APA through U1 small nuclear ribonucleoprotein. When U1 involves in spliceosomecomplex, it prevents the use of intronic proximal poly(A) site. However, when U1 is not present, both splicing and APA machineries are affected. Figure is taken from (Elkon et al., 2013).

1.1.3 APA sites and APA events

Determining the APA sites in the genome may provide better understanding about the genes and their expressions. There are several ways to explore these sites in the genome with the help of bioinformatics.

Expressed sequence tag (EST) databases provide data for searching poly(A) sites. Many unreported polyadenylation sites were discovered by analyzing the EST sequences(Gautheret et al., 1998). Nevertheless, these databases provide limited information because the composition of EST databases is in the frame of biological circumstances. Furthermore, given the number of poly(A) sites predicted, experimental confirmation of these sites becomes important.

Microarray technology is another method to gain experimental evidence about functional APA sites. Tandem 3'UTR APA is the mostly analyzed event. Probe positions may be useful to determine expression level differences of 3'UTR of isoforms.However, there are some limitations concerning the use of this technology. The array probe positions restrict the analysis. In addition, the analysis is made with the known poly(A) sites; therefore, it may miss unknown sites. Furthermore, transcripts having more than two polyadenylation sitescould be difficult to analyze.

Sequence-based approaches are also of interest to study poly(A) sites with the help of advanced RNA sequencing (RNA-seq) technology.Detection of short isoforms and in some cases mapping novelpoly(A) sites can be achieved by analyzing the RNA-seq data. Unlike other APA analysis methods, unannotated transcripts can also be discoveredusing this technique(Pickrell et al., 2010). However, the standard RNA-seq data may notextend beyond the polyadenylation sites and specific optimizations may be needed.

Then, deep-sequencing techniques become more suitable for APA analysis.Genes having three or more poly(A) sites, which may not be probably discovered using microarray techniques, can be detected by deep-sequencing techniques. The difficulties in statistical analysis of multiple poly(A) sites can alsobe overcome with new analysis methods(Fu et al., 2011).

In short, different approaches are being developed to specifically determine the 3' ends of mRNAs. These techniques will provide more accurate APA analysis and will better reveal the extent of APA in normal and in cancer cells(Sun et al., 2012).

1.1.1 Biological processes involving APA

Many studies have reported APAto be involved in various biological processes. Differences in 3'UTR lengths of transcripts have been observed in different cell contexts and in different pathways.

One of the alternatively polyadenylated genes discovered over 30 years ago is the immunoglobulin M (IgM) heavy-chain gene. When mouse B cells were activated, 3'UTR shorteningof IgM transcript was observed (Alt et al., 1980). The use of proximal poly(A) site results in the secreted form of the protein, rather than membrane-bound form (Peterson and Perry, 1989).

Later, high-throughput analysis revealed 3'UTR shortening events in T cells. Expression of short isoforms of 86% of genes was increased in activated and proliferating T cells(Sandberg et al., 2008). These examples suggest a correlation between cell proliferation and tandem shortening of 3'UTR.

Brain-derived neurotrophic factor (BDNF) is another case of APA regulation taking place in neurons. APA does not change the coding sequence of the two isoforms of *BDNF*, but changes the 3'UTR length. These two isoforms, when translated,were found in different subcellular locations. The short isoform localizes at somata, whereas the long one is in dendrities(An et al., 2008). Based on this study, APA seems to alter the subcellular localization of proteins.

Furthermore, transcripts with longer 3'UTRs are more preferred in spinal cord and brain (Zhang et al., 2005b). However, when there are stimulations in the environment, truncated proteins or transcripts with shorter 3'UTR are produced. As a result, the equilibrium in the neuronal tissues is disrupted. This may be due to the fact that these tissues have to respond rapidly to environmental stimuli. The transcription has to be very fast to produce the transcripts of the respondent proteins. Therefore, the proximal poly(A) sites may be chosen (Ji et al., 2011). This case is another example that the transcription and APA is coupled (Figure 1.6).

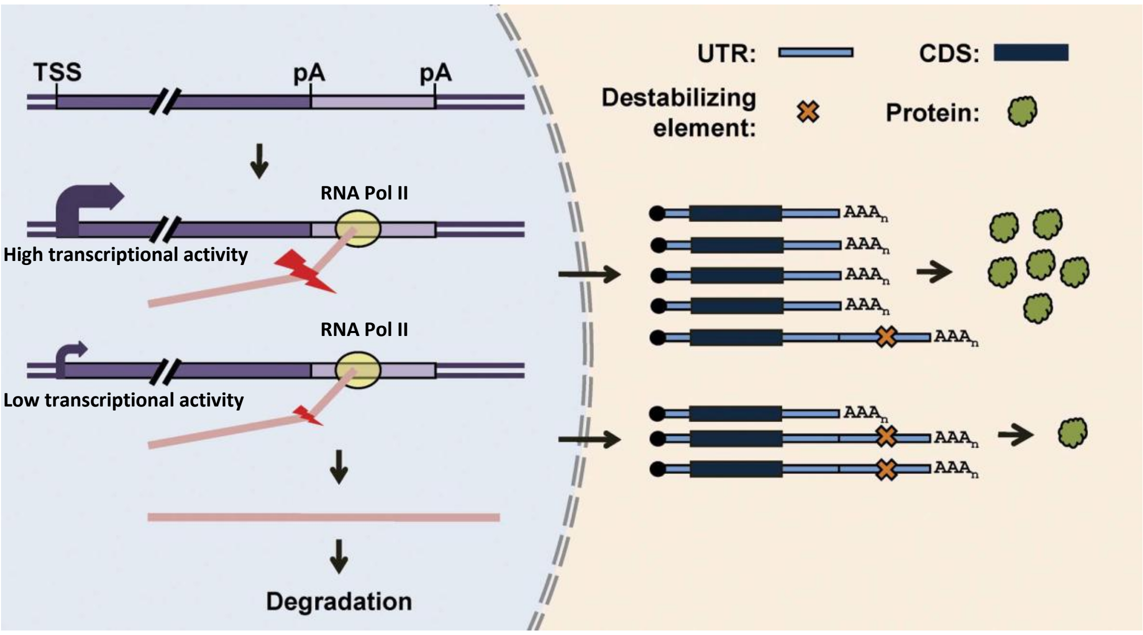


Figure 1.6The relation of APA and transcriptional activity. Shorter transcripts are produced when transcriptional activity is high, whereasduring low transcriptional activity, unprocessed mRNAs which are degraded or longer transcripts are produced. Due to the low number of destabilizing elements in the 3'UTRs of shorter transcripts, they are more stable than longer ones. TSS indicates transcription start site and pA refers to poly(A) site. Figure is taken from (Ji et al., 2011).

On the other hand, in developmental processes, 3'UTR lengthening seems to be more prominent than shortening event. When mouse C2C12 myoblast cells are differentiated into myotubes, proximal poly(A) sites are not processed. This can be due to the fact that during differentiation, 3' end processing genes are downregulated(Ji et al., 2009).

Alternative polyadenylation is also involved in tumorigenesis. Because microRNAs (miRNA) are highly deregulated in cancer, it is expected that the 3'UTRs of transcripts change due to the targeting of these regions by miRNAs. Global 3'UTR shortening is observed in primary breast cancer samples *in vitro*(Fu et al., 2011). For example, when Cyclin D2 is in the shorter isoform, there are more MCF7 cells in the S phase than cells having longer isoform. In addition, IMP-1 (Insulin-like growth factor 2,mRNA binding protein 1) is an RNA-binding protein which is upregulated in many cancers. Its shorter isoform enhances oncogenic processes(Mayr and Bartel, 2009).

Possible mechanisms have been proposed for the use of short 3'UTRs in cancer. One of them is genomic abnormalities. Cyclin D1 has shorter isoforms in mantle cell lymphoma tumors (Rosenwald et al., 2003). The differences between these isoforms are only in 3'UTRs. Genomic deletionsor point mutations resulting in *de novo*poly(A) sites could be the main reasons for shortened isoforms(Wiestner et al., 2007). Furthermore, translocation may occur between the 3'UTRs of the genes resulting in APA. *HMGA2*(high mobility group AT-hook 2) gene is an example to this situation. It changes its 3'UTR with other transcripts by translocation. As a result of this new 3'UTR structure, miRNA regulation is altered(Lee and Dutta, 2007).

1.2 Estrogen receptor signaling

Estrogen receptor (ER) is a nuclear hormone receptor having roles in various parts of the human body. There are two forms of ER: ERα and ERβwhich are transcribed from two different genes on different chromosomes (Bjornstrom and Sjoberg, 2005). They have domains which are evolutionarily conserved. DNA-binding domain (DBD) is the main domain found in both ERα and ERβinteracting with DNA. The domain which ligands are binding is called the ligand-binding domain (LBD). In addition, there are two activation function (AF) domains allowing thebinding of several coregulatory proteins when the ER is bound on DNA (Nilsson et al., 2001). Figure 1.7 is a representative picture of ER structures.

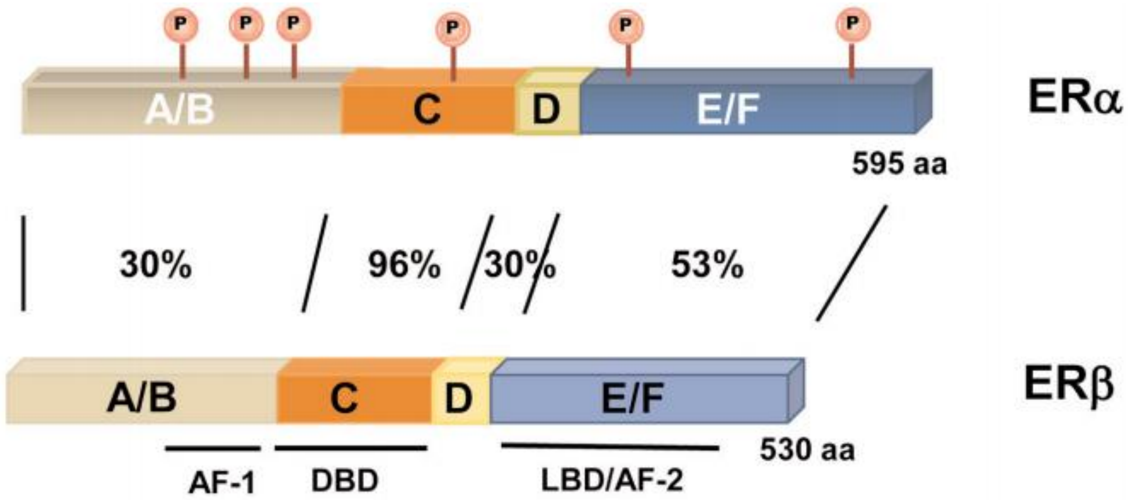


Figure 1.7Structures of ERα and ERβ. Four domains are involved in both receptors. AF-1 and AF-2 domains are the two activation function domains. DBD and LBD are the DNA-binding and ligand-binding domains, respectively. The percentages represent the homology between the two receptors. Several kinases can phosphorylate ERα leading to diverse cellular responses. Figure was retrieved from(Roman-Blas et al., 2009).

Estrogen receptors are expressed in bone, brain, and the mammary gland (Matthews and Gustafsson, 2003). However, ERα and ERβ may not be present in the same tissue. Although they may form heterodimers, ERβis thought to repress the activity of ERα if they are expressed in the same cell(Cheskis et al., 2007).

1.2.1 Types of estrogen receptor signaling

Estrogen receptors are classified as ligand-dependent transcription factors. The main action of these receptors is the ability to bind to 17β-estradiol (E2).Upon binding of the ligand, the LBD changes conformation of the receptors. This changing leads to the activation of coactivator proteins (Rosenfeld and Glass, 2001). As a result, receptors dimerize and bind to specific sequences called estrogen response elements (EREs) found on the promoters of the estrogen-regulated genes (Nilsson et al., 2001).

When the promoters of the genes do not have EREs, they can be regulated by ER through other mechanisms. One of them is that ER can bind to other transcription factors, such as SP1 (specificity protein 1) and DBC1 (deleted in breast cancer 1),in the nucleus, and then they co-bind to specific transcription factor sites(Gottlicher et al., 1998; Porter et al., 1997; Yu et al., 2011). Furthermore, ERs can be activated through growth factor receptors. Activation of these receptors results in the phosphorylation of ER. The phosphorylated receptor can bind to DNA. This mechanism is an example of the ligand-independent activation of ER(Bjornstrom and Sjoberg, 2005). In addition to the genomic association of ERs and estrogen, estradiolcan also affect the ERs on membrane. Protein-kinase cascades are responsible for this action. In this mode, ER does not bind to DNA, but activates other transcription factors to regulate gene expression (Losel and Wehling, 2003). These mechanisms were summarized in Figure 1.8.

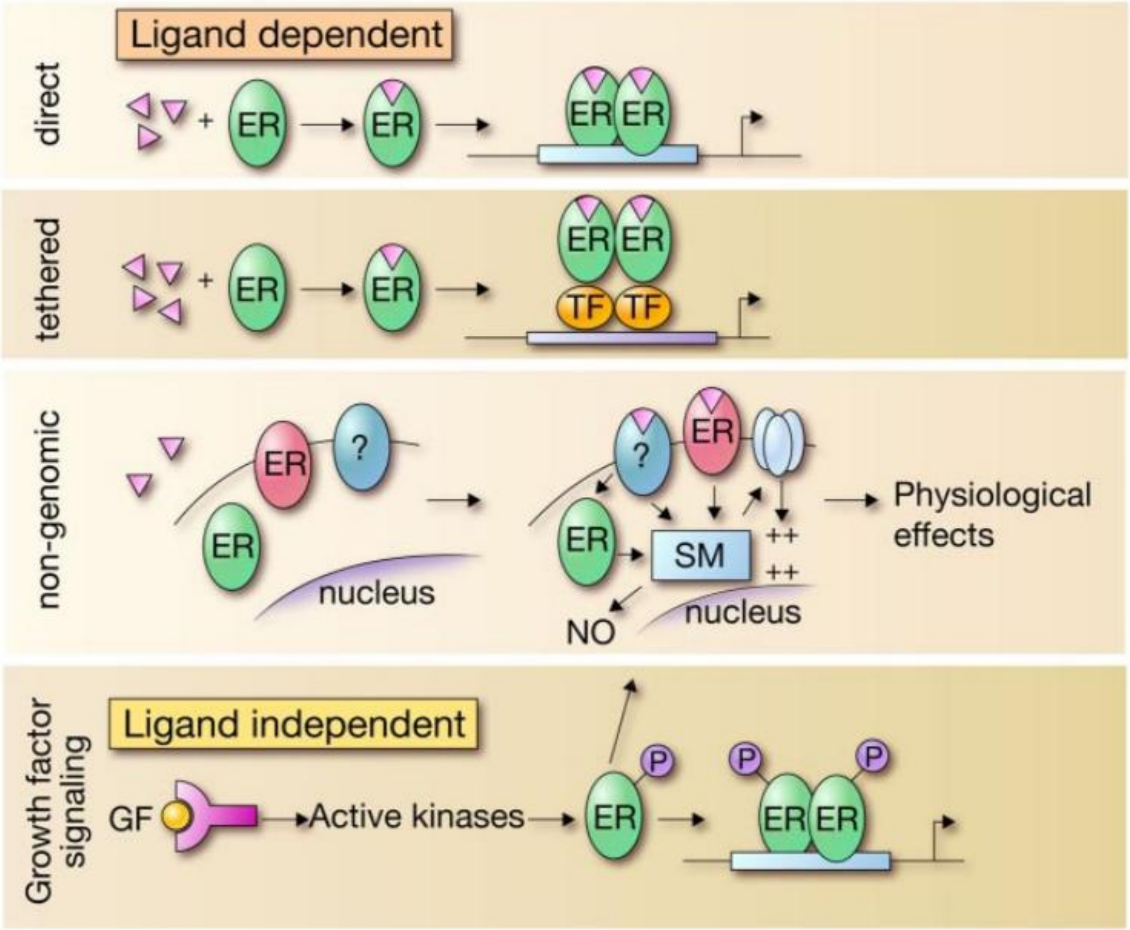


Figure 1.8Signaling mechanisms of estrogen receptor. Classical type of ER signaling is the direct signaling. Dimerized ER due to E2 binds to DNA. ERE-independent mechanism is the tethered signaling. ER binds to DNA via other transcription factors. Membranous actions of ER are the non-genomic signaling through second messengers (SM). ER does not bind to DNA, but it indirectly regulates the physiological effects like nitric oxide (NO) level or ion channel mechanisms in the cell. Ligand-independent type of action occurs via growth factor signaling. ER can bind to DNA and regulate gene expression when phosphorylated with active kinases. Figure was taken from(Heldring et al., 2007).

1.2.2 Estrogen receptor signaling and breast cancer

The relationship between estrogen and carcinogenesis has been investigated in many studies. It appears that ER signaling is associated with many cancer types, such as colon, lung, prostate, ovary, and breast (Folkerd and Dowsett, 2010).

Binding of estrogen to its receptor activates many cellular processes like cell proliferation and progression of cell cycle. For example, the expressions of c-Myc and cyclin D1 are enhanced by the ERα signaling, resulting in increased G1-S transition in cells (Musgrove et al., 2011);(Caldon et al., 2010). ERβ acts in an opposite manner. It suppresses the cell proliferation as a response to E2(Strom et al., 2004).

Among various targets, estrogen receptor signaling also regulates Bcl-2 and Bcl-X_L expression in breast cancer. The upregulation of these anti-apoptotic proteins results in the inhibition of apoptosis (Gompel et al., 2000). Moreover, the indirect activation of MAPK (mitogen activated protein kinase) and PI3K/Akt(Phosphatidylinositol 3-kinase/Protein kinase B) pathways by E2-ERα interaction promotes cell survival (Musgrove and Sutherland, 2009).

Our group also found the first link between E2 and APA. 3’UTR shortened isoform of *CDC6*(cell division cycle 6) was found to be expressed 38% more than the lengthened isoform in response to E2 treatment to MCF7 and T47D breast cancer cell lines. The protein level of CDC6 was also increased upon E2 treatment (Akman et al., 2012).

1.3 Aim of the study

3’UTR length is an important factor for post-transcriptional mechanisms of gene expression. Because shorter 3’UTRs are more generally formed in proliferative states and E2 induces proliferation, we hypothesized that E2 and ER may have roles in APA regulation. Therefore, we used ER(+) patient and model system datasets, present in the public domain, to search for alterations in 3’UTR lengths.

CHAPTER 2

MATERIALS AND METHODS

2.1 Microarray Datasets

The Gene Expression Omnibus (GEO) is a next-generation sequence and high-throughput microarray repository providing publicly available highthrougputdatasets (Barrett et al., 2013). Twelve experiment sets;GSE2034 (Wang et al., 2005), GSE7390 (Desmedt et al., 2007), GSE9761 (Nott et al., 2009), GSE20437(Graham et al., 2010), GSE9574 (Tripathi et al., 2008), GSE6883 (Liu et al., 2007), GSE26910 (Planche et al., 2011), GSE21422 (Kretschmer et al., 2011), GSE3744 (Alimonti et al., 2010), and GSE2361 (Ge et al., 2005)were selected from GEO.

GSE2034 dataset includes286 lymph-node negative breast cancer patients, 209 of them were ER(+). Tumor samples were collected from Erasmus Medical Center (Rotterdam, Netherlands). The patients were followed up between 1980 and 1995. No systemic adjuvant or neoadjuvant therapy was applied to the patients during this term. Modified radical mastectomy or breast-conserving surgery was performed (Wang et al., 2005). Patient characteristics are summarized in Table A.1.

GSE7390 dataset contains 198 lymph-node negative breast cancer patients, 135 of them were ER(+). Patients in this series were followed up between 1980 and 1998. Data were collected from six centers: Institut Gustave Roussy, Villejuif, France; Karolinska Institute, Stockholm and Uppsala University Hospital, Uppsala, Sweden;

Centre Rene´ Huguenin, Saint-Cloud, France; Guy’s Hospital, London, United Kingdom; and John Radcliffe Hospital, Oxford, United Kingdom. All the patients were systemically untreated (Desmedt et al., 2007).Patient characteristics are summarized in Table A.2.

GSE20437, GSE9574, GSE6883, GSE26910, GSE21422, GSE3744, and GSE2361 datasets were used for their normal breast tissue samples. The samples are summarized in Table A.3.

GSE9761 dataset is composed ofMDA-MB-231 breast cancer cell line. The cells were infected with recombinant adenoviruses for the expression of ER. There are 6 groups of cells:ERα expressing cells, ERE-binding defective ERα expressing cells, ERβ expressing cells, ERE-binding defective ERβ expressing cells, and 2 control groups. Then, 10^{-9} M E2 treatment to all groups was performed for 6 hours. Each group has 6 replicates (Nott et al., 2009). Samples are summarized in Table A.4.

2.2 APADetectand SAM

Earlier, a probe-based expression analysis tool (APADetect) was developed by Dr. Tolga Can (Department of Computer Engineering, METU, Turkey)(Akman et al., 2012). The tool detects expression levels of different 3’UTR isoforms. It analyzes Affymetrix Human Genome U133A Arrays (HGU133A, GPL96) and Affymetrix Human Genome U133 Plus 2.0 Arrays (HGU133Plus2, GPL570) using the available chip design data. 2411 and 3683 unique probe sets for a total of 2066 and 3067 genes are present on the HGU133A and HGU133Plus2 platforms. The 11 probes/transcript can be split into proximal and distal probe sets based on the positions of poly(A) sites taken from PolyA_DB(Zhang et al., 2005a)."Short" and "Long" 3'UTR isoform abundance are determined by calculating mean signal intensities of proximal and distal probe sets for individual genes. SLR (Short/Long Ratio) values of the groups were compared with each other to detect changes of APA based isoform abundance. SLR patient values were normalized to that of controls.

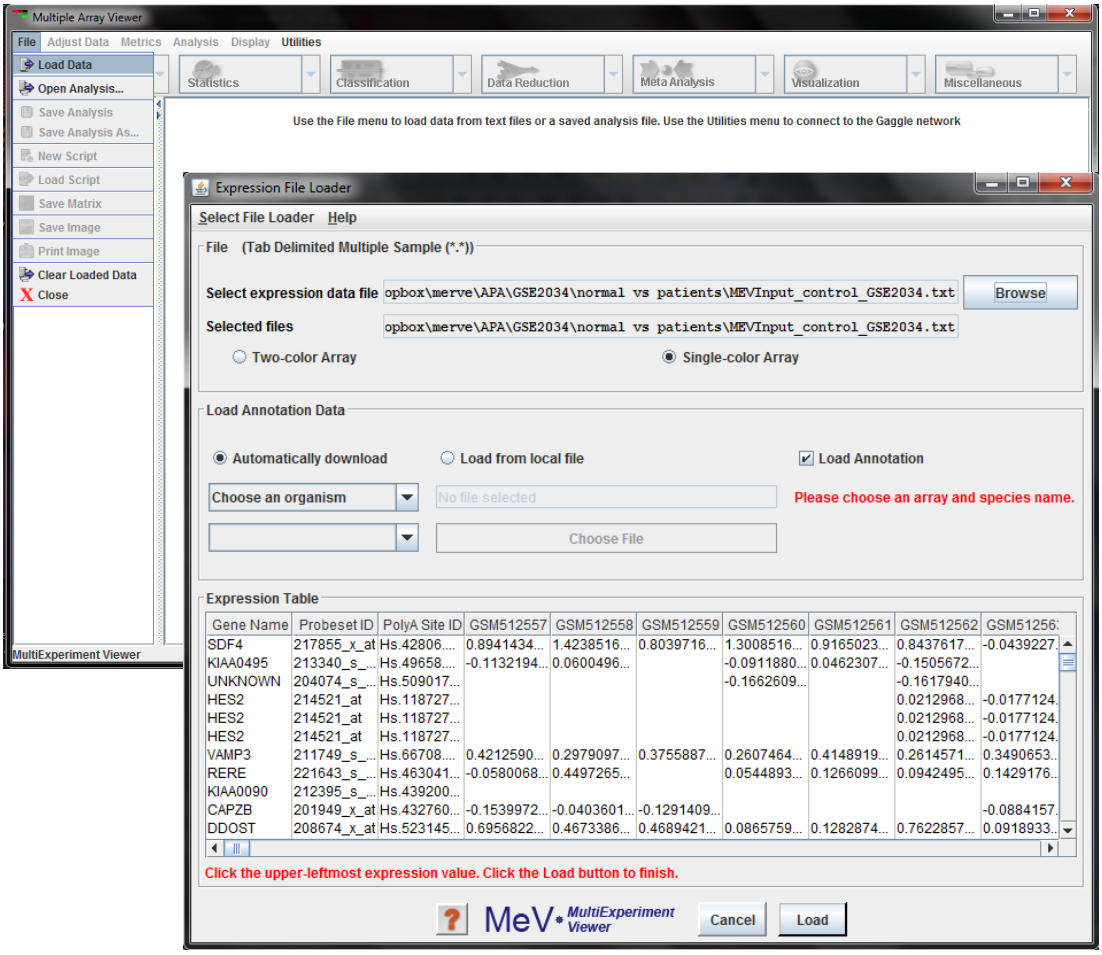
$$\mathrm{SLR} = \frac{\text{mean of proximal probes}}{\text{mean of distal probes}}$$

There are 4 stringency filters for the calculation of SLRs: “Size filter” excludes transcripts detected by only one distal or proximal probe. “Outlier probe filter” excludes individual outlier probes using Iglewicz and Hoaglin's median based outlier detection method. The median absolute deviation (MAD) is calculated as the median of absolute differences of individual probe intensities from their respective medians for each proximal and distal subset. The modified z-score of a probe p is then calculated as:

$$\mathrm{z} = \frac{0.6745\ (\text{intensityofp} - \text{median})}{\mathrm{MAD}}$$

When z-score of probes are greater than 3.5, these probes are defined as outliers and are eliminated from subsequent analyses. Samples with only one proximal/distal probe are filtered out by re-application of the “size filter” at the sample level after removal of the outlier probes. “Outlier sample filter” is applied with the same outlier detection method of outlier probe filter. SLRs of samples deviating significantly from their respective group median are eliminated. “Distal filter” removes probe sets with distal intensities significantly higher than proximal ones since proximal probes should recognize both the short and the long isoform.

The GSM numbers of samples obtained from GEO were usedinAPADetect by Dr. Tolga Can. The output of the APADetectis the input file loaded to TIGR Multiexperiment Viewer (MeV) software(http://www.tm4.org/mev.html). MeV is a microarray data analysistool for clustering, visualization, classification, statistical analysis and biological theme discovery. The data loading is illustrated in Figure 2.1.



Use the File menu to load data from text files or a saved analysis file. Use the Utilities menu to connect to the Gaggle network

File (Tab Delimited Multiple Sample (*.*))

Select expression data file `opbox\merve\APA\GSE2034\normal vs patients\MEVInput_control_GSE2034.txt`

Selected files `opbox\merve\APA\GSE2034\normal vs patients\MEVInput_control_GSE2034.txt`

○ Two-color Array ● Single-color Array

Load Annotation Data

● Automatically download ○ Load from local file ☑ Load Annotation

Choose an organism

No file selected

Please choose an array and species name.

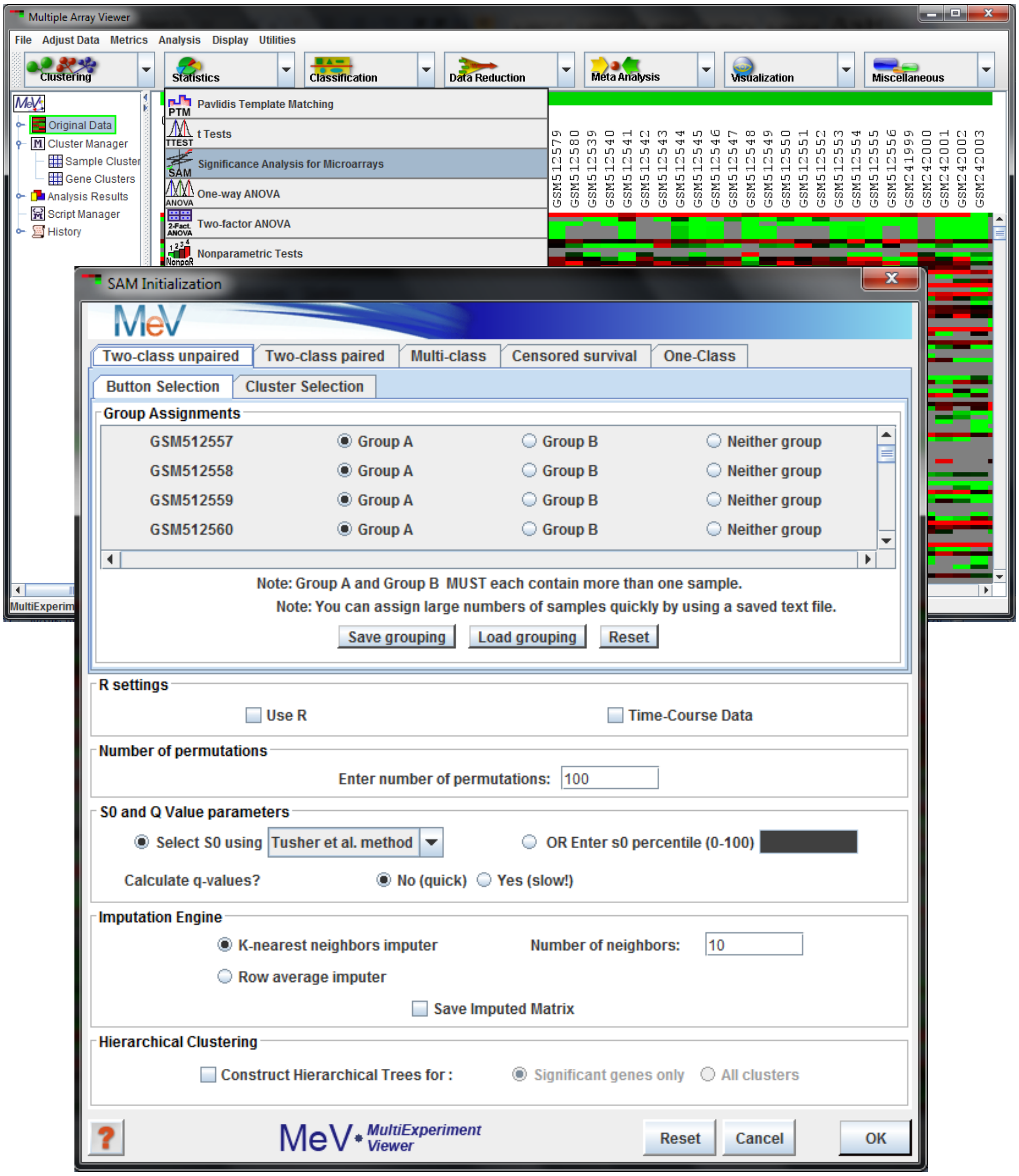
Expression Table

Gene Name	Probeset ID	PolyA Site ID	GSM512557	GSM512558	GSM512559	GSM512560	GSM512561	GSM512562	GSM51256:
SDF4	217855_x_at	Hs.42806....	0.8941434...	1.4238516...	0.8039716...	1.3008516...	0.9165023...	0.8437617...	-0.0439227.
KIAA0495	213340_s_...	Hs.49658....	-0.1132194...	0.0600496...		-0.0911880...	0.0462307...	-0.1505672...	
UNKNOWN	204074_s_...	Hs.509017...				-0.1662609...		-0.1617940...	
HES2	214521_at	Hs.118727...						0.0212968...	-0.0177124.
HES2	214521_at	Hs.118727...						0.0212968...	-0.0177124.
HES2	214521_at	Hs.118727...						0.0212968...	-0.0177124.
VAMP3	211749_s_...	Hs.66708....	0.4212590...	0.2979097...	0.3755887...	0.2607464...	0.4148919...	0.2614571...	0.3490653..
RERE	221643_s_...	Hs.463041...	-0.0580068...	0.4497265...		0.0544893...	0.1266099...	0.0942495...	0.1429176..
KIAA0090	212395_s_...	Hs.439200...							
CAPZB	201949_x_at	Hs.432760...	-0.1539972...	-0.0403601...	-0.1291409...				-0.0884157.
DDOST	208674_x_at	Hs.523145...	0.6956822...	0.4673386...	0.4689421...	0.0865759...	0.1282874...	0.7622857...	0.0918933..

Click the upper-leftmost expression value. Click the Load button to finish.

Figure 2.1 Screenshots of data loading from the MeV tool. From the file menu of MeV, “load data” was chosen.In the next window, MeV input file was loaded. The Affymetrix platform number and species name were chosen. The upper-leftmost expression value was clicked. Then, the “Load” button was chosen.

Significance Analysis of Microarrays (SAM) is a statistical method designed especially for microarrays. Statistically significant genes are detected by SAM taking into account *t* tests specific for the genes. Relative to the standard deviation of each gene, a score is defined according to the change in gene expression. Scores above a threshold are assumed as statistically significant (Tusher et al., 2001).After log normalization of sample to control SLR values, individual SLRs were tested with SAM for statistical significance. In Figure 2.2, SAM application is shown in MeV.



Group Assignments

GSM512557	● Group A	○ Group B	○ Neither group
GSM512558	● Group A	○ Group B	○ Neither group
GSM512559	● Group A	○ Group B	○ Neither group
GSM512560	● Group A	○ Group B	○ Neither group

Note: Group A and Group B MUST each contain more than one sample.

Note: You can assign large numbers of samples quickly by using a saved text file.

R settings

☐ Use R ☐ Time-Course Data

Number of permutations

Enter number of permutations: 100

S0 and Q Value parameters

● Select S0 using Tusher et al. method ○ OR Enter s0 percentile (0-100)

Calculate q-values? ● No (quick) ○ Yes (slow!)

Imputation Engine

● K-nearest neighbors imputer Number of neighbors: 10

○ Row average imputer

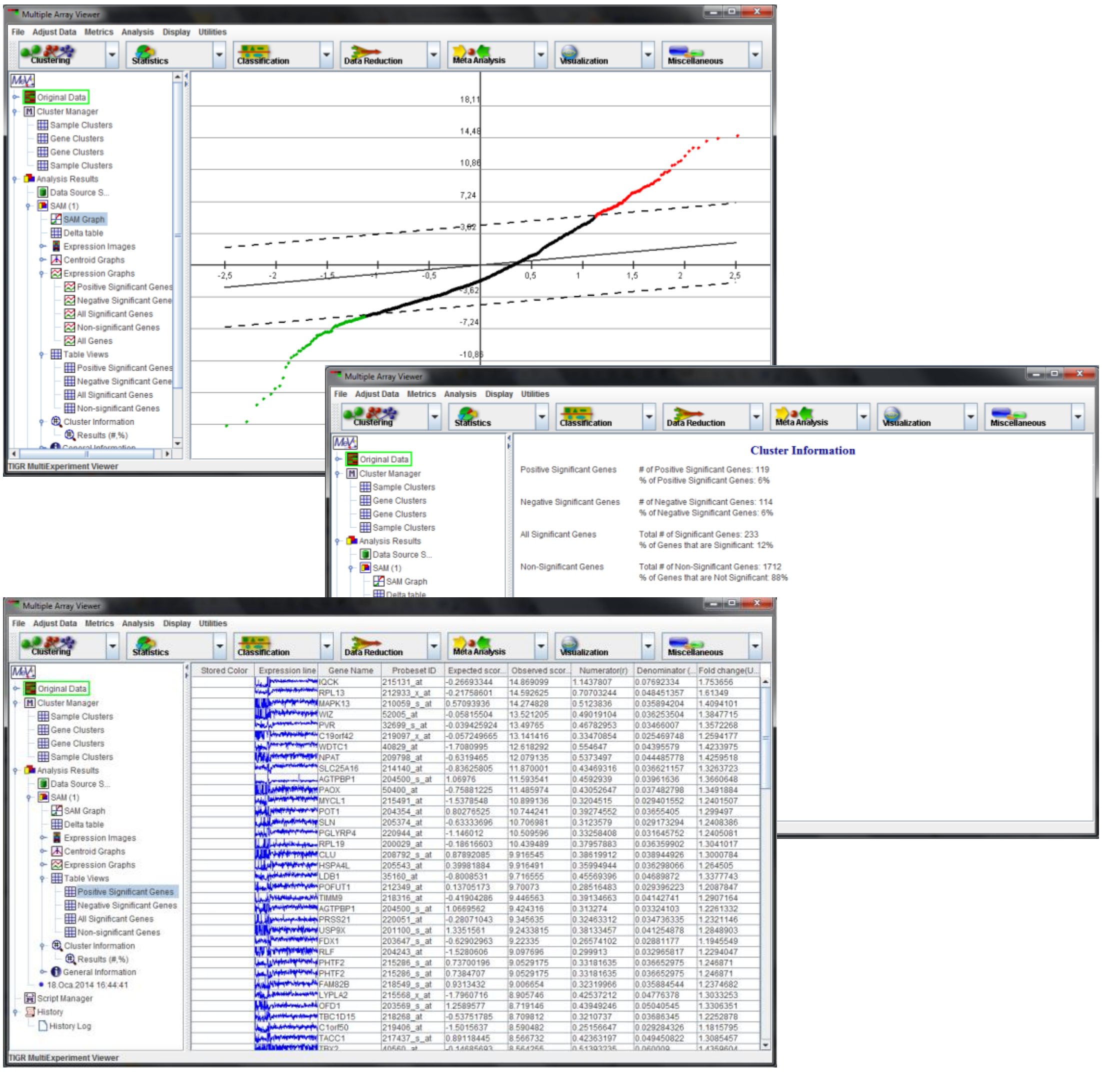
☐ Save Imputed Matrix

Hierarchical Clustering

☐ Construct Hierarchical Trees for : ● Significant genes only ○ All clusters

Figure 2.2Screenshots of SAM from the MeV tool. After data loading, SAM was chosen from “Statistics” section. Then, groups were defined as “Group A” and Group B” according to GSM numbers of the samples.

SAM was performed. The resultswere illustrated as a significance graph. In addition, the significant genes were listed according to positive and negative significance.These properties can be chosen from the toolbar at the right of the window.Figure 2.3 represents SAM application.



Cluster Information

Positive Significant Genes	# of Positive Significant Genes: 119 % of Positive Significant Genes: 6%
Negative Significant Genes	# of Negative Significant Genes: 114 % of Negative Significant Genes: 6%
All Significant Genes	Total # of Significant Genes: 233 % of Genes that are Significant: 12%
Non-Significant Genes	Total # of Non-Significant Genes: 1712 % of Genes that are Not Significant: 88%

Stored Color	Expression line	Gene Name	Probeset ID	Expected scor...	Observed scor...	Numerator(r)	Denominator (...	Fold change(U...
		IQCK	215131_at	-0.26693344	14.869099	1.1437807	0.07692334	1.753656
		RPL13	212933_x_at	-0.21758601	14.592625	0.70703244	0.048451357	1.61349
		MAPK13	210059_s_at	0.57093936	14.274828	0.5123836	0.035894204	1.4094101
		WIZ	52005_at	-0.05815504	13.521205	0.49019104	0.036253504	1.3847715
		PVR	32699_s_at	-0.039425924	13.49765	0.46782953	0.03466007	1.3572268
		C19orf42	219097_x_at	-0.057249665	13.141416	0.33470854	0.025469748	1.2594177
		WDTC1	40829_at	-1.7080995	12.618292	0.554647	0.04395579	1.4233975
		NPAT	209798_at	-0.6319465	12.079135	0.5373497	0.044485778	1.4259518
		SLC25A16	214140_at	-0.83625805	11.870001	0.43469316	0.036621157	1.3263723
		AGTPBP1	204500_s_at	1.06976	11.593541	0.4592939	0.03961636	1.3660648
		PAOX	50400_at	-0.75881225	11.485974	0.43052647	0.037482798	1.3491884
		MYCL1	215491_at	-1.5378548	10.899136	0.3204515	0.029401552	1.2401507
		POT1	204354_at	0.80276525	10.744241	0.39274552	0.03655405	1.299497
		SLN	205374_at	-0.63333696	10.706981	0.3123579	0.029173294	1.2408386
		PGLYRP4	220944_at	-1.146012	10.509596	0.33258408	0.031645752	1.2405081
		RPL19	200029_at	-0.18616603	10.439489	0.37957883	0.036359902	1.3041017
		CLU	208792_s_at	0.87892085	9.916545	0.38619912	0.038944926	1.3000784
		HSPA4L	205543_at	0.39981884	9.916491	0.35994944	0.036298066	1.264505
		LDB1	35160_at	-0.8008531	9.716555	0.45569396	0.04689872	1.3377743
		POFUT1	212349_at	0.13705173	9.70073	0.28516483	0.029396223	1.2087847
		TIMM9	218316_at	-0.41904286	9.446563	0.39134663	0.04142741	1.2907164
		AGTPBP1	204500_s_at	1.0669562	9.424316	0.313274	0.03324103	1.2261332
		PRSS21	220051_at	-0.28071043	9.345635	0.32463312	0.034736335	1.2321146
		USP9X	201100_s_at	1.3351561	9.2433815	0.38133457	0.041254878	1.2848903
		FDX1	203647_s_at	-0.62902963	9.22335	0.26574102	0.02881177	1.1945549
		RLF	204243_at	-1.5280606	9.097696	0.299913	0.032965817	1.2294047
		PHTF2	215286_s_at	0.73700196	9.0529175	0.33181635	0.036652975	1.246871
		PHTF2	215286_s_at	0.7384707	9.0529175	0.33181635	0.036652975	1.246871
		FAM82B	218549_s_at	0.9313432	9.006654	0.32319966	0.035884544	1.2374682
		LYPLA2	215568_x_at	-1.7960716	8.905746	0.42537212	0.04776378	1.3033253
		OFD1	203569_s_at	1.2589577	8.719146	0.43949246	0.05040545	1.3306351
		TBC1D15	218268_at	-0.53751785	8.709812	0.3210737	0.03686345	1.2252878
		C1orf50	219406_at	-1.5015637	8.590482	0.25156647	0.029284326	1.1815795
		TACC1	217437_s_at	0.89118445	8.566732	0.42363197	0.049450822	1.3085457

Figure 2.3Screenshots of MeV after SAM.“SAM graph” shows the distributions,“Cluster information” indicates the number, and “Table Views” section illustrates the lists of the significant genes.

2.3 Statistical Analyses

As a result of the APADetect analysis, log SLR values were determined for each gene in the array. The significant genes according to SAM were further statistically analyzed using these log SLR values. Graphpad Prism (California, USA) was used to plot the graphics. The graphic type was scatter plot in order to show every patientSLR on the graphic. Unpaired t-test was used to compare the means of patient and control groups. Welch's correction was applied with t-test to assume that these two groups do not have equal variances. F-test was also considered to evaluate that the variances of the two groups are significantly different. ROC (Receiver Operator Curve) analysis was performed for sensitivity and specificity of each significant gene (Metz, 1978; Wray et al., 2010). X-axis and y-axis define the false positive rate and the true positive rate, respectively. Interpretation of ROC curves is thorough the area under curve (AUC) value ranging from 0 to 1(Berrar and Flach, 2012).

2.4 Ontology Search

Panther Classification System was used for ontology analysis with statistically significantly 3’UTR lengthened or shortened genes as a result of APADetect and SAM (Mi et al., 2013).

CHAPTER 3

RESULTS AND DISCUSSION

3.1 APA in GSE2034

The relationship between 3’UTR shortening and proliferation was reported previously (Sandberg et al., 2008). Because ER(+) cells are highly proliferative, we investigated whether APA was deregulated in ER(+) patients. To find APA-based 3’UTR length changes, we used a probe-based microarray analysis tool developed by Dr. Tolga Can (Department of Computer Engineering, METU, Ankara). APADetectgroups microarray probes as proximal and distal based on polyadenylationposition information taken from polyA_DB(database) (Zhang et al., 2005a). Based on the sample of proximal to distal ratios of individual transcripts, the SLR (short-long ratio) is calculated as an indication of short vs. long 3’UTR abundance.

In order to statistically determine significantly altered SLRs in patients, we used SAM (Significance Analysis of Microarrays)(Tusher et al., 2001).

3.1.1 APA in ER(+) patients

Microarray datasets containing ER(+) patients were retrieved from GEO (Gene Expression Omnibus) (Barrett et al., 2013). GSE2034 has 209 ER(+) patients which were compared tonormal breast samples (n=82) using APADetectin order to observe the APA changes. APADetect results were then investigated for significance using SAM. We observed both shortening and lengthened events in ER(+) patients compared to controls (Figure 3.1).

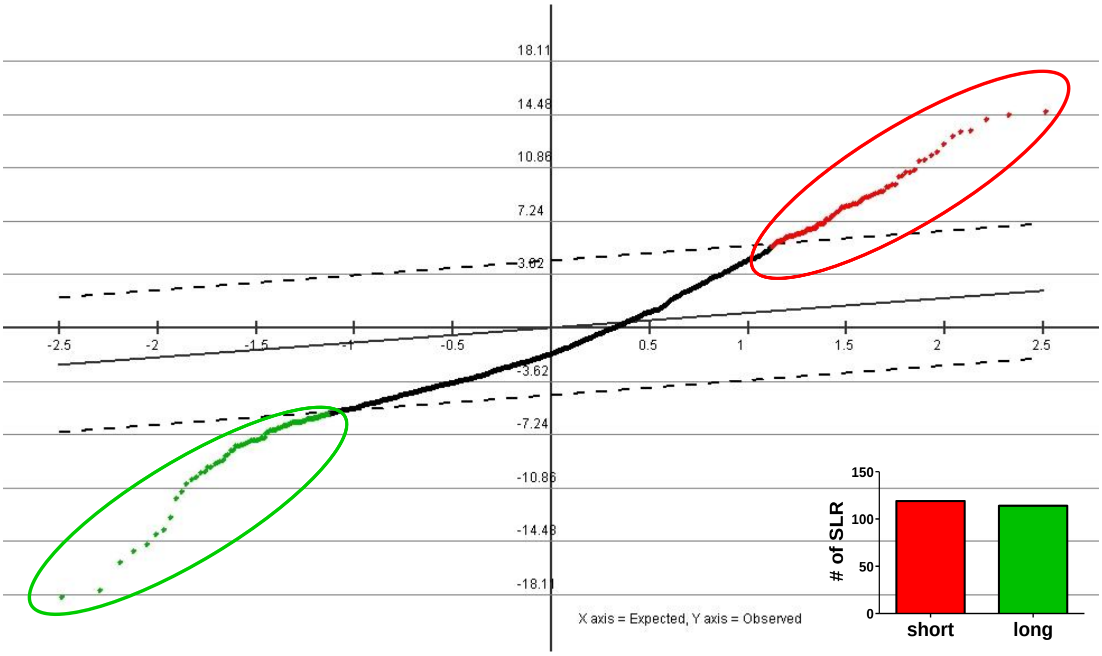


Figure 3.1SAM graph of normal samples vs. ER(+) samples comparison. The red and green dots represent the shortened and lengthened transcripts in patient samples, respectively.In total, there were119 shortening and 114 lengthening cases.

As a result of SAM, 119 shortened and 114 lengthened transcripts were determined (Table B.1 and Table B.2). All of these shortening and lengthening cases were analyzed in terms of ontologyin order to have a general idea about the molecular function of these genes (Figure 3.2).

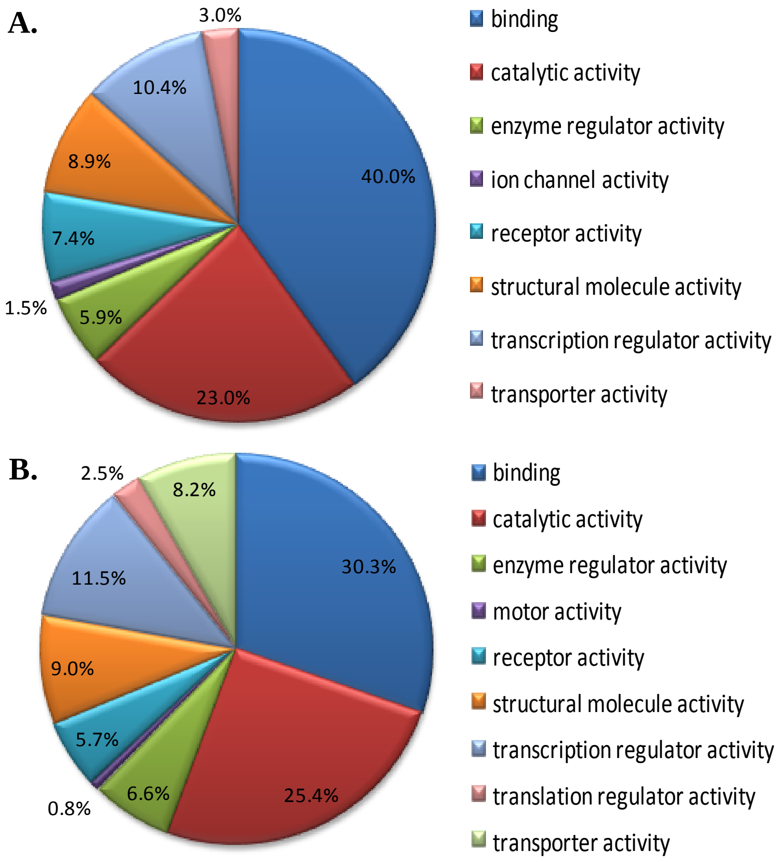


Figure 3.2 Classification of the significant genes. Shortened transcripts **(A)**and lengthened transcripts **(B)**were classified in terms of molecular function. The percentages were indicated on pie charts.The broadest categories are binding and catalytic activity for both 3’UTR shortening and lengthening cases.

Significant genes weregrouped aswithproteins having “binding” and “catalytic activity” function. Binding activity group consisted ofmostly nucleic acid and protein binding proteins. Catalytic activity was dominated by hydrolase and transferase enzymes.

For shortening events, 6 genes (*IQCK*, *MAPK13*, *WIZ*, *PVR*, *C19orf42*, and *WDTC1*) that hadthe highest SLRvalues were selected.These genes were scatter-plottedin order to compare the means and variances of patients with controls(Figure 3.3). Each dot in the plot represents an individual patient or a control sample. All genes had significantly higher SLRs compared to controls ($p<0.0001$).

Next, ROC analysis was performed for 6 shortening cases (Figure 3.4). This analysis confirmed the sensitivity and specificity of the SLR values of these 6 genes in ER(+) patients compared to controls.The area under the curve represents this discrimination between the patients and controls in terms of sensitivity and specificity.

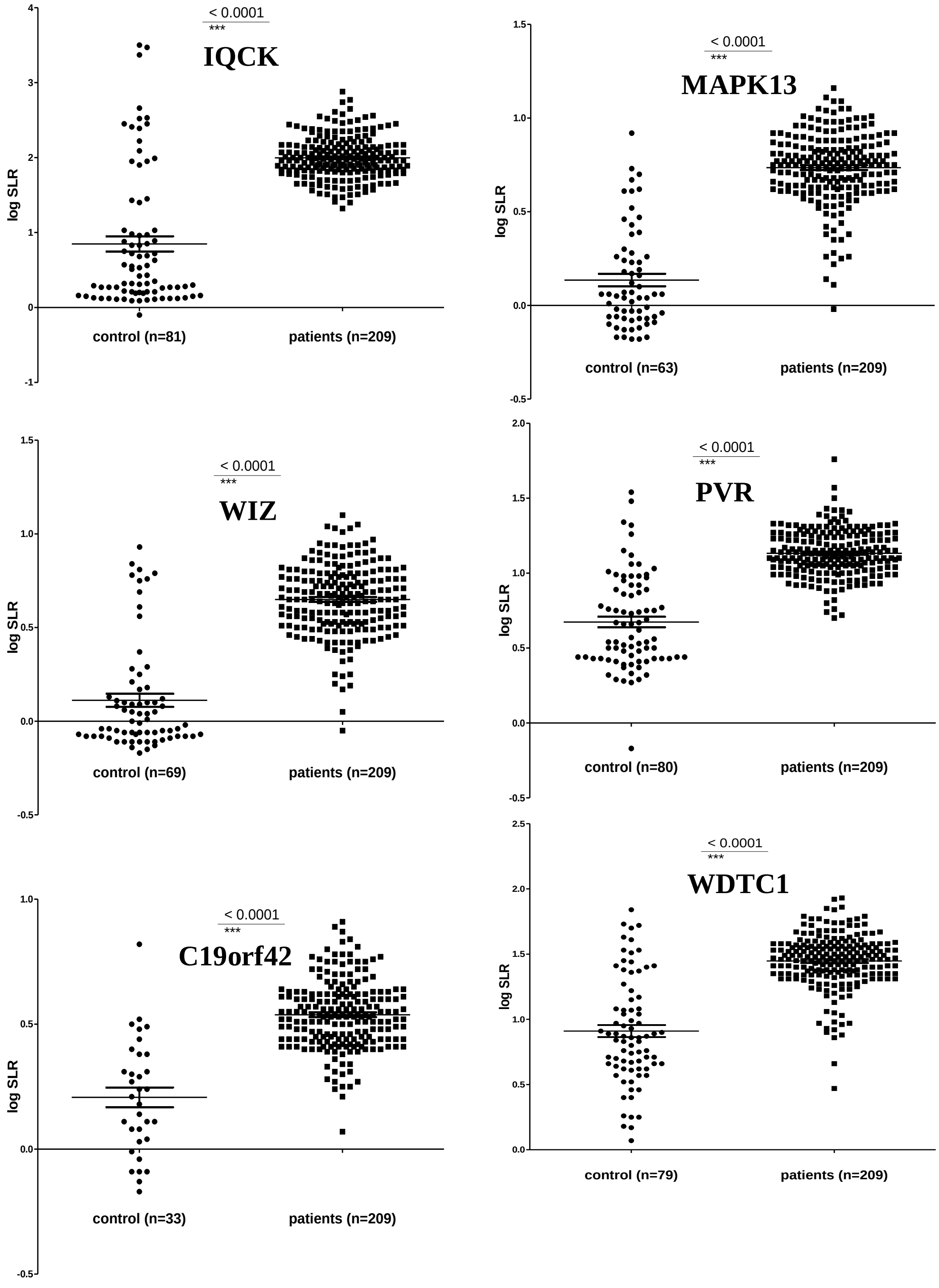


Figure 3.3 Scatter plots of 3'UTR shortened transcripts in ER(+) samples compared to controls. The genes represented are the top 6 shortened genes according to APADetect andSAM. Unpaired t test with Welch's correction was applied to compare means and F-test was used to compare variances of the groups. *** indicates statistical significance for both t-test and F-test with a p value of <0.0001.

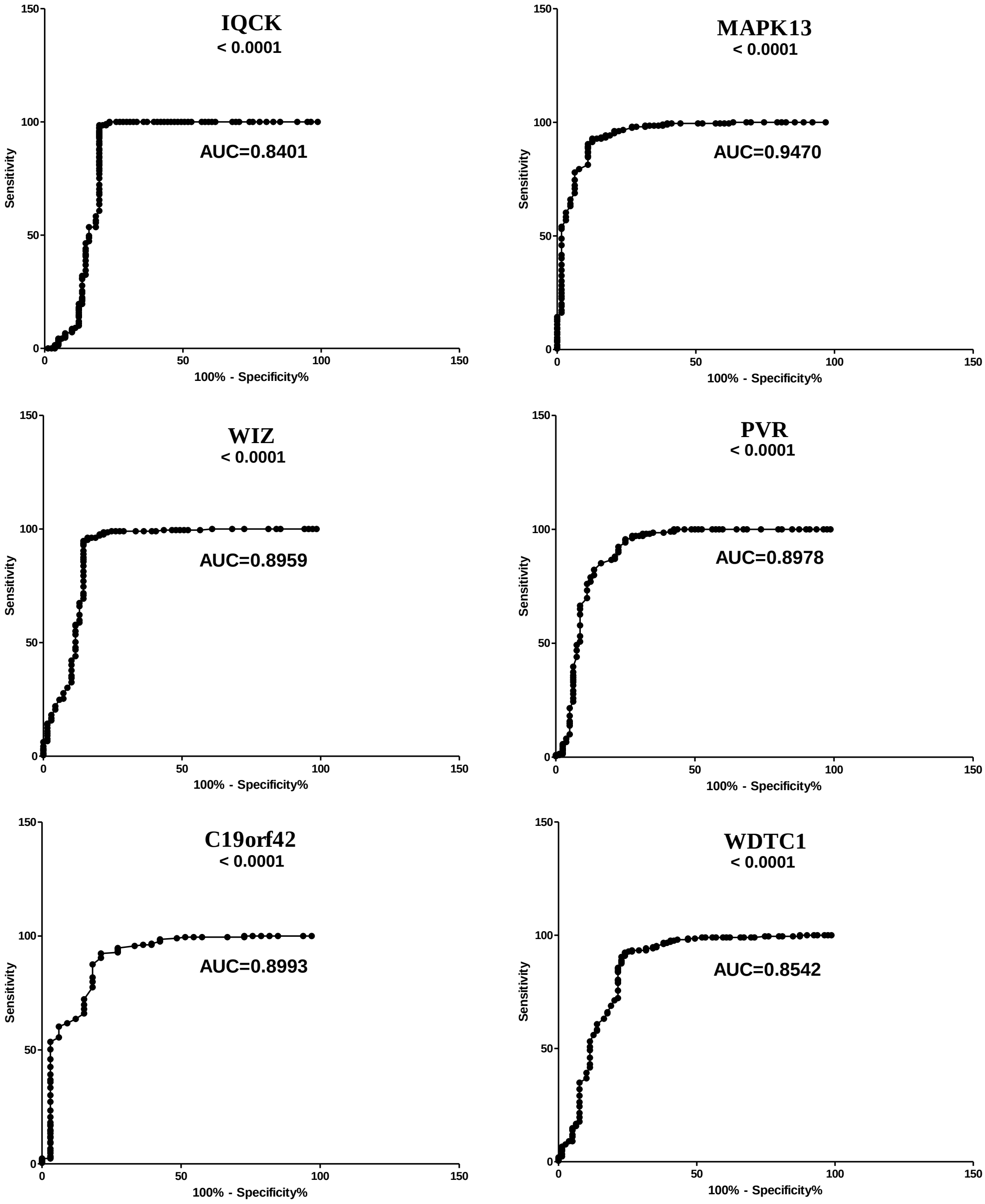


Figure 3.4ROC analyses of the 3’UTR shortened transcripts in ER(+) samples compared to controls. Area under curve (AUC) values are indicated separately for individual genes. For all genes, the p value is <0.0001.

For lengthening cases, 6 cases(*TNPO1*, *SLC46A3*, *WIPI2*, *ERG* 1.5, *ERG* 1.6, and *EIF5B*) having the lowest SLR values were selected. Among these 6 genes, *ERG* had appearedin APADetectand SAM results due to the presence of 2 different poly(A) sites. Therefore, 1.5 and 1.6 represent the different poly(A) sites of *ERG*. As for shortening, scatter plots were generated(Figure 3.5) and ROC analyses were performed (Figure 3.6).

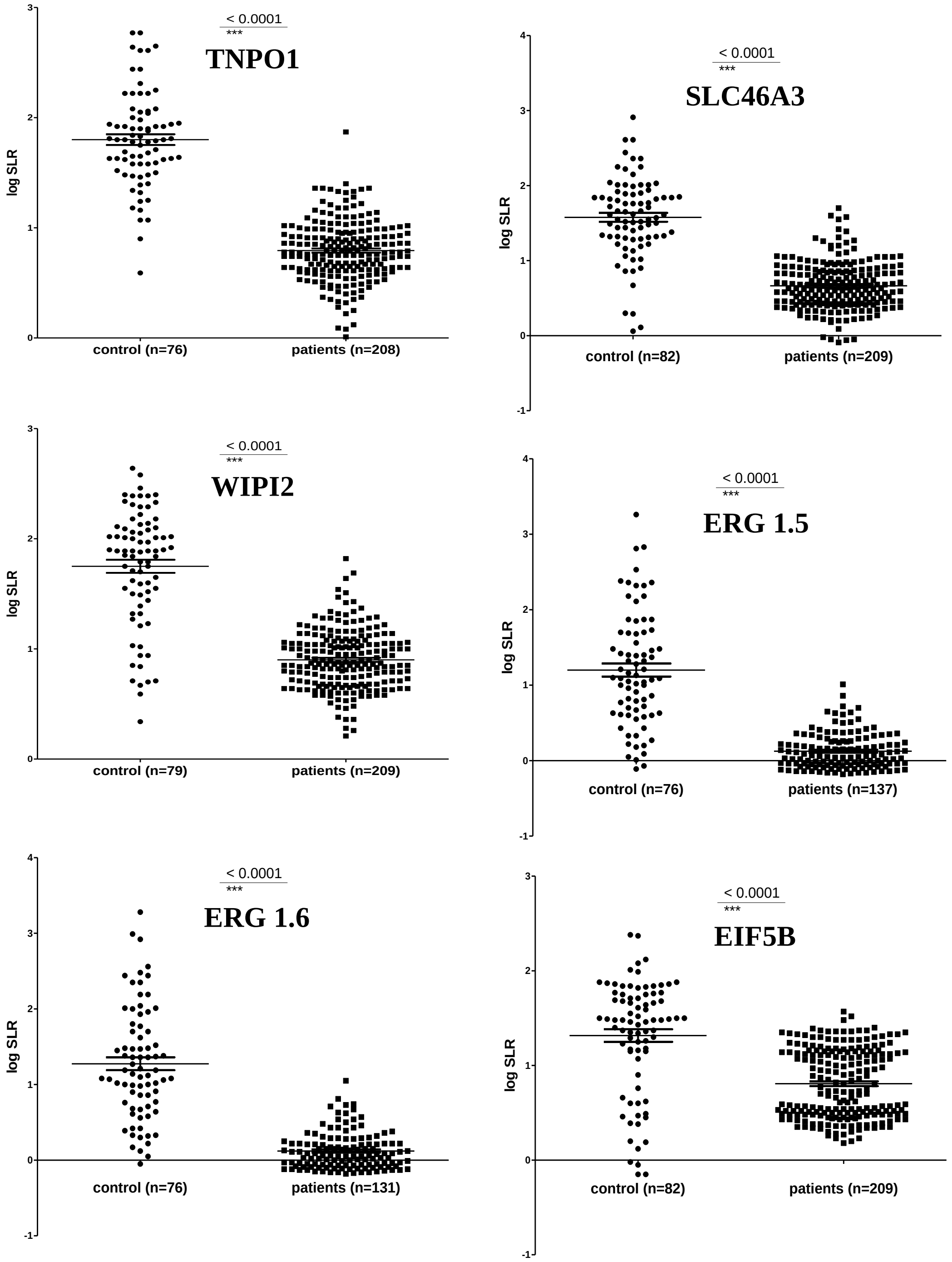


Figure 3.5 Scatter plots of 3'UTR lengthened transcripts in ER(+) samples compared to controls. The genes represented are the top 6 lengthened genes according to APADetect and SAM. Unpaired t test with Welch's correction was applied to compare means and F-test was used to compare variances of the groups. *** indicates statistical significance for both t-test and F-test with a p value of <0.0001.*ERG* 1.5 and *ERG* 1.6 represents the two different poly(A) sites of *ERG*.

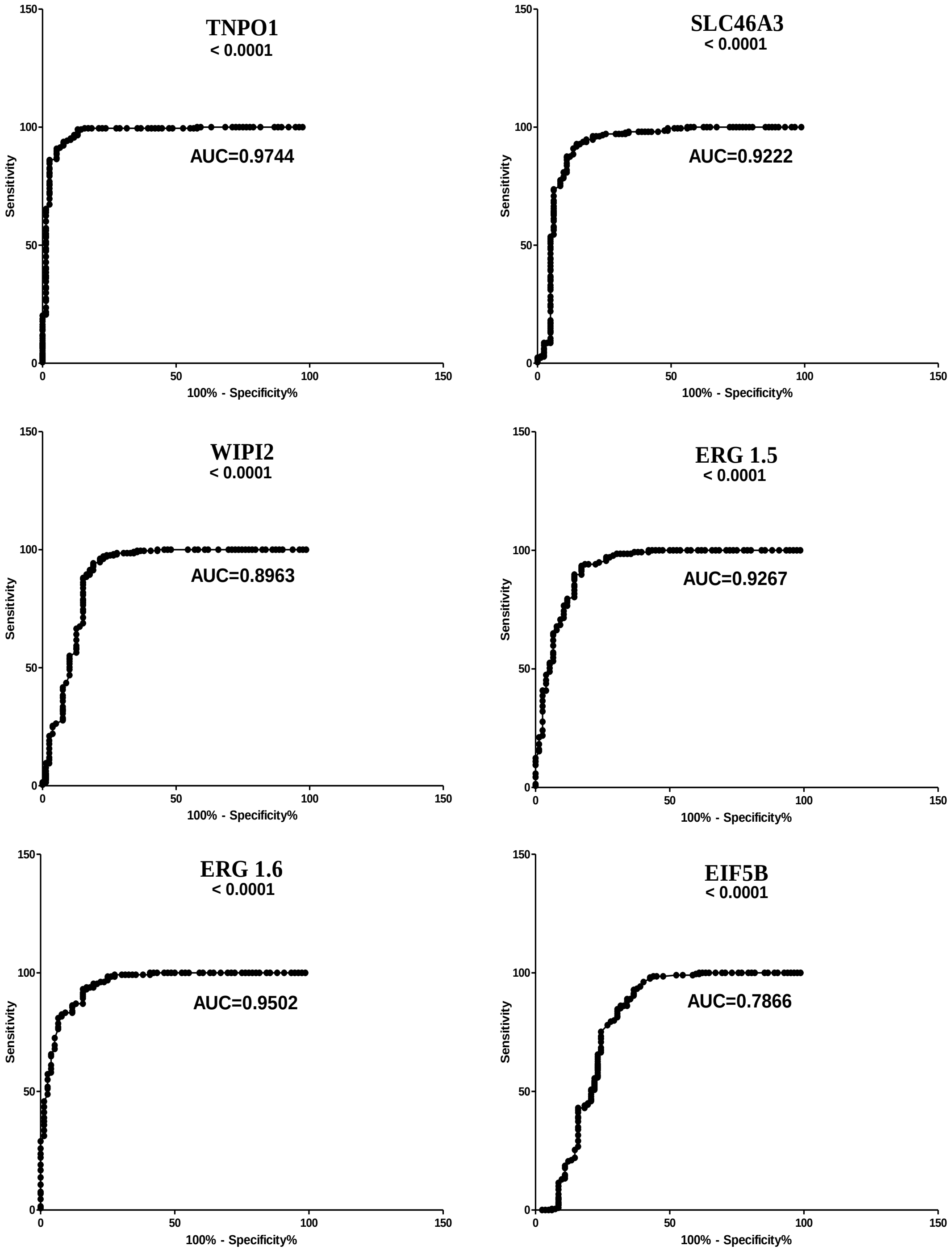


Figure 3.6ROC analyses of the 3'UTR lengthened transcripts in ER(+) samples compared to controls. Area under curve (AUC) values are indicated separately for individual genes. For all the genes, the p value is <0.0001.*ERG* 1.5 and *ERG* 1.6 represents the two different poly(A) sites of *ERG*.

In summary, we confirmed that the selected genes for both shortening and lengthening events had different means and variances in ER(+) breast cancer patients compared to the control group.Furthermore, the specificity and sensitivity of altered 3’UTR lengths in ER(+) patients compared to controls were confirmed with ROC analyses. As a result, we showed that ER(+) patients had deregulations in APA when compared to control group.

To confirm the validity of our approach, we selected 2 genes with unaltered SLR values in patients compared to controls. Individiual scatter plots indeed showed no statistical significance for means of patients vs. controls (Figure 3.7). ROC plots also confirmed that these insignificant SLRs for *SH2D2A* and *LRRC23*were not sensitive or specific to distinguish the 2 groups (Figure 3.8).

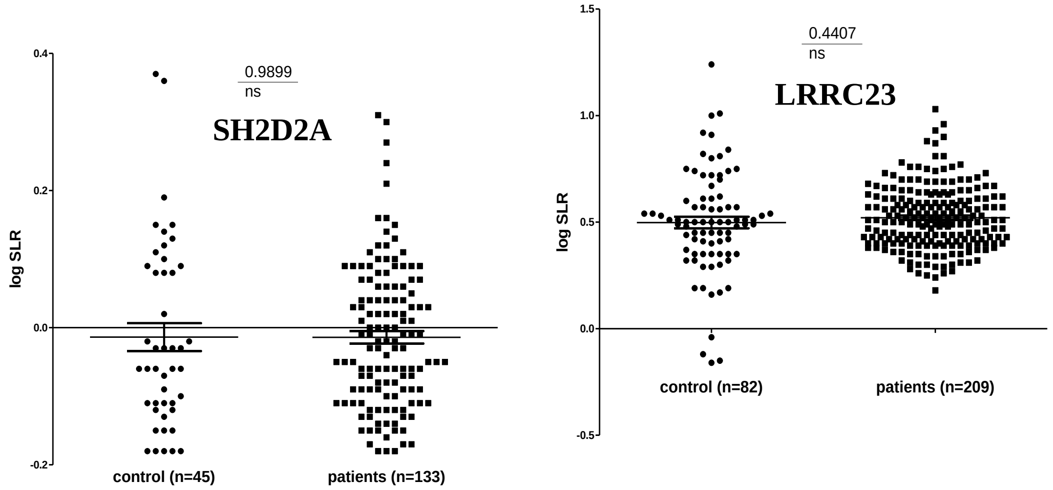


Figure 3.7Scatter plots of 3’UTR unaltered transcripts in ER(+) samples compared to controls. The genes represented are the 2unaltered genes according to APADetect and SAM. Unpaired t test with Welch's correction was applied to compare means and F-test was used to compare variances of the groups.ns indicates statistical insignificance for both t-test and F-test.

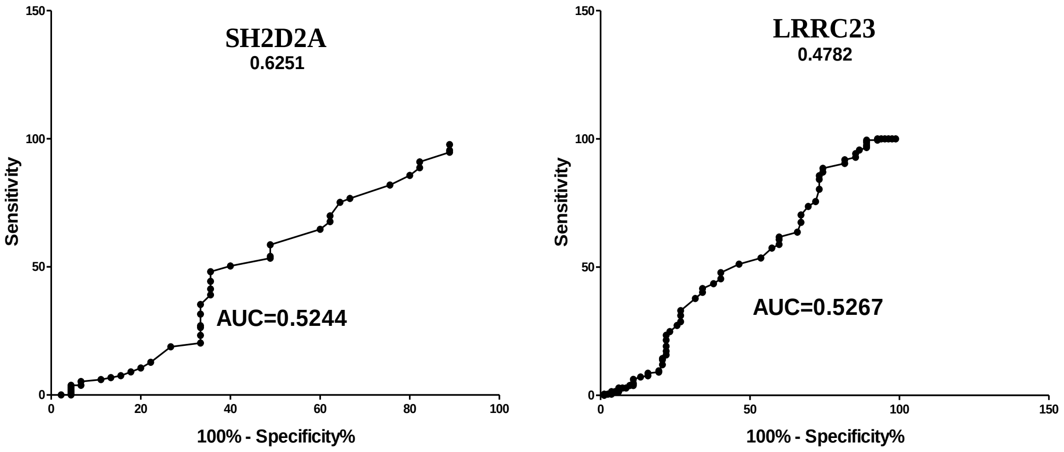


Figure 3.8ROC analyses of the 3'UTR unaltered transcripts in ER(+) samples compared to controls. Area under curve (AUC) values are indicated separately for individual genes. For all the genes, the p value is indicated below the gene name.

3.1.2 APA and Relapse

Given the deregulated APA events in ER(+) patients, we next investigated whether any of these APA events were correlated with the prognosis of patients.In GSE2034 dataset, there are both relapse-free and relapsed patients.In order to examine the changes in 3'UTR lengths among patients in terms of relapse, the patients were divided into two groups: relapse-free patients (n=123) and patients with relapse (n=80). The two groups of patients were compared to each other using APADetect and SAM.There were no significantly altered 3'UTR lengths in relapsed patients compared to relapse-free patients. These observations led to the idea that deregulated APA could be an early event in tumorigenesis.

3.2 APA in GSE7390

To confirm and test our earlier results, we chose to analyze another independent data set of ER(+) breast cancer patientsGSE7390. The difference between this dataset and the previous one is that GSE7390 dataset is a combination of many patient samples from different clinics. However, GSE2034 dataset is uniform; all the patient samples are from the same clinic. Therefore, the two datasets were not combined.

Based on the means of proximal to distal ratios of individual transcripts, the SLR was calculated as an indication of short vs. long 3’UTR abundance. First, we analyzed these patients with controls (normal breast). Next, we investigated whether APA in these patients correlated with relapse status.

3.2.1 APA in ER(+) patients

The normal breast sample set(n=82) was used as the control group. 130 ER(+) breast cancer patient samples of GSE7390 were compared with thecontrols and APADetectanalysis was performed in order to investigate the APA changes between these groups. As a result, shortened and lengthened transcripts according to APADetect and SAM were determined (Figure 3.9).

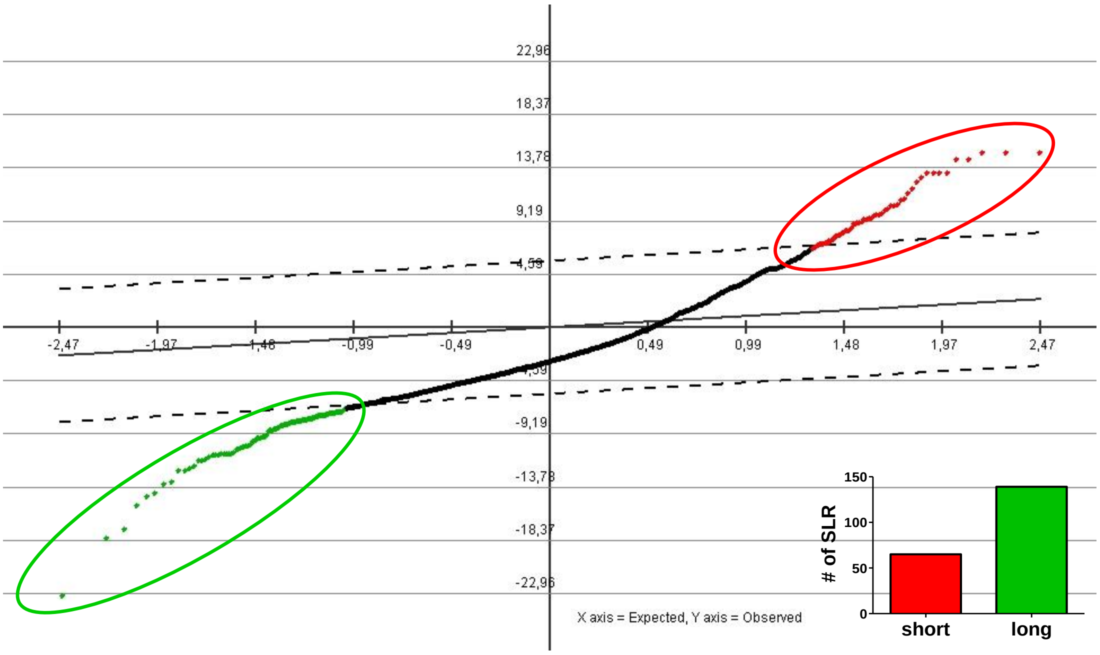


Figure 3.9 SAM graph of normal samples vs. ER(+) samples comparison. The red and green dots represent the shortened and lengthened transcripts in patient samples, respectively. In total, there were65 shortening and 139 lengthening cases.

As a result of SAM, 65 shortened and 139 lengthened transcripts were determined in ER(+) patients(Table B.3 and Table B.4). All of these shortening and lengthening cases were analyzed in terms of ontology. Genes significantly shortened or lengthened also belonged to binding or catalytically active protein groups, similar to our previous findings.

For shortening cases, 6 genes (*GLRX3*, *IQCK*, *APP*, *USP9X*, *GOLPH3*, and *MAPK13*) that had the highest SLR values were selected. These genes were scatter-plotted in order to compare the means and variances of patients with controls and show every data point on the plot (Figure 3.10). Furthermore, ROC analysis was performed for these 6 shortening cases (Figure 3.11).

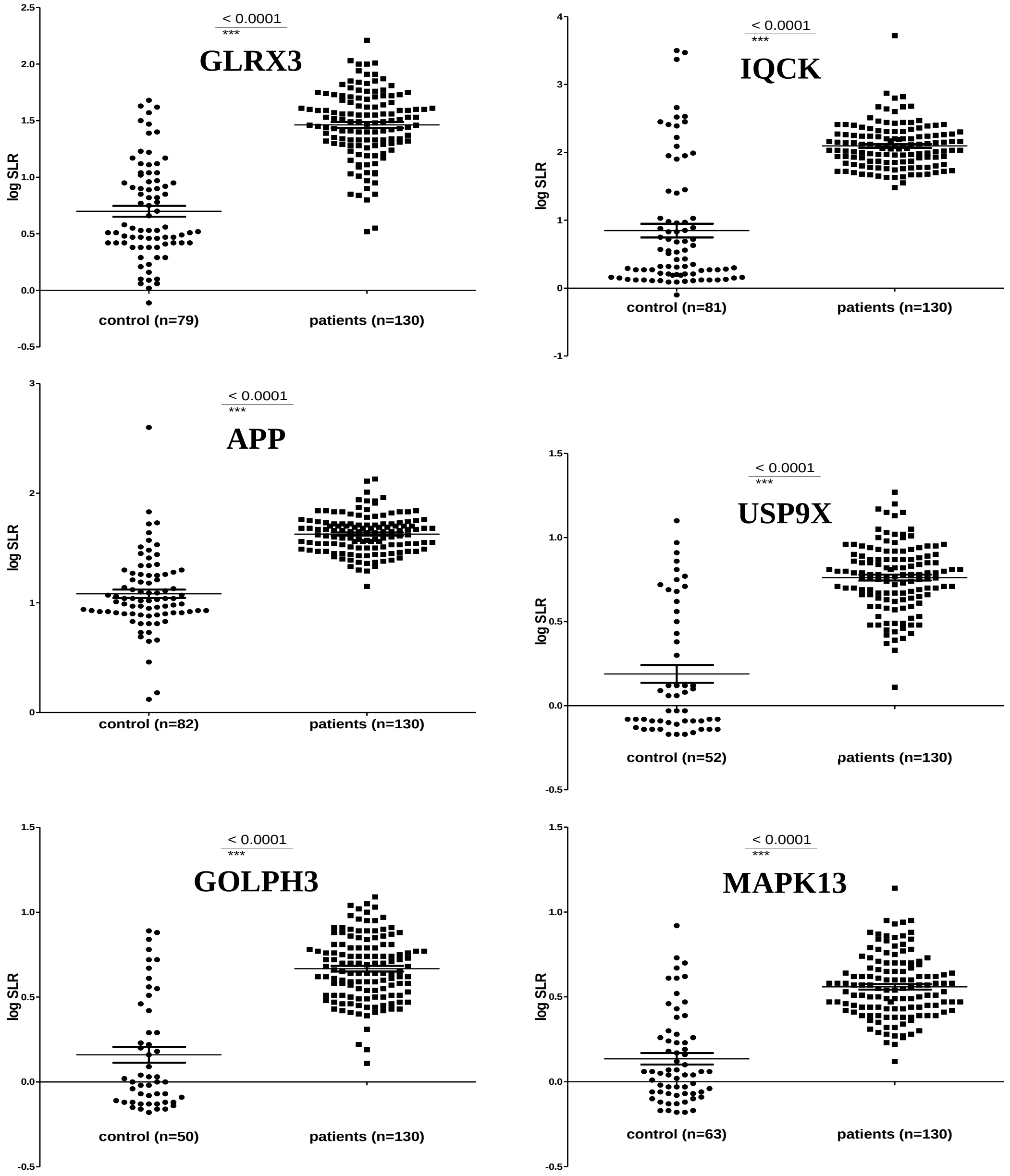


Figure 3.10 Scatter plots of 3'UTR shortened transcripts in ER(+) samples compared to controls. The genes represented are the top 6 shortened genes according to APADetect and SAM. Unpaired t test with Welch's correction was applied to compare means and F-test was used to compare variances of the groups. *** indicates statistical significance for both t-test and F-test with a p value of <0.0001.

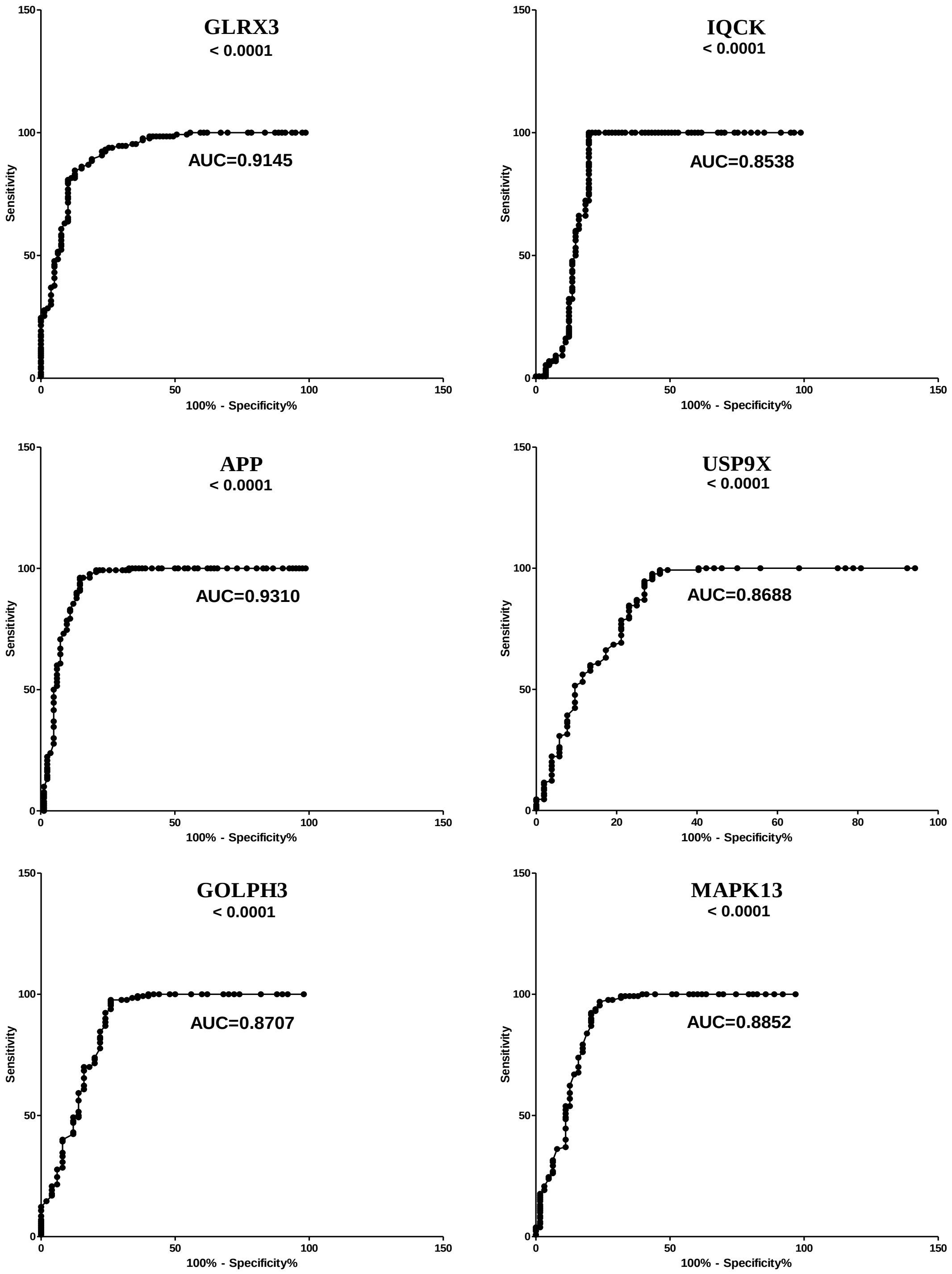


Figure 3.11 ROC analyses of the 3'UTR shortened transcripts in ER(+) samples compared to controls. Area under curve (AUC) values are indicated separately for individual genes. For all the genes, the p value is <0.0001.

For lengthening event, 6 cases (*PABPC3*, *OGT*, *WIPI2* 1.29, *WIPI2* 1.30, *GNS*, and *SLC46A3*) having the lowest SLR values were selected. The two polyadenylation sites for *WIPI2* were significant according to SAM. Therefore, 1.29 and 1.30 represent the differentpolyadenylation sites of *WIPI2*.Scatter plots were generated (Figure 3.12) and ROC analyses were performed (Figure 3.13).

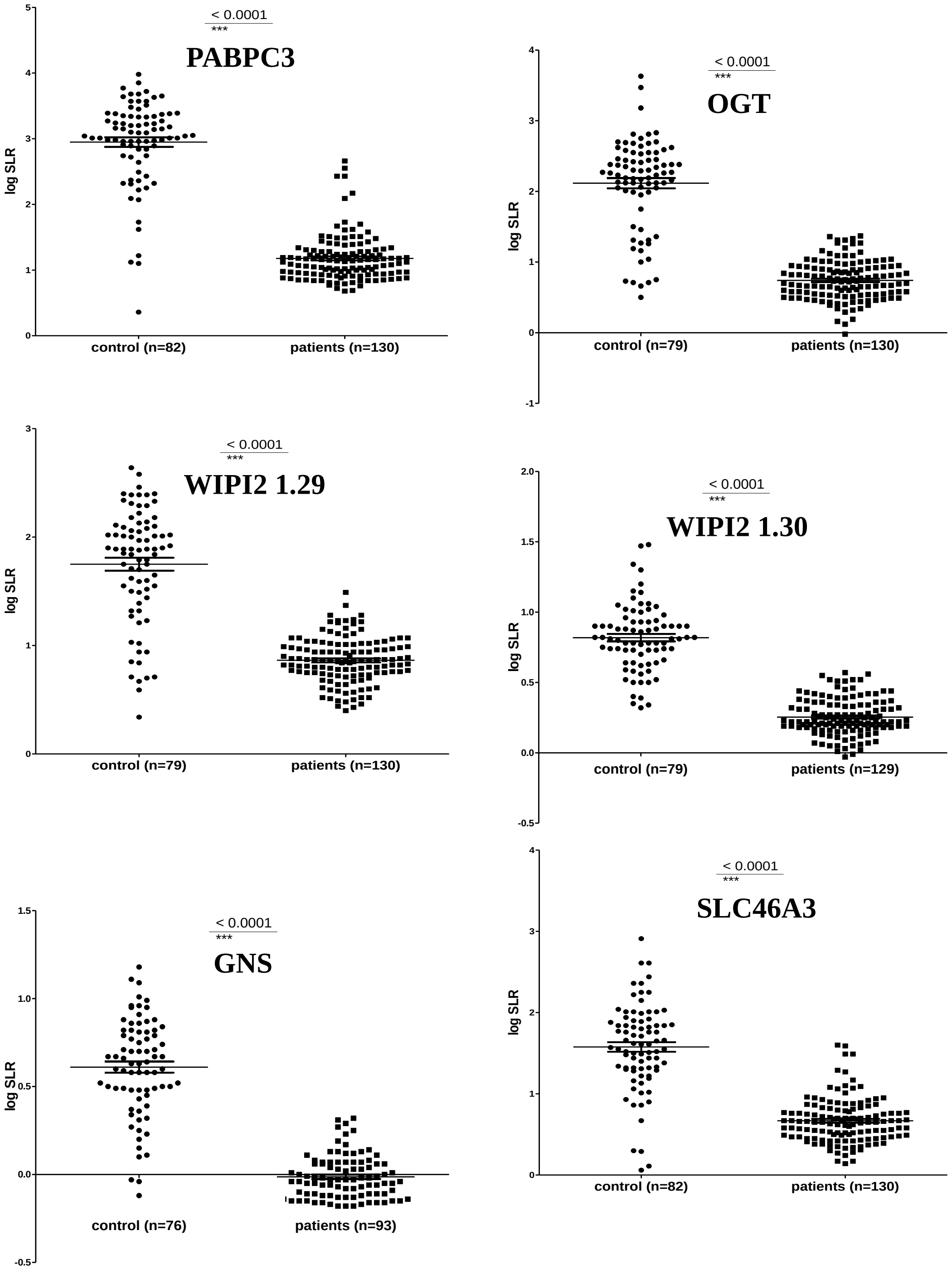


Figure 3.12Scatter plots of 3’UTR lengthened transcripts in ER(+) samples compared to controls. The genes represented are the top 6 lengthened genes according to APADetect and SAM. Unpaired t test with Welch's correction was applied to compare means and F-test was used to compare variances of the groups. *** indicates statistical significance for both t-test and F-test with a p value of <0.0001.*WIPI2* 1.29 and *WIPI2* 1.30 represents the two different poly(A) sites of *WIPI2*.

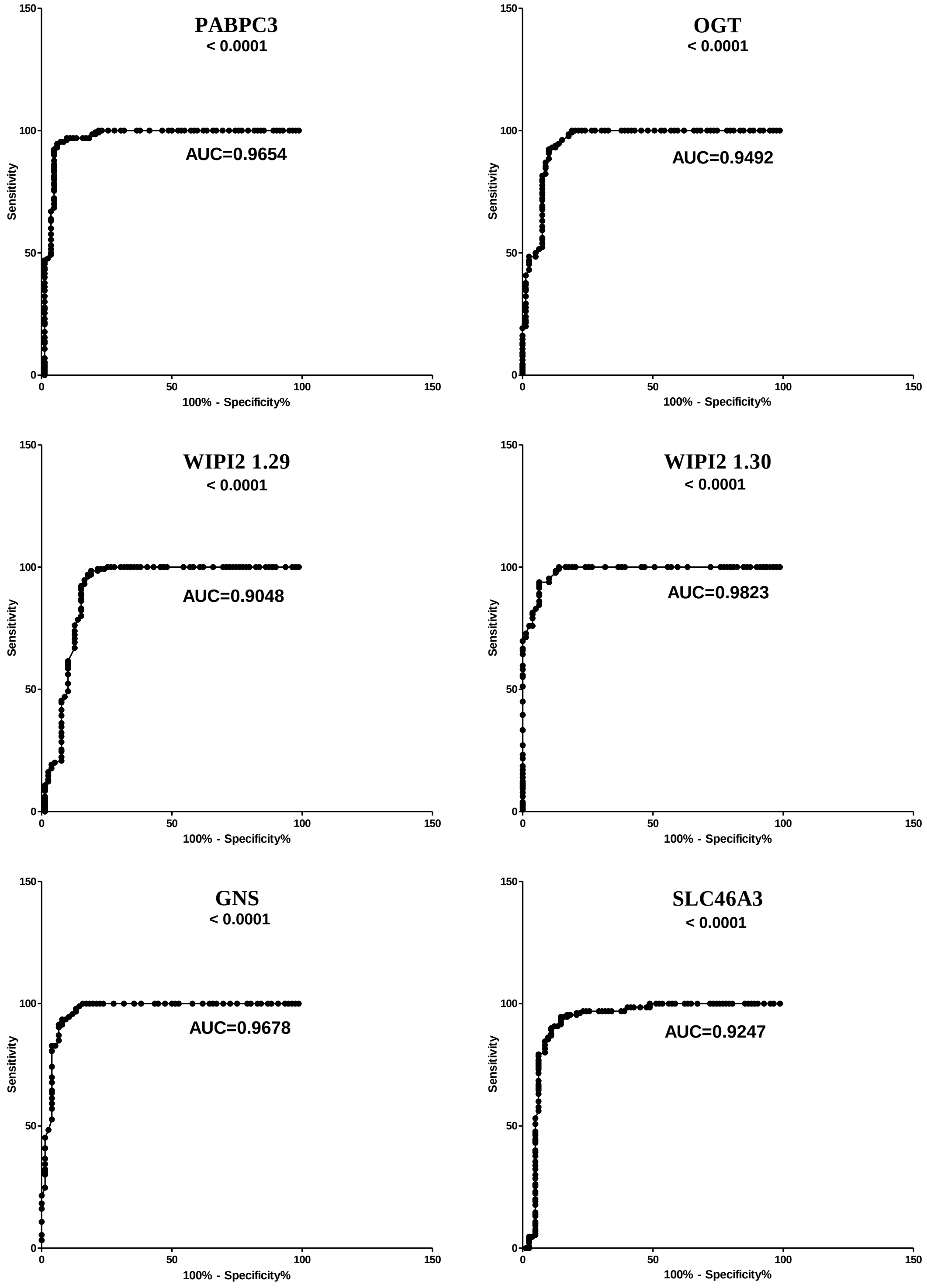


Figure 3.13ROC analyses of the 3'UTR lengthened transcripts in ER(+) samples compared to controls. Area under curve (AUC) values are indicated separately for individual genes. For all the genes, the p value is <0.0001.*WIPI2* 1.29 and *WIPI2* 1.30 represents the two different poly(A) sites of *WIPI2*.

We observed that the selected genes for both shortening and lengthening events have different means and variances in ER(+) breast cancer patients compared to control group. Furthermore, the specificity and sensitivity of these genes were confirmed with ROC analyses.

Next, we investigated whether our GSE2034 and GSE7390 results overlapped in terms of 3’UTR shortening and lengthening events.To determine shared or dataset-specific APA genes for the two analyses, we used theVenny tool (Oliveros, 2007). Our results showed 32 common 3’UTR shortening events (Figure 3.14A) and 35 3’UTR lengthening events (Figure 3.14B). List of common genes are given in Table 3.1. The same group of normal samples was used for the analysis these ER(+) datasets.

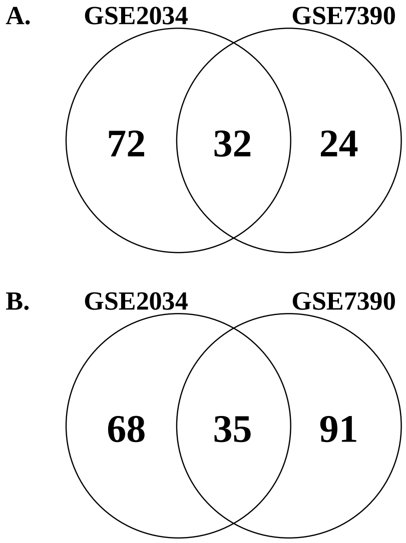


Figure 3.14 The comparison of the GSE2034 and GSE7390 datasets’ APA results. A. Shortening genes of the two datasets were compared. There were 32 common genes shorthened in both of the datasets. **B.** Lengthening genes of the two datasets were compared. There were 35 common genes lengthened in both of the datasets.

Table 3.1 The common alternatively polyadenylatedgenes of the two ER(+) breast cancer patient datasets. 32 and 35 genes were common for shortening and lengthening events, respectively.Bold genes are the selected genes analyzed statistically.

3’UTR shortening			**3’UTR lengthening**		
IQCK	***APP***	***MAPK13***	***SLC46A3***	***WIPI2***	***ERG***
PVR	***USP9X***	*RPL13*	***EIF5B***	***OGT***	*MUT*
HSPA4L	*POFUT1*	*TIMM9*	*AMOT*	*SLC25A31*	*PHKB*
PRSS21	*CLU*	*RLF*	*DOPEY1*	*CCDC93*	*ESF1*
C1orf50	*TPT1*	*METTL9*	*KIF5A*	*ZAK*	*SF4*
WBP5	*IFT57*	*DDX28*	*AZIN1*	*B4GALT1*	*YTHDC2*
HNRNPA1	*SAG*	*RPL31*	*MEF2A*	*AK2*	*TIAL1*
SLC25A17	*SNX3*	*NIPBL*	*ATP6V1C1*	*HFE*	*SUPT7L*
MRPL3	*RPL19*	*DCK*	*STAT6*	*TSHR*	*DKC1*
HMOX2	*COIL*	*G3BP2*	*EBI3*	*TFDP1*	*PLSCR1*
MED14	*ZFP64*		*COL15A1*	*ZMYM1*	*CCND1*
			KDM6A	*NPAS2*	

Although there were common significant genes in both ER(+) patient groups (GSE2034 and GSE7390), we also detected group-specific changes. One of the reasons for this could be the nature of the datasets. For example, while GSE2034 comes from one hospital, GSE7390 samples are collected from 5 hospitals. Another reason could be that breast cancers are quite heterogeneous and then could be some hidden pattern of enriched subtypes.

The genes thatwere found to be statistically significantly different in ER(+) patients compared to controlin both datasets are interesting in terms of their functions and relations with cancer. They were summarized in tablesbelow (Table 3.2 and Table 3.3).

Table 3.2 The 3'UTR shortening genes in the two datasets.

Gene	Function/Relevance to Cancer	References
MAPK13 (mitogen-activated protein kinase 13)	-a member of MAP kinases -proliferation and transcription regulation -induce invasion and motility in cancer	(Tan et al., 2010)
PVR (poliovirus receptor)	-immunoglobulin superfamily -cell-cell attachment,proliferation, migration, and metastasis -upregulated in lung cancer	(Tane et al., 2013)
WDTC1 (WD and tetratricopeptide repeats 1)	-one of the highly mutated genes leading to colorectal cancer	(Alhopuro et al., 2012)
GLRX3 (glutaredoxin 3)	-an oxidoreductase enzyme suppressing apoptosis and promoting cell growth -promoting proliferation, invasion, and survival *in vitro*in breast cancer	(Qu et al., 2011)
APP [amyloid beta (A4) precursor protein]	-a prognostic marker in ER(+) breast cancer patients -high proliferation activity of MCF7 and MDA-MB-231 breast cancer cells	(Takagi et al., 2013)
USP9X (ubiquitin specific peptidase 9, X-linked)	-ubiquitin-specific protease -above 3-fold upregulationin breast cancer -associated with increased beta-catenin levels in MCF7 breast cancer cells -protecting cells against chemotherapy	(Deng et al., 2007),(Murray et al., 2004),(Peddaboina et al., 2012)
GOLPH3 [Golgi phosphoprotein 3 (coat protein)]	-a member of Golgi complex -high expression in approximately 50% of breast cancer -associated with poor survival, high proliferation, and advanced clinical stage -enhancing the expression of the CDK regulator cyclin D1 and inhibited CDK inhibitor p21 (Cip1), p27 (Kip1), and p57 (Kip2)	(Zeng et al., 2012)
IQCK (IQ motif containing K) *WIZ* (widely interspaced zinc finger motifs) *C19orf42* [SMIM7 (small integral membrane protein 7)]	remain to be characterized	

Almost all the shortening genes in ER(+) breast cancer patients are associated with cancer and they have roles in proliferation based on literature (Table 3.2). For example, GLRX3 promotes proliferation and invasion in breast cancer(Qu et al., 2011). Therefore, 3'UTR shortening in certain genes may be considered as a mechanism for oncogene activation.

The functions of lengthening genes in ER(+) patients that we analyzed were also investigated for their relevance to cancer in Table 3.3.

Table 3.3 The 3'UTR lengthening genes in the two datasets.

Gene	Function/Relevance to Cancer	References
TNPO1 (transportin 1)	-the beta subunit of the karyopherin complex -cytoplasm-to-nuclear transport, mitotic spindle and nuclear pore assembly	(Lau et al., 2009)
WIPI2 (WD repeat domain, phosphoinositide interacting 2)	-WD40 repeat proteins -regulation of the assembly of multiprotein complexes by promoting a platform for simultaneous protein-protein interactions -a part of autophagosomes	(Proikas-Cezanne and Robenek, 2011)
ERG (v-ets avian erythroblastosis virus E26 oncogene homolog)	-a member of erythroblast transformation-specific (ETS) family of transcription factors -embryonic development, cell proliferation, differentiation, angiogenesis, inflammation, and apoptosis -overexpressed in prostate carcinoma	(Verdu et al., 2013)
EIF5B (eukaryotic translation initiation factor 5B)	-required for translation in eukaryotes -positioning the initiation methionine tRNA on the mRNA -increased expression due to E2 in MCF7 cells	(Garcia and Nagai, 2011)
PABPC3 [poly(A) binding protein, cytoplasmic 3)	-required for translation initiation and mRNA stability -extensively found in testis during spermatogenesis	(Feral et al., 2001)
OGT [O-linked N-acetylglucosamine (GlcNAc) transferase]	-an enzyme catalyzing the addition of a single N-acetylglucosamine in O-glycosidic linkage to serine and threonine residues -promoting tumor growth both *in vitro* and *in vivo*and decreasing the expression of the cell-cycle inhibitor p27(Kip1) in breast cancer cells -associated with metastasis in breast cancer patients	(Caldwell et al., 2010),(Huang et al., 2013)
GNS [glucosamine (N-acetyl)-6-sulfatase]	-a lysosomal enzyme found in all cells -the accumulation of undegraded substrates in deficiency	(Mok et al., 2003)

As given in Table 3.3, some of these genes were also associated with breast cancer. For example, *OGT* ([O-linked N-acetylglucosamine (GlcNAc) transferase]has been found to promote tumor growth and downregulatethe cell-cycle inhibitor p27(Kip1) in breast cancer cells(Caldwell et al., 2010). Furthermore, *OGT*is known to have 11 transcript variants. It was reported that splicing may lead to use of distal poly(A) sites for these isoforms (Kaida et al., 2010). In connection with our findings, the significance of OGT 3'UTR lengthening and alternative splicing mechanisms remains to be further investigated in breast cancer cells.

3.2.2 APA and Relapse

In order to investigate whether APA alterations correlated with relapse of patients, the patients were divided into two groups: relapse-free patients (n=94) and patients with relapse (n=36). The two groups of patients were compared to each other usingAPADetect and SAM.

As a result of this analysis, we did not observe any 3'UTR lengthdifference in relapsed patients compared to relapse-free patients. This result was consistent with the result of the previous dataset. The reason may be that these events could be an early event in tumorigenesis.

3.3 Role of ER in APA Regulation

Our results showed both 3’UTR shortening and lengthening events in ER(+) patients. It is known that ER expressing breast cancer cells have a high proliferation rate(Fanelli et al., 1996). Interestingly,when MDA-MB-231 cells (triple negative breast cancer cell line) exogenously express ERα, cell proliferation is suppressed, as well as cell motility and invasion (Lazennec et al., 2001). Therefore, to better investigate the role of ER in APA regulation, we chose to use a cell line model system. This model was generated by infecting ERα and ERE-binding defective ERα in to an ER(-) breast cancer cell line, MDA-MB-231 (Nott et al., 2009). Control groups were infected with recombinant adenovirus containing no cDNA. In addition, each group was treated with E2 for 6 hours. Microarray expression data for a set of comparisons is available through GEO (GSE9761). Hence, we compared the groups in terms of APA usingAPADetect and SAM. Based on the means of proximal to distal ratios of individual transcripts, the SLR was calculated as an indication of short vs. long 3’UTR abundance.

3.3.1 APA in ERα-infected MDA-MB-231

We compared the ERα-infected MDA-MB-231 cells with its control group (both treated with E2 for 6 hours) to analyze the APA events.

As a result of APADetect and SAM, 2 shortened transcripts were determined in ERα-infected cells (Table B.5). The scatter plots were graphed forthese two shortened transcripts to test the statistical significance between the groupsin terms of means and variances (Figure 3.15).

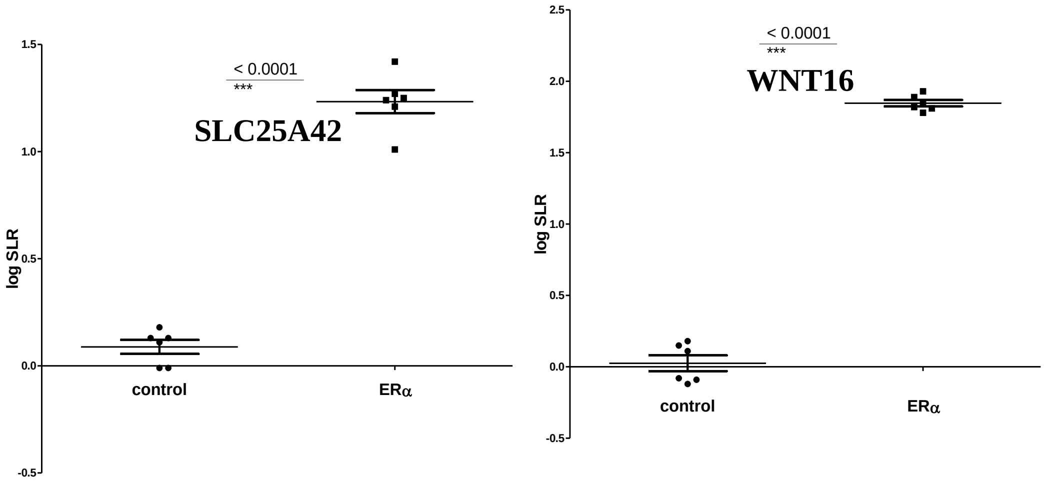


Figure 3.15 Scatter plots of 3’UTR shortened transcripts in ERα-infected MDA-MB-231compared to control group. The genes represented are the shortening genes according to the APADetect and SAM. Unpaired t test with Welch's correction was applied to compare means of the groups. *** indicates statistical significance for t-test with a p value of <0.0001.

For *SLC25A42* and *WNT16*, group means and variances were statistically significantlydifferent in ERα-infected and E2-treated cells compared to just E2-treated cells.

Next, the control group and mutant ERα-infected group was compared to each other. Mutant ERα is unable to bind to ERE; therefore, the effect of ERα is independent of direct DNA binding.APADetect and SAM wereused to find the significantly shortened or lengthened 3’UTRs.

Based on analyses withAPADetect and SAM, no significant genes in the comparison of control and mutant ERα-infected MDA-MB-231 cells were seen. This indicated that 3’UTR length of target genes does not change when ERα does not bind to DNA directly.

To further investigate a potential role of ERα in APA, ERα-infected and mutant ERα-infected groups were compared. This comparison wasexpected to give information about APA events of ERα when it only binds to DNA (Figure 3.16).

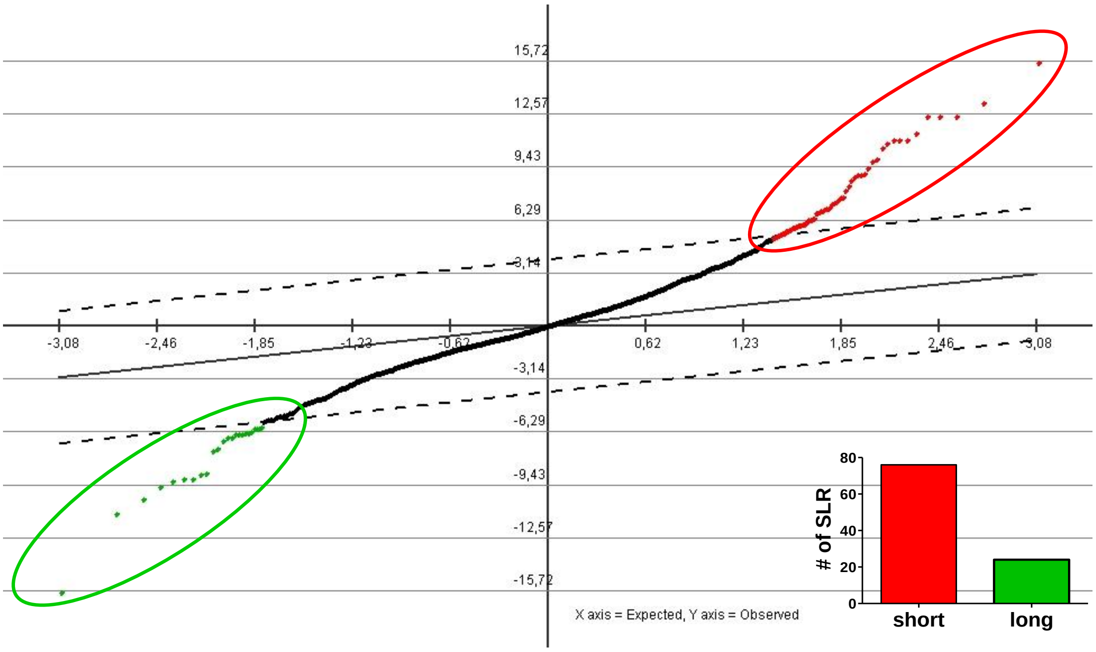


Figure 3.16 SAM graph of ERα-infected MDA-MB-231 compared to mutant ERα-infected group. The red and green dots represent the shortened and lengthened transcripts in ERα-infected cells, respectively. In total, there were 76 shortening and 24 lengthening cases.

A total of 76 shortening and 24 lengthening 3’UTR events were detected in ERα-infected cells compared to DNA-binding defective ERα-infected cells (Table B.6 and Table B.7). For 3’UTR shortening cases, 6 genes (*WNT16*, *CD44*, *LOC440131*, *ST3GAL2*, *MAP7*, and *SRP72*) that had the highest SLR values were selected. These genes were scatter-plotted in order to compare the means and variances of patients with controls (Figure 3.17). For 3’UTR lengthening events, 6 genes (*PML*, *RPS6KB1*, *YME1L1*, *PPP4R1L*, *CLDN22*, and *ADH5*) having the lowest SLR values were selected and scatter plots were generated (Figure 3.18).

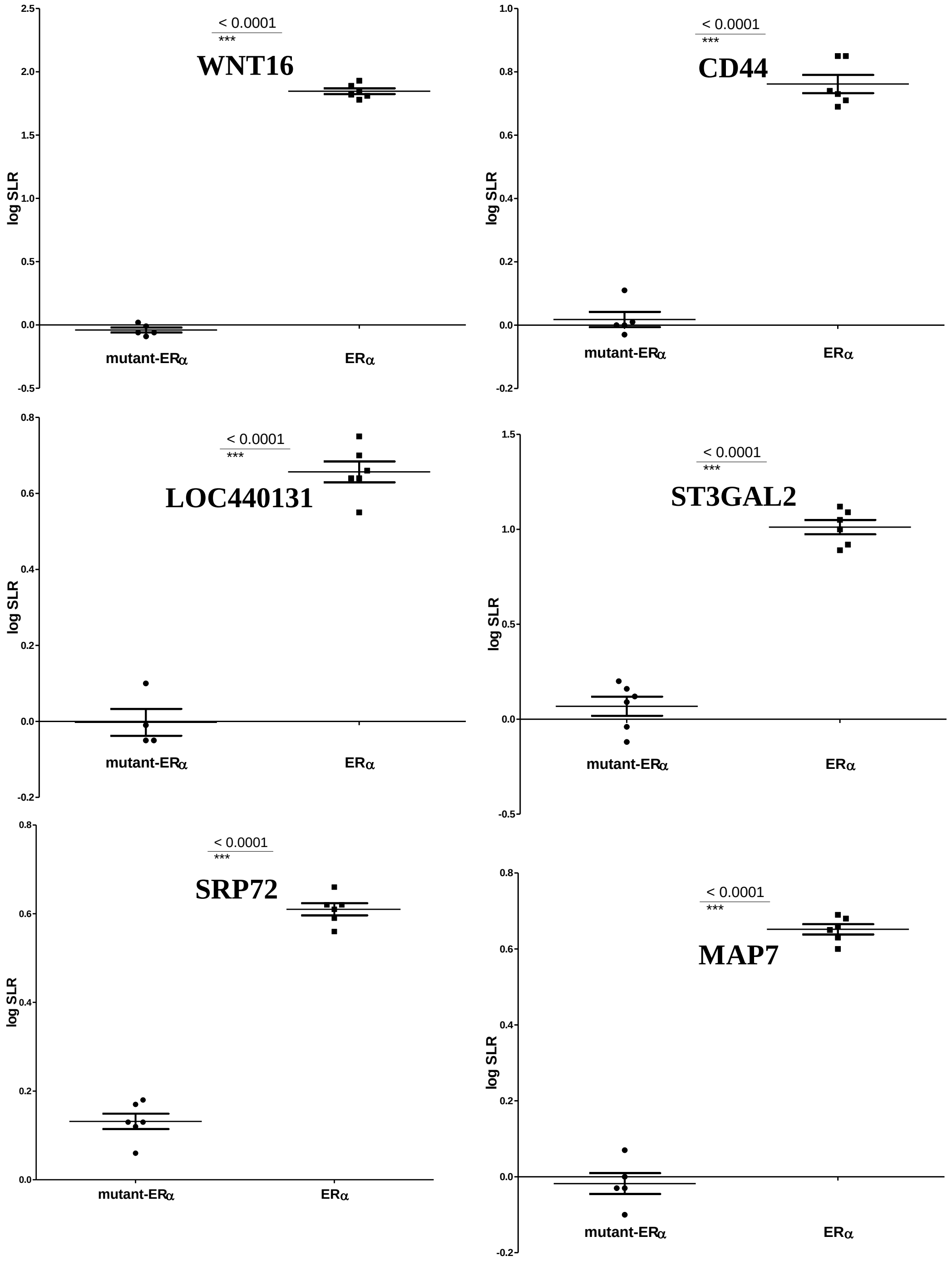


Figure 3.17Scatter plots of 3’UTR shortened transcripts in the ERα-infected MDA-MB-231 compared to mutant ERα-infected cells. The genes represented are the top 6 shortened genes according to the APADetect and SAM. Unpaired t test with Welch's correction was applied to compare means of the groups. *** indicates statistical significance for t-test with a p value of <0.0001.

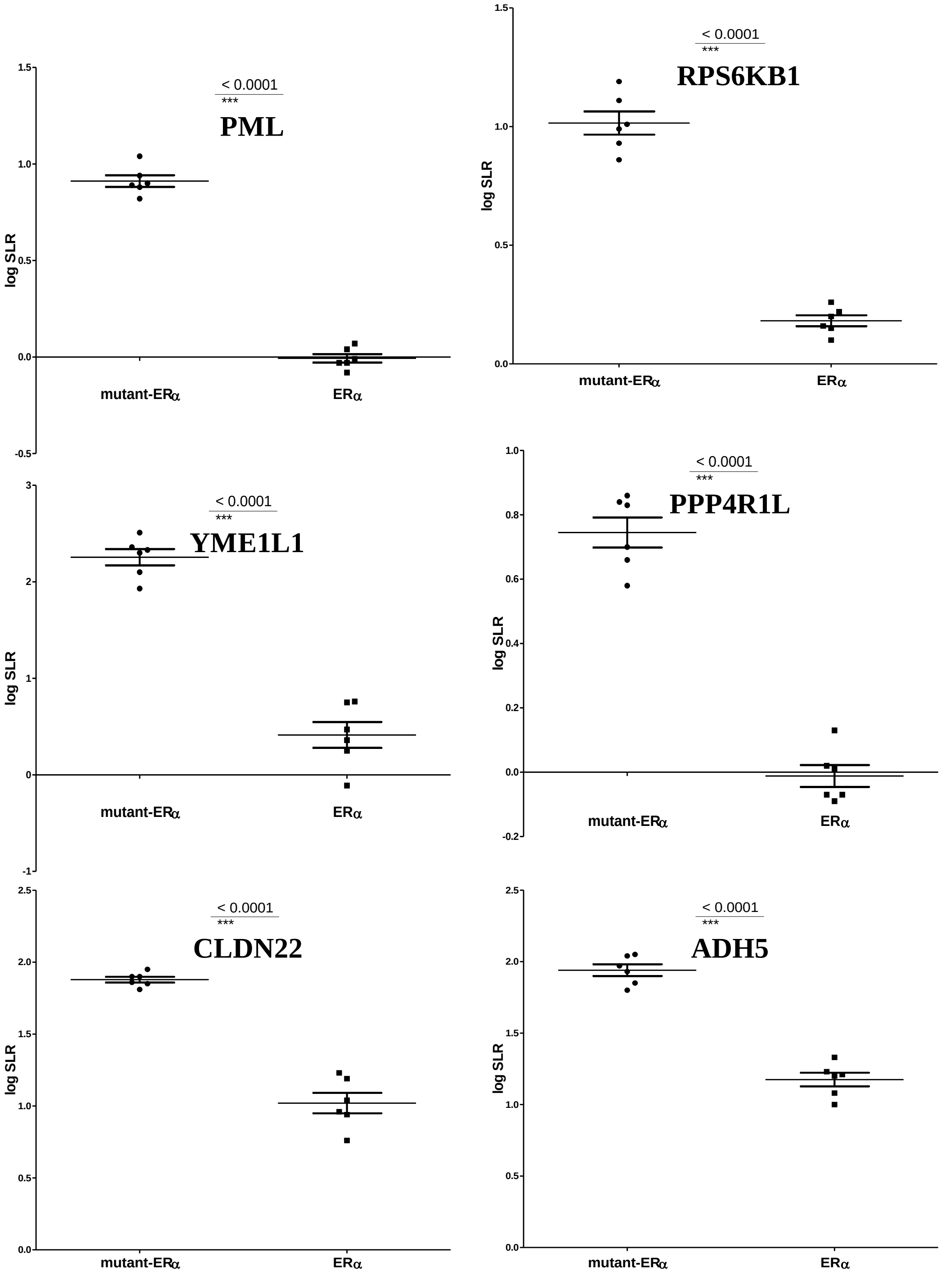


Figure 3.18 Scatter plots of 3'UTR lengthened transcripts in the ERα-infected MDA-MB-231 compared to mutant ERα-infected cells. The genes represented are the 6 3'UTR lengthened genes according to the APADetect and SAM. Unpaired t test with Welch's correction was applied to compare means of the groups. *** indicates statistical significance for t-test with a p value of <0.0001.

We confirmedAPADetect and SAM results for shortened or lengthened 3’UTR genes. Our results suggested higher number of shortening and lengthening cases in ERα-infected MDA-MB-231 cells compared to DNA binding defective cells. This comparison gave information about the APA events in cells that can only bind to DNA. The result was surprising as we expected control vs. ERα-infected cells to demonstrate higher numbers of significant genes. In reality, there were only 2 significant genes in ERα-infected cells.

Both shortening and lengthening genes were searched in literature for their functions and significance in cancer. They were summarized in Table 3.4 and 3.5.

Table 3.4 The 3’UTR shortening genes in the ERα-infected MDA-MB-231 cells.

Gene	Function/Relevance to Cancer	References
SLC25A42 (solute carrier family 25, member 42)	-localized to mitochondria -a mitochondrial transporter for coenzyme A (CoA) and adenosine 3’,5’-diphosphate across the inner mitochondrial membrane	(Fiermonte et al., 2009)
WNT16 (wingless-type MMTV integration site family, member 16)	-a member of WNT gene family encoding secreted signaling proteins -roles in oncogenesis and developmental processes -downregulated or lost in breast cancer cell lines	(Benhaj et al., 2006)
CD44 [CD44 molecule (Indian blood group)]	-a cell-surface glycoprotein -cell-cell interaction, cell adhesion, and migration -increased expression in basal-like breast carcinomas -provides motility to the cells resulting in metastasis	(Kim et al., 2013)
ST3GAL2 (ST3 beta-galactoside alpha-2,3-sialyltransferase 2)	-a type II membrane protein found in Golgi by which sialic acid is catalyzed to galactose-containing substrates -increased in human breast cancer stem cells	(Liang et al., 2013)
MAP7 (microtubule-associated protein 7)	-stabilizing and modulating the microtubules -roles in proliferation in cancer cells with an unknown mechanism	(Yan et al., 2013)
SRP72 (signal recognition particle 72 kDa)	-a subunit of a ribonucleoprotein complex having roles in targeting secretory proteins to the endoplasmic reticulum -enhancing TRAIL (TNF-related apoptosis-inducing ligand)-induced apoptosis in HeLa cells	(Ren et al., 2004)
LOC440131 (long intergenic non-protein coding RNA 544)	Remain to be characterized	

Table 3.5 The 3'UTR lengthening genes in the ERα-infected MDA-MB-231 cells.

Gene	Function/Relevance to Cancer	References
PML (promyelocytic leukemia)	-a member of tripartite (TRIM) family -found in nuclear bodies and functions as a transcription factor and tumor suppressor via modulating the p53 response to oncogenic signals -overexpressed in breast cancer especially in triple-negative cases -associated with poor prognosis	(Carracedo et al., 2012)
RPS6KB1 (ribosomal protein S6 kinase, 70 kDa, polypeptide 1)	-belonging to the ribosomal S6 kinase family of serine/threonine kinases -enhancing cell growth, cell proliferation and protein synthesis -increasing ERα-induced cell proliferation -upregulating anti-apoptotic protein Mcl-1, thus inhibiting apoptosis in MCF7 cells	(Choi et al., 2013)
YME1L1 (YME1-like 1 ATPase)	-role inmitochondrial protein metabolism -promoting apoptotic resistance and cell proliferation, and regulating cristae morphology	(Stiburek et al., 2012)
CLDN22 (claudin 22)	-a member of claudin family which are integral membrane proteins -low expression in breast cancer -related withepithelial-to-mesenchymal transition (EMT) epithelial-to-mesenchymal transition (EMT)	(Prat et al., 2010)
ADH5 [alcohol dehydrogenease 5 (class III), chi polypeptide]	-belonging to alcohol dehydrogenease family -deregulated inpancreatic and colon cancers	(Jelski et al., 2008a, b)
PPP4R1L (protein phosphatase 4, regulatory subunit 1-like)	Remain to be characterized	

In a study conducted in our lab, triple negative breast cancer (TNBC) patients seem to have a tendency towards proximal poly(A) site usage. Because ERα reverses the high proliferation status of MDA-MB-231 cells (triple negative breast cancer cell line) (Lazennec et al., 2001),we hypothesized that the 3’UTR shortened genes may be lengthened in these ERα-infected MDA-MB-231 cells. One of the examples of this situation was*YME1L1* which had significantly shortened 3’UTR in TNBC patients. However, we observed the opposite in the ERα-infected MDA-MB-231 cells. This result may be due to anti-proliferative effect of ERα in this cell line. Some of the other 3’UTR lengthened genes are also highly associated and upregulated in breast cancer. For example, *RPS6KB1* (ribosomal protein S6 kinase, 70 kDa, polypeptide 1) enhances cell growth and proliferation. Therefore, lengthening event of *RPS6KB1* may reverse these functions in this cell line.

Among the 3’UTR shortened genes, there are also highly cancer-related genes. Most of them are involved in breast cancer. Although we expect 3’UTR lengthening for those genes, 3’UTR shortening was observed. Because MDA-MB-231 cells favor distal poly(A) sites, the shortening of these genes may affect some cellular processes that should be investigated experimentally.

Given that transcription is thought to be kinetically coupled to APA (Elkon et al., 2013), involvement of ERα as a transcriptional regulator will be of interest in terms of deregulated APA in ER(+) breast cancer cells. While we do not have any experimental data to answer this question, we speculate that DNA-dependent or DNA-independent ERα actions in cells may operate in complicated ways; there could be additional downstream effects of ER. We also have to consider that E2 treatment and ERα infection in MDA-MB-231 cells may result in an anti-proliferative phenotype which may favor 3’UTR lengthening events. However, MDA-MB-231 cells have already been shown to prefer distal poly(A) sites (Fu et al., 2011). Therefore, these observations should also be repeated in different cell line backgrounds, considering various deregulated other signaling pathways.

CHAPTER 4

CONCLUSION

APA canlead to the production of different transcript isoforms due to the use of alternative poly(A) sites in the 3'UTRs of pre-mRNAs(Liaw et al., 2013). The choice of poly(A) sites may differ in different cellular contexts and physiological states.Therefore, APA deregulation is predicted to be one of the major causes of gene expression abnormalities in various diseases including cancer. In cancer cells, there is a tendency towardsproximal poly(A) site usage and 3'UTR shortening resulting in activation of oncogenes (Mayr and Bartel, 2009).Highly proliferative normal cells also have shorter 3'UTRs to producemore proteins for sustaining the proliferation. On the contrary, in T lymphocytes, forced use of distal poly(A) sites results in decreased protein expression (Sandberg et al., 2008).Proximal poly(A) site choice results with decreased miRNA binding sites due to the shortening of the 3'UTR. Therefore, more transcripts are translated into proteins which is a very favorable situation for oncogenes (An et al., 2013).

Estrogen receptor (ER) overexpression is found approximately 60% of human breast cancers(Boulay et al., 2005).ER is expressed at basal levels in normal mammary gland; however, its expression is higher in breast cancer cells. Two subtypes of ER have been identified as ERα and ERβ. The main ER expressed in breast cancer is ERα,rather than ERβand the main ligand of ER is 17β-estrodiol (E2) (Renoir et al., 2013).E2-ERα complex may regulate the transcription of genes involved in proliferation, differentiation, survival and in the stimulation of invasion, metastasis and angiogenesis.For example, E2 promotes cell proliferation by inducing the expression of anti-apoptotic protein Bcl-2 in MCF7 breast cancer cell line (Perillo et al., 2000). C-myc, cyclins D, A and E are otherexamples of the ER-activated proteins,whilethe cyclin-dependent kinase (CDK)inhibitor p21$^{Waf1/Cip1}$is inhibited by ER-E2 activity (Musgrove et al., 2011).Increased survival and proliferation leads to the E2 dependency of ERα(+) cells; therefore, E2 depletion results in inhibition of the tumor growth. Interestingly, E2-induced proliferation and ERα-regulated transcription activation areinhibited by ERβ if it is present (Gougelet et al., 2005; Paruthiyil et al., 2004; Strom et al., 2004).Given the link between transcription, proliferation and ER, we hypothesized that 3’UTR lengths may be altered due to APA in ER(+) breast cancer patients and cell models.

To address the link between APA and ER, we used publicly available microarray datasetsof ER(+) breast cancer patients from GEO. We separately analyzed two datasets of ER(+) breast cancer patient sampleswith normal breast samples using APADetect and SAM. We selected a few genes that had high SLR for shortening events and low SLR for lengthening events. After statistical analyses, these genes were found to be significantly different between the patient and control groups. These genes are interesting in terms of their functions and relationsto cancer. Almost all the 3’UTR shortening genes enhance proliferation and are upregulated in breast cancer. This pattern was also observed for some lengthening genes.

We also analyzed the APA events in relapsed ER(+) breast cancer patients compared to relapse-free patients using APADetect and SAM. No 3’UTR shortening or lengthening events were detected for both two ER(+) patient datasets. The reason may be that APA events could be an early event in tumorigenesis or that we need to use a higher number of better classified patient samples.

To better investigate the relation between APA and ER, the next approach was the analysis of ERα-infected MDA-MB-231 cells. Because these cells are triple-negative breast cancer cells, we hypothesized that ERα infection may influence the APA events. We compared ERα-infected cells with controls in terms of APA usingAPADetect and SAM. As a result, we detected 2 3’UTR shortening events for *WNT16* and *SLC46A3*. The same analyses were also performed for ERE-binding defective ERα infected cells. When we compared them with the control group, there were no significantly altered genes in terms of APA. However, the comparison of ERα-infected MDA-MB-231 compared to mutant ERα-infected group resulted in 76 shortening and 24 lengthening genesthat are significantly changed. These results suggested to us that most of the genes are 3’UTR shortened or lengthened via directbinding of ERα to DNA.However, we do not have any explanation to why only 2 3’UTR shortening events were detected in E2-treated ERα-infected cells compared to the control group. One contributing factor could be that E2 treatment was for 6 hours; therefore, only very early events could be detected in the analysis.

Our future studies will include more detailed APA analysis in breast cancer. Because ER(+) breast cancer has two subtypes as Luminal A and Luminal B(Schnitt, 2010), subtype specific analysis may give better information about APA patterns. Furthermore, experimental verifications of these shortening and lengthening events will strengthen the results of APADetect. For example, 3’RACE (Rapid Amplification of cDNA Ends) can prove the existance of the short and long isoforms. RT-qPCR studies can show the expression levels of the isoforms.Given the significance of these new findings, we are likely to better understand how different choices of APA may be involved in different physiological and disease states.

REFERENCES

Akman, B. H., T. Can, and A. E. Erson-Bensan, 2012, Estrogen-induced upregulation and 3'-UTR shortening of CDC6: Nucleic Acids Res, v. 40, p. 10679-88.

Alhopuro, P., H. Sammalkorpi, I. Niittymaki, M. Bistrom, A. Raitila, J. Saharinen, K. Nousiainen, H. J. Lehtonen, E. Heliovaara, J. Puhakka, S. Tuupanen, S. Sousa, R. Seruca, A. M. Ferreira, R. M. Hofstra, J. P. Mecklin, H. Jarvinen, A. Ristimaki, T. F. Orntoft, S. Hautaniemi, D. Arango, A. Karhu, and L. A. Aaltonen, 2012, Candidate driver genes in microsatellite-unstable colorectal cancer: Int J Cancer, v. 130, p. 1558-66.

Alimonti, A., A. Carracedo, J. G. Clohessy, L. C. Trotman, C. Nardella, A. Egia, L. Salmena, K. Sampieri, W. J. Haveman, E. Brogi, A. L. Richardson, J. Zhang, and P. P. Pandolfi, 2010, Subtle variations in Pten dose determine cancer susceptibility: Nat Genet, v. 42, p. 454-8.

Alt, F. W., A. L. Bothwell, M. Knapp, E. Siden, E. Mather, M. Koshland, and D. Baltimore, 1980, Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3' ends: Cell, v. 20, p. 293-301.

An, J., X. Zhu, H. Wang, and X. Jin, 2013, A dynamic interplay between alternative polyadenylation and microRNA regulation: implications for cancer (Review): Int J Oncol, v. 43, p. 995-1001.

An, J. J., K. Gharami, G. Y. Liao, N. H. Woo, A. G. Lau, F. Vanevski, E. R. Torre, K. R. Jones, Y. Feng, B. Lu, and B. Xu, 2008, Distinct role of long 3' UTR BDNF mRNA in spine morphology and synaptic plasticity in hippocampal neurons: Cell, v. 134, p. 175-87.

Andreassi, C., and A. Riccio, 2009, To localize or not to localize: mRNA fate is in 3'UTR ends: Trends Cell Biol, v. 19, p. 465-74.

Barrett, T., S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim, M. Tomashevsky, K. A. Marshall, K. H. Phillippy, P. M. Sherman, M. Holko, A. Yefanov, H. Lee, N. Zhang, C. L. Robertson, N. Serova, S. Davis, and A. Soboleva, 2013, NCBI GEO: archive for functional genomics data sets--update: Nucleic Acids Res, v. 41, p. D991-5.

Bava, F. A., C. Eliscovich, P. G. Ferreira, B. Minana, C. Ben-Dov, R. Guigo, J. Valcarcel, and R. Mendez, 2013, CPEB1 coordinates alternative 3'-UTR formation with translational regulation: Nature, v. 495, p. 121-5.

Benhaj, K., K. C. Akcali, and M. Ozturk, 2006, Redundant expression of canonical Wnt ligands in human breast cancer cell lines: Oncol Rep, v. 15, p. 701-7.

Berrar, D., and P. Flach, 2012, Caveats and pitfalls of ROC analysis in clinical microarray research (and how to avoid them): Brief Bioinform, v. 13, p. 83-97.

Bjornstrom, L., and M. Sjoberg, 2005, Mechanisms of estrogen receptor signaling: convergence of genomic and nongenomic actions on target genes: Mol Endocrinol, v. 19, p. 833-42.

Boulay, A., J. Rudloff, J. Ye, S. Zumstein-Mecker, T. O'Reilly, D. B. Evans, S. Chen, and H. A. Lane, 2005, Dual inhibition of mTOR and estrogen receptor signaling in vitro induces cell death in models of breast cancer: Clin Cancer Res, v. 11, p. 5319-28.

Brown, K. M., and G. M. Gilmartin, 2003, A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im: Mol Cell, v. 12, p. 1467-76.

Buratowski, S., 2005, Connections between mRNA 3' end processing and transcription termination: Curr Opin Cell Biol, v. 17, p. 257-61.

Caldon, C. E., R. L. Sutherland, and E. Musgrove, 2010, Cell cycle proteins in epithelial cell differentiation: implications for breast cancer: Cell Cycle, v. 9, p. 1918-28.

Caldwell, S. A., S. R. Jackson, K. S. Shahriari, T. P. Lynch, G. Sethi, S. Walker, K. Vosseller, and M. J. Reginato, 2010, Nutrient sensor O-GlcNAc transferase regulates breast cancer tumorigenesis through targeting of the oncogenic transcription factor FoxM1: Oncogene, v. 29, p. 2831-42.

Carracedo, A., D. Weiss, A. K. Leliaert, M. Bhasin, V. C. de Boer, G. Laurent, A. C. Adams, M. Sundvall, S. J. Song, K. Ito, L. S. Finley, A. Egia, T. Libermann, Z. Gerhart-Hines, P. Puigserver, M. C. Haigis, E. Maratos-Flier, A. L. Richardson, Z. T. Schafer, and P. P. Pandolfi, 2012, A metabolic prosurvival role for PML in breast cancer: J Clin Invest, v. 122, p. 3088-100.

Castelo-Branco, P., A. Furger, M. Wollerton, C. Smith, A. Moreira, and N. Proudfoot, 2004, Polypyrimidine tract binding protein modulates efficiency of polyadenylation: Mol Cell Biol, v. 24, p. 4174-83.

Cheskis, B. J., J. G. Greger, S. Nagpal, and L. P. Freedman, 2007, Signaling by estrogens: J Cell Physiol, v. 213, p. 610-7.

Choi, H. N., H. O. Jin, J. H. Kim, S. E. Hong, H. A. Kim, E. K. Kim, J. K. Lee, I. C. Park, and W. C. Noh, 2013, Inhibition of S6K1 enhances glucose deprivation-induced cell death via downregulation of anti-apoptotic proteins in MCF-7 breast cancer cells: Biochem Biophys Res Commun, v. 432, p. 123-8.

Colgan, D. F., and J. L. Manley, 1997, Mechanism and regulation of mRNA polyadenylation: Genes Dev, v. 11, p. 2755-66.

Cowley, M., A. J. Wood, S. Bohm, R. Schulz, and R. J. Oakey, 2012, Epigenetic control of alternative mRNA processing at the imprinted Herc3/Nap1l5 locus: Nucleic Acids Res, v. 40, p. 8917-26.

D'Ambrogio, A., K. Nagaoka, and J. D. Richter, 2013, Translational control of cell growth and malignancy by the CPEBs: Nat Rev Cancer, v. 13, p. 283-90.

Deng, S., H. Zhou, R. Xiong, Y. Lu, D. Yan, T. Xing, L. Dong, E. Tang, and H. Yang, 2007, Over-expression of genes and proteins of ubiquitin specific peptidases (USPs) and proteasome subunits (PSs) in breast cancer tissue observed by the methods of RFDD-PCR and proteomics: Breast Cancer Res Treat, v. 104, p. 21-30.

Desmedt, C., F. Piette, S. Loi, Y. Wang, F. Lallemand, B. Haibe-Kains, G. Viale, M. Delorenzi, Y. Zhang, M. S. d'Assignies, J. Bergh, R. Lidereau, P. Ellis, A. L. Harris, J. G. Klijn, J. A. Foekens, F. Cardoso, M. J. Piccart, M. Buyse, and C. Sotiriou, 2007, Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series: Clin Cancer Res, v. 13, p. 3207-14.

Di Giammartino, D. C., K. Nishida, and J. L. Manley, 2011, Mechanisms and consequences of alternative polyadenylation: Mol Cell, v. 43, p. 853-66.

Eckmann, C. R., C. Rammelt, and E. Wahle, 2011, Control of poly(A) tail length: Wiley Interdiscip Rev RNA, v. 2, p. 348-61.

Elkon, R., A. P. Ugalde, and R. Agami, 2013, Alternative cleavage and polyadenylation: extent, regulation and function: Nat Rev Genet, v. 14, p. 496-506.

Fanelli, M. A., L. M. Vargas-Roig, F. E. Gago, O. Tello, R. Lucero De Angelis, and D. R. Ciocca, 1996, Estrogen receptors, progesterone receptors, and cell proliferation in human breast cancer: Breast Cancer Res Treat, v. 37, p. 217-28.

Feral, C., G. Guellaen, and A. Pawlak, 2001, Human testis expresses a specific poly(A)-binding protein: Nucleic Acids Res, v. 29, p. 1872-83.

Fiermonte, G., E. Paradies, S. Todisco, C. M. Marobbio, and F. Palmieri, 2009, A novel member of solute carrier family 25 (SLC25A42) is a transporter of coenzyme A and adenosine 3',5'-diphosphate in human mitochondria: J Biol Chem, v. 284, p. 18152-9.

Folkerd, E. J., and M. Dowsett, 2010, Influence of sex hormones on cancer progression: J Clin Oncol, v. 28, p. 4038-44.

Fu, Y., Y. Sun, Y. Li, J. Li, X. Rao, C. Chen, and A. Xu, 2011, Differential genome-wide profiling of tandem 3' UTRs among human breast cancer and normal cells by high-throughput sequencing: Genome Res, v. 21, p. 741-7.

Garcia, S. A., and M. A. Nagai, 2011, Transcriptional regulation of bidirectional gene pairs by 17-beta-estradiol in MCF-7 breast cancer cells: Braz J Med Biol Res, v. 44, p. 112-22.

Gautheret, D., O. Poirot, F. Lopez, S. Audic, and J. M. Claverie, 1998, Alternate polyadenylation in human mRNAs: a large-scale analysis by EST clustering: Genome Res, v. 8, p. 524-30.

Ge, X., S. Yamamoto, S. Tsutsumi, Y. Midorikawa, S. Ihara, S. M. Wang, and H. Aburatani, 2005, Interpreting expression profiles of cancers by genome-wide survey of breadth of expression in normal tissues: Genomics, v. 86, p. 127-41.

Gompel, A., S. Somai, M. Chaouat, A. Kazem, H. J. Kloosterboer, I. Beusman, P. Forgez, M. Mimoun, and W. Rostene, 2000, Hormonal regulation of apoptosis in breast cells and tissues: Steroids, v. 65, p. 593-8.

Gottlicher, M., S. Heck, and P. Herrlich, 1998, Transcriptional cross-talk, the second mode of steroid hormone receptor action: J Mol Med (Berl), v. 76, p. 480-9.

Gougelet, A., C. Bouclier, V. Marsaud, S. Maillard, S. O. Mueller, K. S. Korach, and J. M. Renoir, 2005, Estrogen receptor alpha and beta subtype expression and transactivation capacity are differentially affected by receptor-, hsp90- and immunophilin-ligands in human breast cancer cells: J Steroid Biochem Mol Biol, v. 94, p. 71-81.

Graham, K., A. de las Morenas, A. Tripathi, C. King, M. Kavanah, J. Mendez, M. Stone, J. Slama, M. Miller, G. Antoine, H. Willers, P. Sebastiani, and C. L. Rosenberg, 2010, Gene expression in histologically normal epithelium from breast cancer patients and from cancer-free prophylactic mastectomy patients shares a similar profile: Br J Cancer, v. 102, p. 1284-93.

Heldring, N., A. Pike, S. Andersson, J. Matthews, G. Cheng, J. Hartman, M. Tujague, A. Strom, E. Treuter, M. Warner, and J. A. Gustafsson, 2007, Estrogen receptors: how do they signal and what are their targets: Physiol Rev, v. 87, p. 905-31.

Huang, X., Q. Pan, D. Sun, W. Chen, A. Shen, M. Huang, J. Ding, and M. Geng, 2013, O-GlcNAcylation of Cofilin Promotes Breast Cancer Cell Invasion: J Biol Chem, v. 288, p. 36418-25.

Jelski, W., B. Zalewski, and M. Szmitkowski, 2008a, Alcohol dehydrogenase (ADH) isoenzymes and aldehyde dehydrogenase (ALDH) activity in the sera of patients with pancreatic cancer: Dig Dis Sci, v. 53, p. 2276-80.

Jelski, W., B. Zalewski, and M. Szmitkowski, 2008b, The activity of class I, II, III, and IV alcohol dehydrogenase (ADH) isoenzymes and aldehyde dehydrogenase (ALDH) in liver cancer: Dig Dis Sci, v. 53, p. 2550-5.

Jenal, M., R. Elkon, F. Loayza-Puch, G. van Haaften, U. Kuhn, F. M. Menzies, J. A. Oude Vrielink, A. J. Bos, J. Drost, K. Rooijers, D. C. Rubinsztein, and R. Agami, 2012, The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites: Cell, v. 149, p. 538-53.

Ji, Z., J. Y. Lee, Z. Pan, B. Jiang, and B. Tian, 2009, Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development: Proc Natl Acad Sci U S A, v. 106, p. 7028-33.

Ji, Z., W. Luo, W. Li, M. Hoque, Z. Pan, Y. Zhao, and B. Tian, 2011, Transcriptional activity regulates alternative cleavage and polyadenylation: Mol Syst Biol, v. 7, p. 534.

Kaida, D., M. G. Berg, I. Younis, M. Kasim, L. N. Singh, L. Wan, and G. Dreyfuss, 2010, U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation: Nature, v. 468, p. 664-8.

Kim, K. J., A. Godarova, K. Seedle, M. H. Kim, T. A. Ince, S. I. Wells, J. J. Driscoll, and S. Godar, 2013, Rb Suppresses Collective Invasion, Circulation and Metastasis of Breast Cancer Cells in CD44-Dependent Manner: PLoS One, v. 8, p. e80590.

Kretschmer, C., A. Sterner-Kock, F. Siedentopf, W. Schoenegg, P. M. Schlag, and W. Kemmner, 2011, Identification of early molecular markers for breast cancer: Mol Cancer, v. 10, p. 15.

Lau, C. K., V. A. Delmar, R. C. Chan, Q. Phung, C. Bernis, B. Fichtman, B. A. Rasala, and D. J. Forbes, 2009, Transportin regulates major mitotic assembly events: from spindle to nuclear pore assembly: Mol Biol Cell, v. 20, p. 4043-58.

Lazennec, G., D. Bresson, A. Lucas, C. Chauveau, and F. Vignon, 2001, ER beta inhibits proliferation and invasion of breast cancer cells: Endocrinology, v. 142, p. 4120-30.

Lee, Y. S., and A. Dutta, 2007, The tumor suppressor microRNA let-7 represses the HMGA2 oncogene: Genes Dev, v. 21, p. 1025-30.

Liang, Y. J., Y. Ding, S. B. Levery, M. Lobaton, K. Handa, and S. I. Hakomori, 2013, Differential expression profiles of glycosphingolipids in human breast cancer stem cells vs. cancer non-stem cells: Proc Natl Acad Sci U S A, v. 110, p. 4968-73.

Liaw, H. H., C. C. Lin, H. F. Juan, and H. C. Huang, 2013, Differential microRNA regulation correlates with alternative polyadenylation pattern between breast cancer and normal cells: PLoS One, v. 8, p. e56958.

Licatalosi, D. D., A. Mele, J. J. Fak, J. Ule, M. Kayikci, S. W. Chi, T. A. Clark, A. C. Schweitzer, J. E. Blume, X. Wang, J. C. Darnell, and R. B. Darnell, 2008, HITS-CLIP yields genome-wide insights into brain alternative RNA processing: Nature, v. 456, p. 464-9.

Liu, R., X. Wang, G. Y. Chen, P. Dalerba, A. Gurney, T. Hoey, G. Sherlock, J. Lewicki, K. Shedden, and M. F. Clarke, 2007, The prognostic role of a gene signature from tumorigenic breast-cancer cells: N Engl J Med, v. 356, p. 217-26.

Losel, R., and M. Wehling, 2003, Nongenomic actions of steroid hormones: Nat Rev Mol Cell Biol, v. 4, p. 46-56.

Mansfield, K. D., and J. D. Keene, 2012, Neuron-specific ELAV/Hu proteins suppress HuR mRNA during neuronal differentiation by alternative polyadenylation: Nucleic Acids Res, v. 40, p. 2734-46.

Martincic, K., S. A. Alkan, A. Cheatle, L. Borghesi, and C. Milcarek, 2009, Transcription elongation factor ELL2 directs immunoglobulin secretion in plasma cells by stimulating altered RNA processing: Nat Immunol, v. 10, p. 1102-9.

Matthews, J., and J. A. Gustafsson, 2003, Estrogen signaling: a subtle balance between ER alpha and ER beta: Mol Interv, v. 3, p. 281-92.

Mayr, C., and D. P. Bartel, 2009, Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells: Cell, v. 138, p. 673-84.

Metz, C. E., 1978, Basic principles of ROC analysis: Semin Nucl Med, v. 8, p. 283-98.

Mi, H., A. Muruganujan, and P. D. Thomas, 2013, PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees: Nucleic Acids Res, v. 41, p. D377-86.

Millevoi, S., and S. Vagner, 2010, Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation: Nucleic Acids Res, v. 38, p. 2757-74.

Mok, A., H. Cao, and R. A. Hegele, 2003, Genomic basis of mucopolysaccharidosis type IIID (MIM 252940) revealed by sequencing of GNS encoding N-acetylglucosamine-6-sulfatase: Genomics, v. 81, p. 1-5.

Murray, R. Z., L. A. Jolly, and S. A. Wood, 2004, The FAM deubiquitylating enzyme localizes to multiple points of protein trafficking in epithelia, where it associates with E-cadherin and beta-catenin: Mol Biol Cell, v. 15, p. 1591-9.

Musgrove, E. A., C. E. Caldon, J. Barraclough, A. Stone, and R. L. Sutherland, 2011, Cyclin D as a therapeutic target in cancer: Nat Rev Cancer, v. 11, p. 558-72.

Musgrove, E. A., and R. L. Sutherland, 2009, Biological determinants of endocrine resistance in breast cancer: Nat Rev Cancer, v. 9, p. 631-43.

Nagaike, T., C. Logan, I. Hotta, O. Rozenblatt-Rosen, M. Meyerson, and J. L. Manley, 2011, Transcriptional activators enhance polyadenylation of mRNA precursors: Mol Cell, v. 41, p. 409-18.

Nilsson, S., S. Makela, E. Treuter, M. Tujague, J. Thomsen, G. Andersson, E. Enmark, K. Pettersson, M. Warner, and J. A. Gustafsson, 2001, Mechanisms of estrogen action: Physiol Rev, v. 81, p. 1535-65.

Nott, S. L., Y. Huang, X. Li, B. R. Fluharty, X. Qiu, W. V. Welshons, S. Yeh, and M. Muyan, 2009, Genomic responses from the estrogen-responsive element-dependent signaling pathway mediated by estrogen receptor alpha are required to elicit cellular alterations: J Biol Chem, v. 284, p. 15277-88.

Ozsolak, F., P. Kapranov, S. Foissac, S. W. Kim, E. Fishilevich, A. P. Monaghan, B. John, and P. M. Milos, 2010, Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation: Cell, v. 143, p. 1018-29.

Paruthiyil, S., H. Parmar, V. Kerekatte, G. R. Cunha, G. L. Firestone, and D. C. Leitman, 2004, Estrogen receptor beta inhibits human breast cancer cell proliferation and tumor formation by causing a G2 cell cycle arrest: Cancer Res, v. 64, p. 423-8.

Peddaboina, C., D. Jupiter, S. Fletcher, J. L. Yap, A. Rai, R. P. Tobin, W. Jiang, P. Rascoe, M. K. Rogers, W. R. Smythe, and X. Cao, 2012, The downregulation of Mcl-1 via USP9X inhibition sensitizes solid tumors to Bcl-xl inhibition: BMC Cancer, v. 12, p. 541.

Perillo, B., A. Sasso, C. Abbondanza, and G. Palumbo, 2000, 17beta-estradiol inhibits apoptosis in MCF-7 cells, inducing bcl-2 expression via two estrogen-responsive elements present in the coding sequence: Mol Cell Biol, v. 20, p. 2890-901.

Peterson, M. L., and R. P. Perry, 1989, The regulated production of mu m and mu s mRNA is dependent on the relative efficiencies of mu s poly(A) site usage and the c mu 4-to-M1 splice: Mol Cell Biol, v. 9, p. 726-38.

Pickrell, J. K., J. C. Marioni, A. A. Pai, J. F. Degner, B. E. Engelhardt, E. Nkadori, J. B. Veyrieras, M. Stephens, Y. Gilad, and J. K. Pritchard, 2010, Understanding mechanisms underlying human gene expression variation with RNA sequencing: Nature, v. 464, p. 768-72.

Pinto, P. A., T. Henriques, M. O. Freitas, T. Martins, R. G. Domingues, P. S. Wyrzykowska, P. A. Coelho, A. M. Carmo, C. E. Sunkel, N. J. Proudfoot, and A. Moreira, 2011, RNA polymerase II kinetics in polo polyadenylation signal selection: Embo j, v. 30, p. 2431-44.

Planche, A., M. Bacac, P. Provero, C. Fusco, M. Delorenzi, J. C. Stehle, and I. Stamenkovic, 2011, Identification of prognostic molecular features in the reactive stroma of human breast and prostate cancer: PLoS One, v. 6, p. e18640.

Porter, W., B. Saville, D. Hoivik, and S. Safe, 1997, Functional synergy between the transcription factor Sp1 and the estrogen receptor: Mol Endocrinol, v. 11, p. 1569-80.

Prat, A., J. S. Parker, O. Karginova, C. Fan, C. Livasy, J. I. Herschkowitz, X. He, and C. M. Perou, 2010, Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer: Breast Cancer Res, v. 12, p. R68.

Proikas-Cezanne, T., and H. Robenek, 2011, Freeze-fracture replica immunolabelling reveals human WIPI-1 and WIPI-2 as membrane proteins of autophagosomes: J Cell Mol Med, v. 15, p. 2007-10.

Proudfoot, N. J., 2011, Ending the message: poly(A) signals then and now: Genes Dev, v. 25, p. 1770-82.

Qu, Y., J. Wang, P. S. Ray, H. Guo, J. Huang, M. Shin-Sim, B. A. Bukoye, B. Liu, A. V. Lee, X. Lin, P. Huang, J. W. Martens, A. E. Giuliano, N. Zhang, N. H. Cheng, and X. Cui, 2011, Thioredoxin-like 2 regulates human cancer cell growth and metastasis via redox homeostasis and NF-kappaB signaling: J Clin Invest, v. 121, p. 212-25.

Ren, Y. G., K. W. Wagner, D. A. Knee, P. Aza-Blanc, M. Nasoff, and Q. L. Deveraux, 2004, Differential regulation of the TRAIL death receptors DR4 and DR5 by the signal recognition particle: Mol Biol Cell, v. 15, p. 5064-74.

Renoir, J. M., V. Marsaud, and G. Lazennec, 2013, Estrogen receptor signaling as a target for novel breast cancer therapeutics: Biochem Pharmacol, v. 85, p. 449-65.

Roman-Blas, J. A., S. Castaneda, R. Largo, and G. Herrero-Beaumont, 2009, Osteoarthritis associated with estrogen deficiency: Arthritis Res Ther, v. 11, p. 241.

Rosenfeld, M. G., and C. K. Glass, 2001, Coregulator codes of transcriptional regulation by nuclear receptors: J Biol Chem, v. 276, p. 36865-8.

Rosenwald, A., G. Wright, A. Wiestner, W. C. Chan, J. M. Connors, E. Campo, R. D. Gascoyne, T. M. Grogan, H. K. Muller-Hermelink, E. B. Smeland, M. Chiorazzi, J. M. Giltnane, E. M. Hurt, H. Zhao, L. Averett, S. Henrickson, L. Yang, J. Powell, W. H. Wilson, E. S. Jaffe, R. Simon, R. D. Klausner, E. Montserrat, F. Bosch, T. C. Greiner, D. D. Weisenburger, W. G. Sanger, B. J. Dave, J. C. Lynch, J. Vose, J. O. Armitage, R. I. Fisher, T. P. Miller, M. LeBlanc, G. Ott, S. Kvaloy, H. Holte, J. Delabie, and L. M. Staudt, 2003, The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma: Cancer Cell, v. 3, p. 185-97.

Rozenblatt-Rosen, O., T. Nagaike, J. M. Francis, S. Kaneko, K. A. Glatt, C. M. Hughes, T. LaFramboise, J. L. Manley, and M. Meyerson, 2009, The tumor suppressor Cdc73 functionally associates with CPSF and CstF 3' mRNA processing factors: Proc Natl Acad Sci U S A, v. 106, p. 755-60.

Sandberg, R., J. R. Neilson, A. Sarma, P. A. Sharp, and C. B. Burge, 2008, Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites: Science, v. 320, p. 1643-7.

Schnitt, S. J., 2010, Classification and prognosis of invasive breast cancer: from morphology to molecular taxonomy: Mod Pathol, v. 23 Suppl 2, p. S60-4.

Shi, Y., 2012, Alternative polyadenylation: new insights from global analyses: Rna, v. 18, p. 2105-17.

Shi, Y., D. C. Di Giammartino, D. Taylor, A. Sarkeshik, W. J. Rice, J. R. Yates, 3rd, J. Frank, and J. L. Manley, 2009, Molecular architecture of the human pre-mRNA 3' processing complex: Mol Cell, v. 33, p. 365-76.

Spies, N., C. B. Nielsen, R. A. Padgett, and C. B. Burge, 2009, Biased chromatin signatures around polyadenylation sites and exons: Mol Cell, v. 36, p. 245-54.

Stiburek, L., J. Cesnekova, O. Kostkova, D. Fornuskova, K. Vinsova, L. Wenchich, J. Houstek, and J. Zeman, 2012, YME1L controls the accumulation of respiratory chain subunits and is required for apoptotic resistance, cristae morphogenesis, and cell proliferation: Mol Biol Cell, v. 23, p. 1010-23.

Strom, A., J. Hartman, J. S. Foster, S. Kietz, J. Wimalasena, and J. A. Gustafsson, 2004, Estrogen receptor beta inhibits 17beta-estradiol-stimulated proliferation of the breast cancer cell line T47D: Proc Natl Acad Sci U S A, v. 101, p. 1566-71.

Sun, Y., Y. Fu, Y. Li, and A. Xu, 2012, Genome-wide alternative polyadenylation in animals: insights from high-throughput technologies: J Mol Cell Biol, v. 4, p. 352-61.

Takagi, K., S. Ito, T. Miyazaki, Y. Miki, Y. Shibahara, T. Ishida, M. Watanabe, S. Inoue, H. Sasano, and T. Suzuki, 2013, Amyloid precursor protein in human breast cancer: An androgen-induced gene associated with cell proliferation: Cancer Sci.

Tan, F. L., A. Ooi, D. Huang, J. C. Wong, C. N. Qian, C. Chao, L. Ooi, Y. M. Tan, A. Chung, P. C. Cheow, Z. Zhang, D. Petillo, X. J. Yang, and B. T. Teh, 2010, p38delta/MAPK13 as a diagnostic marker for cholangiocarcinoma and its involvement in cell motility and invasion: Int J Cancer, v. 126, p. 2353-61.

Tane, S., Y. Maniwa, D. Hokka, S. Tauchi, W. Nishio, Y. Okita, and M. Yoshimura, 2013, The role of Necl-5 in the invasive activity of lung adenocarcinoma: Exp Mol Pathol, v. 94, p. 330-5.

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

Tripathi, A., C. King, A. de la Morenas, V. K. Perry, B. Burke, G. A. Antoine, E. F. Hirsch, M. Kavanah, J. Mendez, M. Stone, N. P. Gerry, M. E. Lenburg, and C. L. Rosenberg, 2008, Gene expression abnormalities in histologically normal breast epithelium of breast cancer patients: Int J Cancer, v. 122, p. 1557-66.

Tusher, V. G., R. Tibshirani, and G. Chu, 2001, Significance analysis of microarrays applied to the ionizing radiation response: Proc Natl Acad Sci U S A, v. 98, p. 5116-21.

Verdu, M., I. Trias, R. Roman, N. Rodon, B. Garcia-Pelaez, M. Calvo, A. Dominguez, J. M. Banus, and X. Puig, 2013, ERG expression and prostatic adenocarcinoma: Virchows Arch, v. 462, p. 639-44.

Wang, Y., J. G. Klijn, Y. Zhang, A. M. Sieuwerts, M. P. Look, F. Yang, D. Talantov, M. Timmermans, M. E. Meijer-van Gelder, J. Yu, T. Jatkoe, E. M. Berns, D. Atkins, and J. A. Foekens, 2005, Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer: Lancet, v. 365, p. 671-9.

Wiestner, A., M. Tehrani, M. Chiorazzi, G. Wright, F. Gibellini, K. Nakayama, H. Liu, A. Rosenwald, H. K. Muller-Hermelink, G. Ott, W. C. Chan, T. C. Greiner, D. D. Weisenburger, J. Vose, J. O. Armitage, R. D. Gascoyne, J. M. Connors, E. Campo, E. Montserrat, F. Bosch, E. B. Smeland, S. Kvaloy, H. Holte, J. Delabie, R. I. Fisher, T. M. Grogan, T. P. Miller, W. H. Wilson, E. S. Jaffe, and L. M. Staudt, 2007, Point mutations and genomic deletions in CCND1 create stable truncated cyclin D1 mRNAs that are associated with increased proliferation rate and shorter survival: Blood, v. 109, p. 4599-606.

Wood, A. J., R. G. Roberts, D. Monk, G. E. Moore, R. Schulz, and R. J. Oakey, 2007, A screen for retrotransposed imprinted genes reveals an association between X chromosome homology and maternal germ-line methylation: PLoS Genet, v. 3, p. e20.

Wray, N. R., J. Yang, M. E. Goddard, and P. M. Visscher, 2010, The genetic interpretation of area under the ROC curve in genomic profiling: PLoS Genet, v. 6, p. e1000864.

Yan, X., H. Liang, T. Deng, K. Zhu, S. Zhang, N. Wang, X. Jiang, X. Wang, R. Liu, K. Zen, C. Y. Zhang, Y. Ba, and X. Chen, 2013, The identification of novel targets of miR-16 and characterization of their biological functions in cancer cells: Mol Cancer, v. 12, p. 92.

Yu, E. J., S. H. Kim, K. Heo, C. Y. Ou, M. R. Stallcup, and J. H. Kim, 2011, Reciprocal roles of DBC1 and SIRT1 in regulating estrogen receptor alpha activity and co-activator synergy: Nucleic Acids Res, v. 39, p. 6932-43.

Zeng, Z., H. Lin, X. Zhao, G. Liu, X. Wang, R. Xu, K. Chen, J. Li, and L. Song, 2012, Overexpression of GOLPH3 promotes proliferation and tumorigenicity in breast cancer via suppression of the FOXO1 transcription factor: Clin Cancer Res, v. 18, p. 4059-69.

Zhang, H., J. Hu, M. Recce, and B. Tian, 2005a, PolyA_DB: a database for mammalian mRNA polyadenylation: Nucleic Acids Res, v. 33, p. D116-20.

Zhang, H., J. Y. Lee, and B. Tian, 2005b, Biased alternative polyadenylation in human tissues: Genome Biol, v. 6, p. R100.

APPENDIX A

PATIENT CHARACTERISTICS

Table A.1Patient characteristics of GSE2034 dataset.

GEO accession number	lymph node status	time to relapse or last follow-up (months)	relapse
GSM36777	negative	79	0
GSM36779	negative	132	0
GSM36781	negative	147	0
GSM36782	negative	66	0
GSM36783	negative	52	0
GSM36785	negative	57	0
GSM36786	negative	66	0
GSM36787	negative	72	0
GSM36790	negative	107	0
GSM36794	negative	87	0
GSM36796	negative	118	0
GSM36799	negative	75	0
GSM36801	negative	125	0
GSM36802	negative	92	0
GSM36803	negative	93	0
GSM36804	negative	88	0
GSM36805	negative	100	0
GSM36806	negative	92	0
GSM36807	negative	88	0
GSM36810	negative	105	0
GSM36817	negative	113	0
GSM36819	negative	131	0
GSM36820	negative	141	0
GSM36821	negative	88	0
GSM36823	negative	119	0
GSM36824	negative	86	0
GSM36825	negative	99	0
GSM36829	negative	157	0
GSM36830	negative	117	0
GSM36831	negative	87	0
GSM36832	negative	119	0
GSM36834	negative	109	0
GSM36836	negative	99	0
GSM36837	negative	137	0
GSM36840	negative	86	0
GSM36841	negative	134	0
GSM36842	negative	81	0
GSM36843	negative	77	0
GSM36844	negative	121	0
GSM36845	negative	54	0

Table A.1 (continued)

GEO accession number	lymph node status	time to relapse or last follow-up (months)	relapse
GSM36848	negative	105	0
GSM36849	negative	110	0
GSM36850	negative	121	0
GSM36851	negative	100	0
GSM36852	negative	107	0
GSM36853	negative	105	0
GSM36856	negative	86	0
GSM36857	negative	95	0
GSM36859	negative	130	0
GSM36861	negative	155	0
GSM36864	negative	84	0
GSM36866	negative	107	0
GSM36867	negative	108	0
GSM36868	negative	123	0
GSM36869	negative	126	0
GSM36871	negative	84	0
GSM36873	negative	100	0
GSM36878	negative	98	0
GSM36880	negative	97	0
GSM36882	negative	99	0
GSM36883	negative	88	0
GSM36884	negative	129	0
GSM36887	negative	108	0
GSM36889	negative	108	0
GSM36890	negative	116	0
GSM36892	negative	110	0
GSM36893	negative	124	0
GSM36894	negative	109	0
GSM36895	negative	96	0
GSM36896	negative	102	0
GSM36899	negative	169	0
GSM36900	negative	109	0
GSM36901	negative	101	0
GSM36907	negative	86	0
GSM36910	negative	125	0
GSM36913	negative	109	0
GSM36914	negative	84	0
GSM36916	negative	125	0
GSM36917	negative	144	0
GSM36919	negative	153	0
GSM36921	negative	108	0
GSM36922	negative	113	0
GSM36925	negative	104	0
GSM36929	negative	93	0
GSM36930	negative	84	0
GSM36932	negative	96	0
GSM36933	negative	110	0
GSM36934	negative	171	0
GSM36936	negative	109	0
GSM36938	negative	108	0
GSM36942	negative	152	0
GSM36944	negative	112	0
GSM36945	negative	92	0
GSM36948	negative	87	0
GSM36951	negative	104	0
GSM36958	negative	111	0
GSM36962	negative	114	0
GSM36963	negative	60	0
GSM36965	negative	134	0
GSM36970	negative	116	0
GSM36975	negative	110	0
GSM36979	negative	107	0
GSM36980	negative	122	0
GSM36982	negative	92	0
GSM36984	negative	138	0

Table A.1 (continued)

GEO accession number	lymph node status	time to relapse or last follow-up (months)	relapse
GSM36987	negative	97	0
GSM36988	negative	96	0
GSM36990	negative	143	0
GSM36992	negative	104	0
GSM36993	negative	98	0
GSM36995	negative	113	0
GSM37000	negative	137	0
GSM37009	negative	122	0
GSM37010	negative	101	0
GSM37012	negative	132	0
GSM37014	negative	143	0
GSM37015	negative	122	0
GSM37019	negative	146	0
GSM37024	negative	95	0
GSM37025	negative	107	0
GSM37032	negative	86	0
GSM37033	negative	161	0
GSM37044	negative	112	0
GSM37046	negative	86	0
GSM37047	negative	108	0
GSM37057	negative	59	0
GSM37059	negative	83	0
GSM37060	negative	81	0
GSM37062	negative	54	0
GSM36778	negative	50	1
GSM36784	negative	57	1
GSM36789	negative	51	1
GSM36792	negative	71	1
GSM36811	negative	62	1
GSM36813	negative	49	1
GSM36814	negative	61	1
GSM36815	negative	72	1
GSM36818	negative	62	1
GSM36826	negative	58	1
GSM36838	negative	34	1
GSM36839	negative	32	1
GSM36858	negative	14	1
GSM36860	negative	30	1
GSM36870	negative	30	1
GSM36872	negative	7	1
GSM36874	negative	30	1
GSM36877	negative	43	1
GSM36881	negative	11	1
GSM36885	negative	40	1
GSM36888	negative	28	1
GSM36897	negative	25	1
GSM36898	negative	7	1
GSM36902	negative	28	1
GSM36903	negative	8	1
GSM36908	negative	15	1
GSM36911	negative	80	1
GSM36920	negative	20	1
GSM36924	negative	29	1
GSM36927	negative	8	1
GSM36928	negative	23	1
GSM36931	negative	37	1
GSM36939	negative	14	1
GSM36943	negative	19	1
GSM36946	negative	26	1
GSM36947	negative	16	1
GSM36950	negative	28	1
GSM36954	negative	19	1
GSM36956	negative	32	1
GSM36957	negative	47	1
GSM36960	negative	51	1

Table A.1 (continued)

GEO accession number	lymph node status	time to relapse or last follow-up (months)	relapse
GSM36967	negative	15	1
GSM36971	negative	19	1
GSM36972	negative	37	1
GSM36973	negative	25	1
GSM36974	negative	23	1
GSM36976	negative	38	1
GSM36983	negative	75	1
GSM36985	negative	55	1
GSM36986	negative	77	1
GSM36989	negative	11	1
GSM36994	negative	2	1
GSM36996	negative	24	1
GSM36997	negative	44	1
GSM36998	negative	5	1
GSM36999	negative	37	1
GSM37001	negative	33	1
GSM37003	negative	19	1
GSM37004	negative	35	1
GSM37005	negative	56	1
GSM37006	negative	71	1
GSM37007	negative	59	1
GSM37008	negative	70	1
GSM37011	negative	60	1
GSM37013	negative	66	1
GSM37018	negative	20	1
GSM37026	negative	36	1
GSM37027	negative	17	1
GSM37028	negative	39	1
GSM37029	negative	47	1
GSM37030	negative	44	1
GSM37031	negative	18	1
GSM37035	negative	14	1
GSM37036	negative	15	1
GSM37037	negative	19	1
GSM37038	negative	60	1
GSM37039	negative	48	1
GSM37041	negative	9	1
GSM37051	negative	18	1
GSM37058	negative	64	1

Table A.2Patient characteristics of GSE7390 dataset

Sample ID	node	grade	t.rfs	e.rfs	t.os	e.os	t.dmfs (day)	e.dmfs	t.tdm (day)	e.tdm	hospital
GSM177886	0	3	183	1	6591	0	6591	0	6591	0	GUY
GSM177890	0	2	6507	0	6507	0	6507	0	6507	0	GUY
GSM177893	0	3	6007	0	6007	0	6007	0	6007	0	GUY
GSM177896	0	1	5984	0	5984	0	5984	0	5984	0	GUY
GSM177904	0	2	5693	0	5693	0	5693	0	5693	0	GUY
GSM177905	0	2	5823	0	5823	0	5823	0	5823	0	GUY
GSM177907	0	2	5598	1	5661	0	5661	0	5661	0	GUY
GSM177910	0	1	5110	1	5110	0	5110	0	5110	0	GUY
GSM177911	0	2	5533	0	5533	0	5533	0	5533	0	GUY
GSM177912	0	1	1429	1	5571	0	5571	0	5571	0	GUY
GSM177914	0	3	5379	0	5379	0	5379	0	5379	0	GUY
GSM177915	0	3	4675	0	4675	0	4675	0	4675	0	GUY
GSM177916	0	2	1666	1	4652	0	4652	0	4652	0	GUY
GSM177917	0	2	4227	0	4227	0	4227	0	4227	0	GUY
GSM177919	0	2	4472	0	4472	0	4472	0	4472	0	GUY
GSM177921	0	2	5632	1	7057	0	7057	0	7057	0	IGR
GSM177922	0	2	958	0	958	0	958	0	958	0	IGR
GSM177928	0	2	5685	0	5685	0	5685	0	5685	0	IGR
GSM177930	0	2	4791	0	4791	0	4791	0	4791	0	IGR
GSM177931	0	1	1215	0	1215	0	1215	0	1215	0	IGR
GSM177932	0	1	3779	0	3779	0	3779	0	3779	0	IGR
GSM177933	0	3	5306	0	5306	0	5306	0	5306	0	IGR
GSM177934	0	3	5678	0	5678	0	5678	0	5678	0	IGR
GSM177936	0	2	840	1	5612	0	5612	0	5612	0	IGR
GSM177938	0	1	1104	0	1104	0	1104	0	1104	0	IGR
GSM177940	0	2	1609	1	5238	0	5238	0	5238	0	IGR
GSM177941	0	2	5101	0	5101	0	5101	0	5101	0	IGR
GSM177942	0	3	2809	0	2809	0	2809	0	2809	0	IGR
GSM177944	0	2	546	1	3555	0	3555	0	3555	0	IGR
GSM177945	0	2	3525	1	5433	0	5433	0	5433	0	IGR
GSM177946	0	3	854	1	5228	0	5228	0	5228	0	IGR
GSM177947	0	2	1813	1	5316	0	5316	0	5316	0	IGR
GSM177948	0	1	4925	0	4925	0	4925	0	4925	0	IGR
GSM177949	0	3	3853	1	5012	0	5012	0	5012	0	IGR
GSM177950	0	1	4642	0	4642	0	4642	0	4642	0	IGR
GSM177951	0	1	3311	0	3311	0	3311	0	3311	0	IGR
GSM177952	0	2	4863	0	4863	0	4863	0	4863	0	IGR
GSM177955	0	2	2105	0	2105	0	2105	0	2105	0	IGR
GSM177957	0	1	4830	0	4830	0	4830	0	4830	0	IGR
GSM177959	0	1	4720	0	4720	0	4720	0	4720	0	IGR
GSM177963	0	3	1740	0	1740	0	1740	0	1740	0	IGR
GSM177965	0	2	2536	1	3900	0	3900	0	3900	0	IGR
GSM177967	0	3	3544	0	3544	0	3544	0	3544	0	IGR
GSM177968	0	1	3721	0	3721	0	3721	0	3721	0	IGR
GSM177969	0	2	3771	0	3771	0	3771	0	3771	0	IGR
GSM177974	0	2	4945	1	4952	0	4952	0	4952	0	JRH
GSM177984	0	1	4327	0	4327	0	4327	0	4327	0	JRH
GSM177986	0	3	2083	1	4181	0	4181	0	4181	0	JRH
GSM177987	0	1	3188	0	3188	0	3188	0	3188	0	JRH
GSM177988	0	3	2672	0	2672	0	2672	0	2672	0	JRH
GSM177990	0	1	1591	0	1591	0	1591	0	1591	0	JRH
GSM177991	0	1	1920	0	1920	0	1920	0	1920	0	JRH
GSM177992	0	2	2647	0	2647	0	2647	0	2647	0	JRH
GSM177995	0	2	4568	0	4568	0	4568	0	4568	0	KAR
GSM177996	0	1	4357	0	4357	0	4357	0	4357	0	KAR
GSM177997	0	2	5035	0	5035	0	5035	0	5035	0	KAR
GSM178001	0	3	2549	0	2549	0	2549	0	2549	0	KAR
GSM178002	0	2	4085	0	4085	0	4085	0	4085	0	KAR
GSM178004	0	NA	2441	1	5236	0	5236	0	5236	0	KAR
GSM178005	0	1	5406	0	5406	0	5406	0	5406	0	KAR
GSM178007	0	2	1520	1	5359	0	5359	0	5359	0	KAR
GSM178008	0	1	3964	0	3964	0	3964	0	3964	0	KAR
GSM178011	0	NA	4972	1	6399	0	6399	0	6399	0	KAR
GSM178012	0	3	5683	0	5683	0	5683	0	5683	0	KAR
GSM178014	0	2	6161	0	6161	0	6161	0	6161	0	KAR

Table A.2 (continued)

Sample ID	node	grade	t.rfs	e.rfs	t.os	e.os	t.dmfs (day)	e.dmfs	t.tdm (day)	e.tdm	hospital
GSM178015	0	2	6132	0	6132	0	6132	0	6132	0	KAR
GSM178019	0	2	5014	0	5014	0	5014	0	5014	0	KAR
GSM178024	0	3	6135	0	6135	0	6135	0	6135	0	KAR
GSM178026	0	3	5917	0	5917	0	5917	0	5917	0	KAR
GSM178027	0	2	5909	0	5909	0	5909	0	5909	0	KAR
GSM178029	0	2	5860	0	5860	0	5860	0	5860	0	KAR
GSM178031	0	2	5779	0	5779	0	5779	0	5779	0	KAR
GSM178032	0	3	5779	0	5779	0	5779	0	5779	0	KAR
GSM178033	0	2	5723	0	5723	0	5723	0	5723	0	KAR
GSM178034	0	2	5714	0	5714	0	5714	0	5714	0	KAR
GSM178035	0	3	5653	0	5653	0	5653	0	5653	0	KAR
GSM178037	0	1	5678	0	5678	0	5678	0	5678	0	KAR
GSM178040	0	2	5602	0	5602	0	5602	0	5602	0	KAR
GSM178041	0	2	5768	0	5768	0	5768	0	5768	0	KAR
GSM178042	0	2	5684	0	5684	0	5684	0	5684	0	KAR
GSM178043	0	3	5530	0	5530	0	5530	0	5530	0	KAR
GSM178045	0	1	6352	0	6352	0	6352	0	6352	0	KAR
GSM178046	0	2	722	1	6295	0	6295	0	6295	0	RH
GSM178048	0	2	6281	0	6281	0	6281	0	6281	0	RH
GSM178053	0	2	1857	1	9108	0	9108	0	9108	0	RH
GSM178056	0	3	5181	0	5181	0	5181	0	5181	0	RH
GSM178057	0	3	3855	0	3855	0	3855	0	3855	0	RH
GSM178058	0	3	5156	0	5156	0	5156	0	5156	0	RH
GSM178062	0	2	5038	0	5038	0	5038	0	5038	0	RH
GSM178064	0	3	4555	0	4555	0	4555	0	4555	0	RH
GSM178067	0	2	4866	0	4866	0	4866	0	4866	0	RH
GSM178068	0	3	4178	0	4178	0	4178	0	4178	0	RH
GSM178070	0	2	2541	1	4052	0	4052	0	4052	0	RH
GSM178071	0	2	1432	1	3782	0	3782	0	3782	0	RH
GSM178072	0	1	3492	0	3492	0	3492	0	3492	0	RH
GSM178076	0	2	5137	0	5137	0	5137	0	5137	0	RH
GSM178077	0	2	6961	1	7023	0	7023	0	7023	0	RH
GSM178080	0	1	2225	0	2225	0	2225	0	2225	0	RH
GSM178081	0	2	2722	0	2722	0	2722	0	2722	0	RH
GSM177888	0	3	2192	1	6255	1	6255	1	6255	0	GUY
GSM177889	0	2	3822	1	4133	1	3822	1	3822	1	GUY
GSM177892	0	1	5816	1	5816	1	5816	1	5816	0	GUY
GSM177894	0	2	422	1	1484	1	1233	1	1233	1	GUY
GSM177895	0	3	1136	1	1911	1	1136	1	1136	1	GUY
GSM177900	0	3	994	1	3830	1	994	1	994	1	GUY
GSM177908	0	2	2909	1	2909	1	2909	1	2909	0	GUY
GSM177920	0	3	350	1	1275	1	1206	1	1206	1	GUY
GSM177926	0	3	528	1	6036	1	528	1	528	1	IGR
GSM177929	0	3	1023	1	1754	1	1023	1	1023	1	IGR
GSM177953	0	1	1780	1	3347	1	1815	1	1815	1	IGR
GSM177956	0	3	530	1	929	1	530	1	530	1	IGR
GSM177958	0	3	1721	1	2148	1	1721	1	1721	1	IGR
GSM177960	0	2	1259	1	2453	1	1259	1	1259	1	IGR
GSM177961	0	2	3976	1	4438	0	3976	1	3976	1	IGR
GSM177962	0	2	3331	1	4469	0	4050	1	4050	1	IGR
GSM177964	0	1	3218	1	3608	1	3608	1	3608	0	IGR
GSM177973	0	3	649	1	1029	1	649	1	649	1	JRH
GSM177983	0	2	996	1	2111	1	1731	1	1731	1	JRH
GSM177989	0	2	730	1	794	1	730	1	730	1	JRH
GSM177998	0	2	1465	1	3207	1	2604	1	2604	1	KAR
GSM178000	0	2	3512	1	8251	0	6777	1	6777	0	KAR
GSM178016	0	2	4212	1	5337	1	4212	1	4212	1	KAR
GSM178018	0	2	1703	1	1703	1	1703	1	1703	0	KAR
GSM178020	0	3	880	1	1924	1	880	1	880	1	KAR
GSM178025	0	2	805	1	1803	1	805	1	805	1	KAR
GSM178030	0	2	2349	1	3650	1	2349	1	2349	1	KAR
GSM178038	0	2	1742	1	2697	1	1742	1	1742	1	KAR
GSM178039	0	1	865	1	3614	1	3614	1	3614	0	KAR
GSM178047	0	2	6590	1	7260	0	7185	1	7185	1	RH

Table A.2 (continued)

Sample ID	node	grade	t.rfs	e.rfs	t.os	e.os	t.dmfs (day)	e.dmfs	t.tdm (day)	e.tdm	hospital
GSM178049	0	2	3313	1	4594	1	3313	1	3313	1	RH
GSM178051	0	2	966	1	2972	1	2972	1	2972	0	RH
GSM178054	0	1	4353	1	5445	1	4353	1	4353	1	RH
GSM178060	0	1	794	1	1308	1	794	1	794	1	RH
GSM178061	0	2	1598	1	1864	1	1598	1	1598	1	RH
GSM178074	0	3	2454	1	5284	1	2454	1	2454	1	RH

Table A.3Normal breast samples fromGSE20437, GSE9574, GSE6883, GSE26910, GSE21422, GSE3744, and GSE2361 datasets

GSM number	Sample name
GSM512557	histologically normal breast epithelium from ER+ breast cancer patient sample 1
GSM512558	histologically normal breast epithelium from ER+ breast cancer patient sample 2
GSM512559	histologically normal breast epithelium from ER+ breast cancer patient sample 3
GSM512560	histologically normal breast epithelium from ER+ breast cancer patient sample 4
GSM512561	histologically normal breast epithelium from ER+ breast cancer patient sample 5
GSM512562	histologically normal breast epithelium from ER+ breast cancer patient sample 6
GSM512563	histologically normal breast epithelium from ER+ breast cancer patient sample 7
GSM512564	histologically normal breast epithelium from ER+ breast cancer patient sample 8
GSM512565	histologically normal breast epithelium from ER+ breast cancer patient sample 9
GSM512566	histologically normal breast epithelium from ER- breast cancer patient sample 1
GSM512567	histologically normal breast epithelium from ER- breast cancer patient sample 2
GSM512568	histologically normal breast epithelium from ER- breast cancer patient sample 3
GSM512569	histologically normal breast epithelium from ER- breast cancer patient sample 4
GSM512570	histologically normal breast epithelium from ER- breast cancer patient sample 5
GSM512571	histologically normal breast epithelium from ER- breast cancer patient sample 6
GSM512572	histologically normal breast epithelium from ER- breast cancer patient sample 7
GSM512573	histologically normal breast epithelium from ER- breast cancer patient sample 8
GSM512574	histologically normal breast epithelium from ER- breast cancer patient sample 9
GSM512575	histologically normal breast epithelium from prophylactic mastectomy patient sample 1
GSM512576	histologically normal breast epithelium from prophylactic mastectomy patient sample 2
GSM512577	histologically normal breast epithelium from prophylactic mastectomy patient sample 3
GSM512578	histologically normal breast epithelium from prophylactic mastectomy patient sample 4
GSM512579	histologically normal breast epithelium from prophylactic mastectomy patient sample 5
GSM512580	histologically normal breast epithelium from prophylactic mastectomy patient sample 6
GSM512539	reduction mammoplasty breast epithelium sample 1
GSM512540	reduction mammoplasty breast epithelium sample 2
GSM512541	reduction mammoplasty breast epithelium sample 3
GSM512542	reduction mammoplasty breast epithelium sample 4
GSM512543	reduction mammoplasty breast epithelium sample 5
GSM512544	reduction mammoplasty breast epithelium sample 6
GSM512545	reduction mammoplasty breast epithelium sample 7
GSM512546	reduction mammoplasty breast epithelium sample 8
GSM512547	reduction mammoplasty breast epithelium sample 9
GSM512548	reduction mammoplasty breast epithelium sample 10
GSM512549	reduction mammoplasty breast epithelium sample 11
GSM512550	reduction mammoplasty breast epithelium sample 12
GSM512551	reduction mammoplasty breast epithelium sample 13
GSM512552	reduction mammoplasty breast epithelium sample 14
GSM512553	reduction mammoplasty breast epithelium sample 15
GSM512554	reduction mammoplasty breast epithelium sample 16
GSM512555	reduction mammoplasty breast epithelium sample 17
GSM512556	reduction mammoplasty breast epithelium sample 18
GSM241999	reduction mammoplasty patient 288
GSM242000	reduction mammoplasty patient 278
GSM242001	reduction mammoplasty patient 309
GSM242002	reduction mammoplasty patient 310
GSM242003	reduction mammoplasty patient 314
GSM242004	reduction mammoplasty patient 334
GSM242005	reduction mammoplasty patient 340
GSM242006	reduction mammoplasty patient 350
GSM242007	reduction mammoplasty patient 357
GSM242008	reduction mammoplasty patient 347
GSM242009	reduction mammoplasty patient 352
GSM242010	reduction mammoplasty patient 361
GSM242011	reduction mammoplasty patient 328
GSM242012	reduction mammoplasty patient 360
GSM242013	reduction mammoplasty patient 368
GSM158644	Normal (022_Breast)
GSM158645	Normal (023_Breast)
GSM158646	Normal (068_Breast)
GSM158656	Normal breast (022_Breast_B)
GSM158657	Normal breast (023_Breast_B)
GSM158658	Normal breast (068_Breast_B)

Table A.3 (continued)

GSM number	Sample name
GSM662768	breast normal 1
GSM662770	breast normal 2
GSM662772	breast normal 3
GSM662774	breast normal 4
GSM662776	breast normal 5
GSM662778	breast normal 6
GSM535613	healthy breast 1, biological rep. 1
GSM535614	healthy breast 2, biological rep. 2
GSM535615	healthy breast 3, biological rep. 3
GSM535616	healthy breast 4, biological rep. 4
GSM535617	healthy breast 5, biological rep. 5
GSM85513	NB42 U133p2
GSM85514	NB58 U133p2
GSM85515	NB60 U133p2
GSM85516	NB64 U133p2
GSM85517	NB69 U133p2
GSM85518	NB83 U133p2
GSM85519	NB87 U133p2
GSM44683	Normal Breast

Table A.4Samples of GSE9761 dataset.

sample name	replicates
GSM246267	CMV_rep1
GSM246268	CMV_rep2
GSM246269	CMV_rep3
GSM246270	CMV_rep4
GSM246271	CMV_rep5
GSM246272	CMV_rep6
GSM246273	Era_rep1
GSM246274	Era_rep2
GSM246275	Era_rep3
GSM246276	Era_rep4
GSM246277	Era_rep5
GSM246278	Era_rep6
GSM246279	ERa3411 rep1
GSM246280	ERa3411 rep2
GSM246281	ERa3411 rep3
GSM246282	ERa3411 rep4
GSM246283	ERa3411 rep5
GSM246284	ERa3411 rep6

APPENDIX B

SIGNIFICANTLY 3’UTR ALTERED GENES AS A RESULT OF APADETECT AND SAM

Table B.1 Significantly 3’UTR shortened genes in ER(+) patients from GSE2034 compared to normal breast samples.

Gene name	**Probeset ID**	**Expected score**	**Observed score**	**Numerator**	**Denominator**	**Fold change (unlogged)**
IQCK	215131_at	-0.26693344	14.869099	1.1437807	0.07692334	1.753656
RPL13	212933_x_at	-0.21758601	14.592625	0.70703244	0.048451357	1.61349
MAPK13	210059_s_at	0.57093936	14.274828	0.5123836	0.035894204	1.4094101
WIZ	52005_at	-0.05815504	13.521205	0.49019104	0.036253504	1.3847715
PVR	32699_s_at	0.039425924	13.49765	0.46782953	0.03466007	1.3572268
C19orf42	219097_x_at	0.057249665	13.141416	0.33470854	0.025469748	1.2594177
WDTC1	40829_at	-1.7080995	12.618292	0.554647	0.04395579	1.4233975
NPAT	209798_at	-0.6319465	12.079135	0.5373497	0.044485778	1.4259518
SLC25A16	214140_at	-0.83625805	11.870001	0.43469316	0.036621157	1.3263723
AGTPBP1	204500_s_at	1.06976	11.593541	0.4592939	0.03961636	1.3660648
PAOX	50400_at	-0.75881225	11.485974	0.43052647	0.037482798	1.3491884
MYCL1	215491_at	-1.5378548	10.899136	0.3204515	0.029401552	1.2401507
POT1	204354_at	0.80276525	10.744241	0.39274552	0.03655405	1.299497
SLN	205374_at	-0.63333696	10.706981	0.3123579	0.029173294	1.2408386
PGLYRP4	220944_at	-1.146012	10.509596	0.33258408	0.031645752	1.2405081
RPL19	200029_at	-0.18616603	10.439489	0.37957883	0.036359902	1.3041017
CLU	208792_s_at	0.87892085	9.916545	0.38619912	0.038944926	1.3000784
HSPA4L	205543_at	0.39981884	9.916491	0.35994944	0.036298066	1.264505
LDB1	35160_at	-0.8008531	9.716555	0.45569396	0.04689872	1.3377743
POFUT1	212349_at	0.13705173	9.70073	0.28516483	0.029396223	1.2087847
TIMM9	218316_at	-0.41904286	9.446563	0.39134663	0.04142741	1.2907164
AGTPBP1	204500_s_at	1.0669562	9.424316	0.313274	0.03324103	1.2261332
PRSS21	220051_at	-0.28071043	9.345635	0.32463312	0.034736335	1.2321146
USP9X	201100_s_at	1.3351561	9.2433815	0.38133457	0.041254878	1.2848903
FDX1	203647_s_at	-0.62902963	9.22335	0.26574102	0.02881177	1.1945549
RLF	204243_at	-1.5280606	9.097696	0.299913	0.032965817	1.2294047
PHTF2	215286_s_at	0.73700196	9.0529175	0.33181635	0.036652975	1.246871
PHTF2	215286_s_at	0.7384707	9.0529175	0.33181635	0.036652975	1.246871
FAM82B	218549_s_at	0.9313432	9.006654	0.32319966	0.035884544	1.2374682
LYPLA2	215568_x_at	-1.7960716	8.905746	0.42537212	0.04776378	1.3033253
OFD1	203569_s_at	1.2589577	8.719146	0.43949246	0.05040545	1.3306351
TBC1D15	218268_at	-0.53751785	8.709812	0.3210737	0.03686345	1.2252878
C1orf50	219406_at	-1.5015637	8.590482	0.25156647	0.029284326	1.1815795
TACC1	217437_s_at	0.89118445	8.566732	0.42363197	0.049450822	1.3085457

Table B.1 (continued)

Gene name	Probeset ID	Expected score	Observed score	Numerator	Denominator	Fold change (unlogged)
TBX2	40560_at	-0.14685693	8.564255	0.51393235	0.060009	1.4359604
NRG1	206343_s_at	0.88926274	8.52073	0.33134496	0.038886923	1.2657403
TBX2	40560_at	-0.1477507	8.476936	0.40431833	0.04769628	1.2590455
EGR1	201694_s_at	0.5031329	8.4257965	0.35795766	0.04248354	1.2827574
PTPLA	219654_at	-0.8852287	8.416165	0.25627756	0.030450633	1.1899526
TIMM9	218316_at	-0.4200468	8.381658	0.33096883	0.039487276	1.2339793
TPT1	212284_x_at	-0.47648934	8.377408	0.1736727	0.020731078	1.1246094
METTL9	217868_s_at	-0.26392618	8.33841	0.5333246	0.063959986	1.3889018
PVR	214444_s_at	-0.04031699	8.222619	0.33882824	0.04120685	1.2251234
WBP5	217975_at	1.4871516	8.148816	0.40614477	0.049840953	1.3351613
TPMT	203671_at	0.54041535	8.078473	0.19709487	0.02439754	1.1418058
PALMD	218736_s_at	-1.2455142	7.96067	0.27157688	0.034114826	1.1993032
IVL	214599_at	-1.1494005	7.940681	0.22156075	0.027901983	1.1567066
SARDH	211322_s_at	1.2215593	7.8279085	0.2663512	0.034025844	1.1963962
MFAP3	214588_s_at	0.5146847	7.7609735	0.30899316	0.039813712	1.2041906
MGAT4A	219797_at	0.032474793	7.752417	0.2905742	0.037481755	1.2227991
SFTPB	209810_at	0.024118332	7.574096	0.18124482	0.023929564	1.1263815
KRT12	207811_at	-0.18126981	7.5378127	0.20001416	0.026534773	1.1487439
EXOSC8	215136_s_at	-0.4809082	7.529675	0.27012968	0.035875343	1.1941551
IFT57	218100_s_at	0.2558981	7.4957404	0.2684511	0.03581382	1.1992615
FAM82B	218549_s_at	0.9333978	7.3660865	0.29217163	0.039664432	1.2285534
UGCG	221765_at	1.1353837	7.3658023	0.33864877	0.045975816	1.2493011
MPZL2	203779_s_at	-0.6179394	7.2442994	0.21095188	0.029119708	1.154771
DDX28	40255_at	-0.23714711	7.224147	0.20233501	0.028008154	1.141862
HNRNPA1	214280_x_at	-0.5641692	7.197557	0.40616742	0.05643129	1.3374013
HNRNPA1	214280_x_at	-0.56267875	7.197557	0.40616742	0.05643129	1.3374013
HNRNPA1	214280_x_at	-0.56136537	7.197557	0.40616742	0.05643129	1.3374013
SAG	206671_at	0.11475154	7.146125	0.21828692	0.030546194	1.1562064
RPL31	221593_s_at	0.037939034	7.1178665	0.23845404	0.033500776	1.1622843
YARS2	218470_at	-0.5796557	6.957309	0.3468712	0.049857095	1.2349056
SLC25A17	211754_s_at	0.21678136	6.905259	0.31535548	0.045668885	1.2588645
SRP72	208800_at	0.3571506	6.8988733	0.24633408	0.03570642	1.1817627
LETM1	218939_at	0.32290265	6.886277	0.18962267	0.02753631	1.1329005
NIN	219285_s_at	-0.43087614	6.8584003	0.26933587	0.039270945	1.1762735
NEIL3	219502_at	0.42621672	6.8490834	0.28822485	0.04208225	1.2183038
RET	215771_x_at	-0.86333525	6.8439627	0.26886442	0.039284904	1.2060405
RET	215771_x_at	-0.8651639	6.8439627	0.26886442	0.039284904	1.2060405
FZD10	219764_at	-0.5000687	6.8142424	0.29609346	0.043452147	1.2095503
MKI67	212022_s_at	-0.7712661	6.7941933	0.23654786	0.03481618	1.1690986
SNX3	213545_x_at	0.6201064	6.7117214	0.35243082	0.05250975	1.1928873
NIPBL	212483_at	0.44325668	6.6758566	0.24323922	0.036435656	1.1643995
PCMT1	208857_s_at	0.6556854	6.655611	0.2578592	0.03874313	1.1757983
MRPL3	208787_at	0.28139162	6.627436	0.21322565	0.03217317	1.1622691
SERPINB8	206034_at	-0.09201695	6.580668	0.18826324	0.02860853	1.1361245
BICD1	204741_at	-0.58254224	6.5751505	0.24476182	0.03722528	1.181069
GLRA3	210661_at	0.42282984	6.5533133	0.28405774	0.043345667	1.1904067
TNN	215271_at	-1.0585968	6.5432324	0.27891535	0.042626537	1.2049855
PSAP	200871_s_at	-0.8293218	6.535729	0.22779618	0.034853984	1.1695193
SMAD6	207069_s_at	-0.31943175	6.5300903	0.2581066	0.03952573	1.1804172
SMAD6	207069_s_at	-0.31846628	6.5300903	0.2581066	0.03952573	1.1804172
CD5L	206680_at	-1.1058527	6.4847617	0.27231544	0.04199313	1.1866862
APP	200602_at	0.16696748	6.449291	0.24729216	0.038344085	1.1610193
YARS2	218470_at	-0.57834166	6.434623	0.28810084	0.04477354	1.2147344
LRCH3	214739_at	0.31967998	6.408542	0.21978715	0.03429597	1.1640075
SMAD6	207069_s_at	-0.31732875	6.40584	0.23596099	0.036835294	1.1743534
DCK	203302_at	0.36249307	6.3869004	0.29562962	0.04628687	1.2243396
HMOX2	218121_at	-0.27578744	6.3733997	0.2683984	0.042112283	1.1824423
MGAT2	203102_s_at	-0.43538448	6.3628335	0.2952143	0.046396673	1.1975689
TOP2A	201291_s_at	-0.18304668	6.3537865	0.24130306	0.037977837	1.1943532
NLRP3	207075_at	-0.9214197	6.3512645	0.16723514	0.026330998	1.1199591
DIRAS3	215506_s_at	-1.3125739	6.345331	0.17653528	0.027821288	1.1294609
PFDN4	205362_s_at	0.15682939	6.3147616	0.2647639	0.041927774	1.1632806
PFDN4	205362_s_at	0.15779015	6.3147616	0.2647639	0.041927774	1.1632806
ACTR8	218658_s_at	0.24864465	6.265595	0.18322831	0.029243562	1.1235654

Table B.1 (continued)

Gene name	Probeset ID	Expected score	Observed score	Numerator	Denominator	Fold change (unlogged)
MRP63	204386_s_at	-0.4968153	6.2077475	0.21331811	0.034363206	1.1375813
C2orf18	219783_at	-0.00378709	6.185712	0.21685785	0.035057865	1.1432338
SLC16A3	217691_x_at	-0.1289822	6.141049	0.4108246	0.066898115	1.3579204
COIL	203654_s_at	-0.15512627	6.119961	0.24686179	0.040337153	1.1849078
G3BP2	208841_s_at	0.3694926	6.092217	0.21011353	0.034488846	1.1452626
ARSJ	219973_at	0.3966821	6.0529175	0.15638843	0.025836868	1.1158233
TBC1D9	212956_at	0.4019018	6.0515957	0.28610176	0.047277078	1.2195276
SEC22A	218703_at	0.27051756	6.0441413	0.33006507	0.054609094	1.2395762
MED14	202610_s_at	1.3241192	5.998848	0.30834752	0.051401123	1.2263118
CETN1	207209_at	-0.12337953	5.99096	0.19670671	0.032833923	1.1358333
SHOX2	208443_x_at	0.29599816	5.9882226	0.17490456	0.029208092	1.126643
ZFP64	218968_s_at	0.15487224	5.9797807	0.20519595	0.034314964	1.1564603
C11orf24	52164_at	-0.6755164	5.977472	0.25075063	0.04194928	1.1587437
ACACB	214584_x_at	-0.51553005	5.9698567	0.26251233	0.04397297	1.1838033
SNX3	200067_x_at	0.61864305	5.8936825	0.2869867	0.048693955	1.1491308
SPA17	205406_s_at	-0.6107301	5.8487926	0.15721337	0.02687963	1.1128808
NR0B1	206644_at	1.297714	5.8422666	0.1525614	0.026113393	1.1039187
ANKRD12	216550_x_at	0.117988594	5.803189	0.3413775	0.058825847	1.2400688
MYST2	200049_at	-0.16360782	5.710427	0.2714969	0.047544062	1.1628419
CUGBP2	202158_s_at	-0.90555644	5.6692576	0.26924247	0.047491662	1.1660706

Table B. 2Significantly 3’UTR lengthened genes in ER(+) patients from GSE2034 compared to normal breast samples.

Gene name	Probeset ID	Expected score	Observed score	Numerator	Denominator	Fold change (unlogged)
TNPO1	209225_x_at	0.46078923	-18.105446	-0.90323246	0.049887337	0.52796805
TNPO1	209225_x_at	0.45954993	-17.676332	-0.9324622	0.052752018	0.50416213
SLC46A3	214719_at	-0.4911221	-15.762062	-0.9131478	0.05793327	0.5106096
WIPI2	214699_x_at	0.6935498	-15.0025015	-0.8183749	0.05454923	0.54177785
ERG	211626_x_at	0.17835727	-14.503462	-0.96083415	0.06624861	0.45469356
ERG	211626_x_at	0.17736891	-13.86469	-0.9118577	0.065768346	0.46969014
EIF5B	201024_x_at	0.033359885	-13.564391	-0.4812258	0.03547714	0.6959434
MUT	202960_s_at	0.5852868	-12.696313	-0.5960024	0.046942953	0.62230015
OGT	207564_x_at	1.4249951	-11.439954	-0.80444336	0.07031876	0.53429055
CTBS	218923_at	-1.2777423	-10.986672	-0.43140453	0.03926617	0.73332506
FAM59A	219377_at	-0.11112964	-10.385724	-0.16632476	0.01601475	0.8897145
AMOT	209521_s_at	1.5868655	-10.123422	-0.64404714	0.06361951	0.63283116
OGT	207564_x_at	1.4301554	-9.94297	-0.48965055	0.0492459	0.70081085
SLC25A31	221037_s_at	0.39878312	-9.626065	-0.23564397	0.02447978	0.84148943
PHKB	202739_s_at	-0.25355566	-9.569605	-0.31196645	0.03259972	0.79966235
DOPEY1	40612_at	0.60163933	-9.316416	-0.34329566	0.03684847	0.7700408
CCDC93	219774_at	0.04667492	-9.286576	-0.2650298	0.02853902	0.82333505
ESF1	218859_s_at	0.1284136	-9.086328	-0.3802428	0.0418478	0.7572752
KIF5A	205318_at	-0.5525743	-8.960994	-0.2026903	0.022619177	0.8681284
ZAK	218833_at	0.069119245	-8.927961	-0.572788	0.06415664	0.6720713
MCFD2	212246_at	0.011244561	-8.833683	-0.16482331	0.018658504	0.890154
SF4	209547_s_at	-0.04953985	-8.542827	-0.43669534	0.051118366	0.7242954
COX4NB	218057_x_at	-0.22116618	-8.417612	-0.15101556	0.017940428	0.8991654
AZIN1	201772_at	0.9646151	-8.174138	-0.15388341	0.018825645	0.894401
RPS17	211487_x_at	-0.30009007	-8.150904	-0.17832112	0.021877466	0.8840969
EIF5B	201024_x_at	0.034331013	-7.8552217	-0.5087983	0.06477199	0.67238724
PTN	209465_x_at	0.81632006	-7.827134	-0.2533066	0.032362625	0.8281286
CLUAP1	204576_s_at	-0.2797226	-7.8036275	-0.16000867	0.020504396	0.89607435
B4GALT1	211631_x_at	1.0163137	-7.7691875	-0.21393351	0.02753615	0.85961723
BRD8	210352_at	0.5018454	-7.7516594	-0.19033904	0.02455462	0.8729534
YTHDC2	205836_s_at	0.48386422	-7.674269	-0.28199968	0.03674613	0.8017305
GATA3	209603_at	-0.90730554	-7.6021214	-0.20842266	0.027416382	0.8628329
TRAK1	214924_s_at	0.23604089	-7.50877	-0.1643351	0.021885755	0.8929949
MEF2A	208328_s_at	-0.29099768	-7.4545035	-0.2723928	0.036540706	0.80253226
AK2	205996_s_at	-1.6035434	-7.430312	-0.45841682	0.0616955	0.7012107
ARL4A	205020_s_at	0.6996866	-7.429115	-0.16815245	0.022634251	0.8888941
PRKAR1A	200603_at	-0.14202192	-7.4273043	-0.23656431	0.03185063	0.8206399
LPXN	216250_s_at	-0.69108844	-7.420362	-0.2072217	0.027926091	0.8605394
FEZ2	215000_s_at	0.002144305	-7.411118	-0.106961496	0.014432572	0.9275853
CREBBP	202160_at	-0.2777147	-7.336307	-0.31375313	0.042767175	0.79384863
APC	216933_x_at	0.48274702	-7.333587	-0.41329575	0.05635656	0.7352935
TIAL1	202405_at	-0.7884015	-7.3335543	-0.17156711	0.02339481	0.889265
ATP6V1C1	202872_at	0.9701419	-7.0223384	-0.2531943	0.036055554	0.8232901
HFE	211332_x_at	0.5525901	-6.9532223	-0.12588388	0.018104395	0.91462994
SLC6A16	219820_at	-0.03500556	-6.928284	-0.16324951	0.023562761	0.89282286
PPP1R11	201500_s_at	0.5575347	-6.862861	-0.11354558	0.016544934	0.9239979
SEC11A	201290_at	-0.29592064	-6.7868958	-0.13067186	0.019253554	0.910526
IPO13	203546_at	-1.4712919	-6.745475	-0.2127963	0.031546526	0.8452857
PTPN11	212610_at	-0.51199067	-6.7433653	-0.14158213	0.020995768	0.9071333
SUPT7L	201837_s_at	-0.001385372	-6.738581	-0.14305055	0.021228585	0.9064449
STAT6	201331_s_at	-0.5551185	-6.7378316	-0.33494455	0.04971103	0.77077824
KDM5B	201548_s_at	-0.98916453	-6.70826	-0.26082468	0.038881123	0.8141915
CDKN1B	209112_at	-0.5923835	-6.6245117	-0.10736015	0.0162065	0.92767495
MAP3K7IP2	212184_s_at	0.6529613	-6.610662	-0.15225255	0.023031363	0.8916863
MAP3K7IP2	212184_s_at	0.65147567	-6.610662	-0.15225255	0.023031363	0.8916863
EBAG9	204274_at	0.9769356	-6.5405884	-0.17348556	0.026524458	0.8809718
TSHR	210055_at	-0.39507926	-6.5041447	-0.1454696	0.022365678	0.90402955
NRG1	208230_s_at	0.88729453	-6.4933596	-0.15794145	0.024323532	0.8955876
COG5	203630_s_at	0.7886968	-6.485954	-0.15434164	0.02379629	0.89720863
MEF2A	208328_s_at	-0.28995076	-6.4453692	-0.22986561	0.035663683	0.83155584
DKC1	216212_s_at	1.9830532	-6.433341	-0.20526257	0.03190606	0.859966
EBI3	219424_at	-0.07107137	-6.4160023	-0.1660773	0.025884857	0.8900807

Table B.2 (continued)

Gene name	Probeset ID	Expected score	Observed score	Numerator	Denominator	Fold change (unlogged)
UBAP2	221839_s_at	1.0240314	-6.399785	-0.13519794	0.021125387	0.91101944
SSX2IP	203019_x_at	-1.2686129	-6.3026257	-0.1797665	0.028522478	0.8804776
TRAPPC3	203511_s_at	-1.5547765	-6.2908354	-0.14897567	0.023681382	0.8999684
ATP6V1C1	202872_at	0.9669238	-6.275585	-0.26995167	0.043016177	0.81132627
TFDP1	204147_s_at	-0.45129114	-6.2740364	-0.18813786	0.029986734	0.87605256
MAPK6	207121_s_at	-0.33925897	-6.264071	-0.13207465	0.021084476	0.9146859
PTK2	207821_s_at	0.98908687	-6.256107	-0.16129768	0.025782436	0.8916596
EGFL6	219454_at	1.2536263	-6.2429366	-0.1537441	0.024626888	0.8973278
CXCL3	207850_at	0.36743	-6.23566	-0.1404111	0.022517439	0.9093991
PLSCR1	202430_s_at	0.29195005	-6.2346945	-0.2976034	0.047733437	0.8025127
SRPX2	205499_at	1.4576367	-6.2168756	-0.13587284	0.021855487	0.9110368
GGNBP2	218079_s_at	-0.18811125	-6.213671	-0.14668459	0.023606751	0.9023478
QDPR	209123_at	0.3419478	-6.2005725	-0.1519603	0.024507463	0.8992791
SPCS2	201240_s_at	-0.6628756	-6.179752	-0.11808455	0.0191083	0.91859806
HIPK1	212293_at	-1.1754303	-6.17884	-0.15070896	0.02439114	0.8984788
PKP4	201928_at	0.0588165	-6.1547856	-0.15410604	0.025038408	0.8967266
SERF2	217756_x_at	-0.35366198	-6.145681	-0.09904216	0.016115734	0.93218136
RBM15	219286_s_at	-1.1979829	-6.1211243	-0.13319151	0.021759322	0.9116592
SLC47A1	219525_at	-0.19884372	-6.050712	-0.15269256	0.025235469	0.8945783
HSP90AA1	211969_at	-0.37815994	-6.0418344	-0.122014344	0.020194918	0.918558
COL15A1	203477_at	1.1116642	-6.032374	-0.14682534	0.024339562	0.9030873
ZMYM1	220206_at	-1.5757039	-6.0318527	-0.14392546	0.023860903	0.90102875
CD58	205173_x_at	-1.1687167	-6.001715	-0.11721629	0.019530466	0.91953534
YIPF2	221939_at	-0.06544364	-6.0005307	-0.105901405	0.017648673	0.9262192
ITSN2	209907_s_at	-0.009334193	-5.996298	-0.13031797	0.021733072	0.9154365
DLGAP2	210227_at	0.84208244	-5.98293	-0.1494219	0.024974702	0.8998387
DUT	208956_x_at	-0.35064635	-5.981703	-0.17092028	0.02857385	0.8786377
DOPEY1	213271_s_at	0.6028955	-5.952086	-0.11640925	0.019557724	0.922969
CCND1	208712_at	-0.67383236	-5.9475083	-0.1786204	0.030032812	0.865458
CHSY1	203044_at	-0.2888774	-5.9206085	-0.13055602	0.022051115	0.91445786
KDM6A	203991_s_at	1.3479817	-5.916387	-0.1661983	0.028091181	0.88339806
GADD45B	207574_s_at	-0.07479551	-5.9151826	-0.1501038	0.025376022	0.9011336
GTPBP1	205275_at	0.2092342	-5.9099092	-0.16374257	0.027706444	0.890558
FAM13A	202973_x_at	0.38319775	-5.881051	-0.13109946	0.022291843	0.9149358
EIF5	208708_x_at	-0.37598342	-5.867073	-0.13056879	0.022254502	0.913165
SATB1	203408_s_at	0.22932987	-5.851749	-0.108919345	0.018613126	0.9281603
KBTBD11	204301_at	0.84400415	-5.817691	-0.12254364	0.021063965	0.9204104
ZNF143	221873_at	-0.74013704	-5.809289	-0.18991154	0.032691013	0.86972296
TPT1	212869_x_at	-0.4775202	-5.8026867	-0.14991672	0.025835743	0.8957349
TBXAS1	208130_s_at	0.8251567	-5.774535	-0.14368452	0.02488244	0.902657
TBXAS1	208130_s_at	0.823103	-5.774535	-0.14368452	0.02488244	0.902657
GRK6	211543_s_at	0.52459556	-5.772625	-0.13000214	0.022520455	0.9125358
CLUAP1	204576_s_at	-0.27875853	-5.7558155	-0.12072681	0.020974753	0.920866
ACTR6	218395_at	-0.5212515	-5.737386	-0.14133768	0.024634507	0.9061336
NPAS2	39549_at	0.037019245	-5.7297926	-0.14900471	0.026005255	0.8989943
COX7A2L	201256_at	0.007897385	-5.7149224	-0.095072076	0.016635759	0.93574053
MMP7	204259_at	-0.63751173	-5.701715	-0.5419816	0.095055886	0.68226415
SFRS14	213505_s_at	-0.05251337	-5.6997685	-0.1023908	0.017964028	0.9284584
SFRS14	213505_s_at	-0.051570974	-5.6997685	-0.1023908	0.017964028	0.9284584
MBTPS1	201620_at	-0.2262711	-5.679056	-0.09860098	0.017362213	0.9340688
UBE2I	208760_at	-0.28285527	-5.6775618	-0.14076495	0.024793204	0.9073674
GSTM1	204550_x_at	-1.2048852	-5.676961	-0.11616771	0.02046301	0.9224915

Table B.3Significantly 3’UTR shortened genes in ER(+) patients from GSE7390 compared to normal breast samples.

Gene name	Probeset ID	Poly(A) site ID	Expected score	Observed score	Numerator	Denominator	Fold change (unlogged)
HNRNPA1	214280_x_at	Hs.546261.1.24	-0.5474836	15.283277	0.965465	0.063171	1.9274905
HNRNPA1	214280_x_at	Hs.546261.1.23	-0.5487645	15.283277	0.965465	0.063171	1.9274905
HNRNPA1	214280_x_at	Hs.546261.1.20	-0.549977	15.283277	0.965465	0.063171	1.9274905
G3BP2	208841_s_at	Hs.303676.1.3	0.35860148	14.729386	0.641015	0.043519	1.5513788
RPL13	212933_x_at	Hs.410817.1.11	-0.20844637	14.71295	0.895472	0.060863	1.8627723
WBP5	217975_at	Hs.533287.1.3	1.482113	13.518285	0.744868	0.055101	1.6708009
GLRX3	214205_x_at	Hs.42644.1.14	-0.74183613	13.477963	0.770273	0.057151	1.6666461
GLRX3	214205_x_at	Hs.42644.1.16	-0.74006283	13.477963	0.770273	0.057151	1.6666461
GLRX3	214205_x_at	Hs.42644.1.17	-0.7384402	13.477963	0.770273	0.057151	1.6666461
IQCK	215131_at	Hs.460217.1.5	-0.25728709	13.126949	1.240696	0.094515	1.8900921
APP	200602_at	Hs.434980.1.2	0.16212215	12.705506	0.543549	0.042781	1.4237798
USP9X	201100_s_at	Hs.77578.1.84	1.3256184	12.11886	0.518964	0.042823	1.4099373
CNBP	206158_s_at	Hs.518249.1.4	0.26460963	11.773301	0.689798	0.05859	1.5847447
GOLPH3	217803_at	Hs.408909.1.2	0.4219005	11.279692	0.448643	0.039774	1.3500788
MAPK13	210059_s_at	Hs.178695.1.25	0.55835325	11.177422	0.411913	0.036852	1.3218129
RPL19	200029_at	Hs.381061.1.6	-0.17760162	10.815645	0.567822	0.0525	1.515099
RPL31	221593_s_at	Hs.469473.1.5	0.0379352	10.733009	0.426562	0.039743	1.3270297
LMAN2	200805_at	Hs.75864.1.4	0.50701946	10.673798	0.479423	0.044916	1.4008951
ADSL	210250_x_at	Hs.75527.1.19	0.20635948	10.371504	0.537513	0.051826	1.4057109
MRPL3	208787_at	Hs.205163.1.4	0.27232802	10.330479	0.394967	0.038233	1.3216634
CLU	208792_s_at	Hs.436657.1.10	0.8654564	10.025716	0.3783	0.037733	1.2846111
METTL9	217868_s_at	Hs.279583.1.15	-0.25467512	9.947806	0.667794	0.06713	1.5107725
DDRGK1	218159_at	Hs.471975.1.3	0.11901656	9.908653	0.421152	0.042503	1.3359392
LXN	218729_at	Hs.478067.1.6	0.29012737	9.836243	0.392297	0.039883	1.2912375
HNRNPA1	200016_x_at	Hs.546261.1.20	-0.551378	9.664443	0.380054	0.039325	1.2874571
SPP1	209875_s_at	Hs.313.1.13	0.36976305	9.63958	0.424017	0.043987	1.3539408
SNX3	213545_x_at	Hs.12102.1.3	0.6109966	9.544586	0.602253	0.063099	1.4202209
SNX3	200067_x_at	Hs.12102.1.3	0.6098383	9.487029	0.555337	0.058536	1.387254
CKS2	204170_s_at	Hs.83758.1.6	1.0700601	9.3363695	0.26712	0.028611	1.2012334
PRSS21	220051_at	Hs.72026.1.3	-0.27319172	9.254785	0.383879	0.041479	1.2862176
MED14	202610_s_at	Hs.407604.1.7	1.3156382	9.207467	0.488956	0.053104	1.3807857
VASP	202205_at	Hs.515469.1.28	-0.03231617	9.144816	0.421884	0.046134	1.3200918
CNBP	206158_s_at	Hs.518249.1.3	0.26367977	9.129408	0.48726	0.053373	1.3786234
GLUL	215001_s_at	Hs.518525.1.4	-0.98703086	8.785584	0.393426	0.044781	1.3491642
ATP6V1D	208899_x_at	Hs.272630.1.17	-0.39417022	8.678844	0.396495	0.045685	1.3206781
IFT57	218100_s_at	Hs.412196.1.6	0.24841578	8.550838	0.310804	0.036348	1.2267473
SPIN1	217813_s_at	Hs.146804.1.15	1.0637699	8.547181	0.313971	0.036734	1.2470126
DDX28	40255_at	Hs.458313.1.2	-0.22841386	8.52653	0.256349	0.030065	1.1824886
FMR1	215245_x_at	Hs.103183.1.36	1.8851423	8.373392	0.314381	0.037545	1.2417293
WARS	200628_s_at	Hs.497599.1.4	-0.37115327	8.35754	0.345452	0.041334	1.2715523
DCK	203302_at	Hs.709.1.27	0.3514836	8.259256	0.416944	0.050482	1.3348552
SMARCA4	208794_s_at	Hs.327527.1.81	-0.06057165	8.206937	0.324594	0.039551	1.2485884
PPFIA1	202066_at	Hs.530749.1.60	-0.64876103	8.190737	0.286623	0.034994	1.2146345
TPT1	214327_x_at	Hs.374596.1.15	-0.46137166	8.066726	0.545821	0.067663	1.5071801
DNPEP	38703_at	Hs.258551.1.5	0.10169233	8.046389	0.356429	0.044297	1.2614763
TIMM9	218316_at	Hs.440525.1.4	-0.4072631	7.9888597	0.392065	0.049076	1.2951349
RLF	204243_at	Hs.205627.1.5	-1.4787644	7.7894983	0.234587	0.030116	1.1691171
CNBP	206158_s_at	Hs.518249.1.5	0.26561546	7.760622	0.292728	0.03772	1.2243514
TIMM9	218316_at	Hs.440525.1.3	-0.40835676	7.710879	0.338978	0.043961	1.2408638
HMOX2	218121_at	Hs.284279.1.14	-0.26843545	7.646531	0.386293	0.050519	1.2889805
SAG	206671_at	Hs.32721.1.21	0.10980464	7.4971094	0.232246	0.030978	1.1675955
SLC25A17	211754_s_at	Hs.474938.1.9	0.20802753	7.4955516	0.305579	0.040768	1.2265078
YME1L1	216304_x_at	Hs.499145.1.10	-0.8453865	7.4704103	0.508516	0.068071	1.36344
HIST1H1C	209398_at	Hs.7644.1.2	0.53186595	7.44674	0.238097	0.031973	1.1754853
STAM	203544_s_at	Hs.441498.1.40	-0.85768837	7.4331284	0.299631	0.04031	1.2094488
ANXA6	200982_s_at	Hs.412117.1.7	0.49992004	7.430307	0.248605	0.033458	1.1849054
PVR	214444_s_at	Hs.171844.1.18	-0.03602296	7.424177	0.345998	0.046604	1.2300892
ZFP64	218968_s_at	Hs.473082.1.19	0.14942865	7.3761024	0.248094	0.033635	1.1880822
NIPBL	212483_at	Hs.481927.1.71	0.4285732	7.3222885	0.321961	0.04397	1.2340492
POFUT1	212349_at	Hs.472409.1.31	0.13477793	7.261289	0.242279	0.033366	1.1737554
HSPA4L	205543_at	Hs.135554.1.18	0.3868922	7.2205977	0.304999	0.04224	1.218558
C1orf50	219406_at	Hs.148845.1.9	-1.4523133	7.197904	0.23398	0.032507	1.167575

Table B.3 (continued)

Gene name	Probeset ID	Poly(A) site ID	Expected score	Observed score	Numerator	Denominator	Fold change (unlogged)
COIL	203654_s_at	Hs.532795.1.6	-0.14778629	7.14301	0.307614	0.043065	1.2343034
CCNG2	202769_at	Hs.13291.1.25	0.36697868	7.0888762	0.420334	0.059295	1.3444141
BGN	201262_s_at	Hs.821.1.15	1.9842215	7.053238	0.388488	0.055079	1.3610016

Table B.4Significantly 3’UTR lenghened genes in ER(+) patients from GSE7390 compared to normal breast samples.

Gene name	Probeset ID	Poly(A) site ID	Expected score	Observed score	Numerator	Denominator	Fold change (unlogged)
PABPC3	208113_x_at	Hs.458280.1.1	-0.47981	-22.9631	-1.77	0.07708	0.27945822
OGT	207564_x_at	Hs.405410.1.62	1.412618	-17.9894	-1.3312	0.073999	0.36659104
WIPI2	214699_x_at	Hs.122363.1.30	0.684733	-17.2008	-0.56049	0.032585	0.6717721
GNS	203676_at	Hs.334534.1.5	-0.53145	-15.1702	-0.54735	0.036081	0.67505664
SLC46A3	214719_at	Hs.117167.1.1	-0.47643	-14.4142	-0.90964	0.063107	0.5088842
WIPI2	214699_x_at	Hs.122363.1.29	0.683073	-14.0892	-0.85304	0.060546	0.52480537
VAMP3	211749_s_at	Hs.66708.1.11	-1.96016	-13.3625	-0.37786	0.028277	0.76540786
CHST5	64900_at	Hs.156784.1.13	-0.22305	-13.1427	-0.69279	0.052713	0.5934663
ESF1	218859_s_at	Hs.369284.1.19	0.125591	-12.1923	-0.49912	0.040937	0.6939443
ARF6	203312_x_at	Hs.525330.1.12	-0.42138	-12.1603	-0.7191	0.059135	0.5878332
ERG	211626_x_at	Hs.473819.1.5	0.17172	-11.9198	-0.93293	0.078267	0.46288368
AMOT	209521_s_at	Hs.528051.1.1	1.591824	-11.7865	-0.75087	0.063705	0.5800485
EIF5B	201024_x_at	Hs.158688.1.13	0.034351	-11.2905	-0.81667	0.072333	0.5414848
ADRBK1	38447_at	Hs.83636.1.51	-0.65421	-11.2507	-0.54348	0.048306	0.66532904
SRP72	208802_at	Hs.237825.1.39	0.347536	-11.2066	-0.62643	0.055898	0.6210181
MEF2A	208328_s_at	Hs.268675.1.28	-0.28276	-11.0302	-0.46514	0.042169	0.7013186
KIF5A	205318_at	Hs.151219.1.36	-0.53852	-10.8285	-0.27945	0.025807	0.8218636
ERG	211626_x_at	Hs.473819.1.6	0.172602	-10.816	-0.85761	0.079291	0.49155885
ZMYM1	220206_at	Hs.471243.1.13	-1.54159	-10.7618	-0.27351	0.025415	0.8219444
EYA3	206379_at	Hs.185774.1.3	-1.62864	-10.7206	-0.333	0.031062	0.7861263
TSPYL1	221493_at	Hs.458358.1.4	0.617603	-10.7184	-0.19827	0.018498	0.869482
TSPYL1	221493_at	Hs.458358.1.3	0.616406	-10.7184	-0.19827	0.018498	0.869482
SET	213047_x_at	Hs.436687.1.19	1.185643	-10.6991	-0.60879	0.056902	0.622509
PLSCR1	202430_s_at	Hs.130759.1.13	0.283241	-10.6632	-0.51244	0.048057	0.68431926
MUT	202960_s_at	Hs.485527.1.6	0.573772	-10.6445	-0.59413	0.055816	0.6237134
BBS9	37547_at	Hs.372360.1.39	0.70519	-10.2754	-0.349	0.033965	0.77229625
EYA3	206379_at	Hs.185774.1.6	-1.61733	-10.27	-0.32448	0.031594	0.78957665
MKS1	218630_at	Hs.408843.1.8	-0.14511	-10.2467	-0.45386	0.044293	0.70913875
STK3	211078_s_at	Hs.492333.1.29	0.941575	-10.1324	-0.43309	0.042743	0.7200486
CYP4A11	207407_x_at	Hs.1645.1.4	-1.36699	-10.1144	-0.32309	0.031943	0.7947006
DOPEY1	40612_at	Hs.520246.1.28	0.591698	-10.0495	-0.41068	0.040865	0.73328274
AK2	205996_s_at	Hs.470907.1.3	-1.57182	-9.99116	-0.63865	0.063922	0.60994035
LMO7	202674_s_at	Hs.207631.1.53	-0.45253	-9.89989	-0.6423	0.06488	0.6131617
SPTBN1	200671_s_at	Hs.503178.1.83	0.015809	-9.66539	-0.42238	0.043701	0.7260106
SAFB	201747_s_at	Hs.23978.1.24	-0.06488	-9.5996	-0.38934	0.040558	0.74386686
MEF2A	208328_s_at	Hs.268675.1.29	-0.28198	-9.53381	-0.38611	0.040499	0.7445498
B4GALT1	211631_x_at	Hs.272011.1.15	1.0048	-9.37485	-0.25217	0.026899	0.8342673
SLC48A1	48106_at	Hs.438867.1.13	-0.55839	-9.33035	-0.4518	0.048422	0.7148561
ATP6V1C1	202872_at	Hs.86905.1.22	0.959287	-9.32821	-0.36903	0.039561	0.75757575
SF4	209547_s_at	Hs.515274.1.3	-0.0446	-9.25215	-0.46735	0.050512	0.7025299
COL5A3	52255_s_at	Hs.235368.1.2	-0.06399	-9.25132	-0.26352	0.028484	0.82893914
TFDP1	204147_s_at	Hs.79353.1.21	-0.43807	-9.22294	-0.26232	0.028442	0.82896316
OGT	207564_x_at	Hs.405410.1.63	1.418391	-9.13814	-0.47028	0.051463	0.7045664
DCAF6	217908_s_at	Hs.435741.1.51	-1.04924	-8.7692	-0.1797	0.020492	0.8789563
HBB	209116_x_at	Hs.523443.1.8	-0.72252	-8.69856	-0.20389	0.023439	0.86596084
SIGMAR1	214484_s_at	Hs.522087.1.7	1.021321	-8.69525	-0.21773	0.02504	0.85299915
SLC45A2	220245_at	Hs.278962.1.3	0.425494	-8.63762	-0.29395	0.034032	0.80520815
HFE	211332_x_at	Hs.233325.1.9	0.536351	-8.5451	-0.18783	0.021981	0.87346727
CAPN3	210944_s_at	Hs.143261.1.34	-0.3462	-8.54288	-0.1541	0.018038	0.89857715
HSPA4	211016_x_at	Hs.90093.1.1	0.481216	-8.51604	-0.28917	0.033956	0.807177
TMBIM4	219206_x_at	Hs.505934.1.8	-0.52899	-8.50687	-0.58921	0.069263	0.60374
RABGEF1	218310_at	Hs.530053.1.33	0.721857	-8.46215	-0.5045	0.059618	0.66714895
TTC4	46167_at	Hs.412482.1.25	-1.31109	-8.36111	-0.17773	0.021257	0.8807351
YTHDC2	205836_s_at	Hs.231942.1.36	0.471029	-8.24471	-0.33105	0.040153	0.7725908
BUB1	209642_at	Hs.469649.1.11	0.041338	-8.24128	-0.21136	0.025646	0.86157775
SERPINB3	209720_s_at	Hs.227948.1.6	-0.08986	-8.22858	-0.21122	0.025669	0.86321026
HSPBAP1	219284_at	Hs.29169.1.2	0.259985	-8.18616	-0.17636	0.021543	0.8830752
IL1R2	211372_s_at	Hs.25333.1.10	0.038773	-8.17706	-0.26094	0.031911	0.8270676
NPAS2	39549_at	Hs.156832.1.43	0.037035	-8.14922	-0.21004	0.025775	0.8589163
OXR1	218197_s_at	Hs.148778.1.30	0.961397	-8.14913	-0.15905	0.019517	0.89353395
PA2G4	214794_at	Hs.524498.1.17	-0.54521	-8.05848	-0.46964	0.058279	0.6831055
C17orf42	219146_at	Hs.317340.1.7	-0.18397	-8.03975	-0.18419	0.02291	0.8790497

Table B.4 (continued)

Gene name	Probeset ID	Poly(A) site ID	Expected score	Observed score	Numerator	Denominator	Fold change (unlogged)
CCDC93	219774_at	Hs.107845.1.3	0.04567	-8.002	-0.24865	0.031074	0.8307782
C17orf85	218896_s_at	Hs.120963.1.2	-0.20657	-7.99752	-0.31491	0.039376	0.7942325
GABRA3	207210_at	Hs.123024.1.5	1.916286	-7.96061	-0.22694	0.028508	0.8466355
SRD5A1	211056_s_at	Hs.552.1.20	0.419853	-7.95988	-0.30926	0.038852	0.79941297
SRD5A1	211056_s_at	Hs.552.1.19	0.418877	-7.95988	-0.30926	0.038852	0.79941297
TSHR	210055_at	Hs.160411.1.17	-0.38529	-7.89616	-0.18064	0.022877	0.8799382
PHKB	202739_s_at	Hs.78060.1.71	-0.24436	-7.87534	-0.27312	0.03468	0.82067275
ZNF764	222120_at	Hs.132227.1.3	-0.24622	-7.85905	-0.39411	0.050147	0.73676383
ABHD6	45288_at	Hs.476454.1.20	0.242837	-7.84098	-0.19318	0.024638	0.87347114
NDUFS7	211752_s_at	Hs.211914.1.16	-0.07913	-7.84031	-0.18153	0.023154	0.87538093
SOX10	209842_at	Hs.376984.1.7	0.201782	-7.8383	-0.1662	0.021204	0.89070714
MRPL35	218890_x_at	Hs.433439.1.6	0.027189	-7.83619	-0.17094	0.021814	0.8887486
STAT6	201331_s_at	Hs.524518.1.4	-0.54001	-7.82279	-0.22309	0.028518	0.849438
LARP1	212193_s_at	Hs.292078.1.54	0.502442	-7.79877	-0.24764	0.031754	0.8334033
KDM6A	203991_s_at	Hs.522616.1.43	1.334203	-7.79424	-0.25383	0.032566	0.8321544
AZIN1	201772_at	Hs.459106.1.14	0.955	-7.76924	-0.13494	0.017368	0.90926313
COL15A1	203477_at	Hs.409034.1.43	1.105354	-7.75194	-0.19617	0.025305	0.8700244
PCYT1A	204209_at	Hs.435767.1.5	0.305997	-7.74904	-0.17514	0.022602	0.8844559
SLC5A1	206628_at	Hs.1964.1.12	0.198994	-7.73309	-0.23744	0.030705	0.8374668
POLE	216026_s_at	Hs.524871.1.3	-0.48321	-7.72556	-0.2216	0.028684	0.8518298
SUPT7L	201837_s_at	Hs.6232.1.10	0.002404	-7.72082	-0.1458	0.018884	0.9036612
GAS7	202192_s_at	Hs.462214.1.3	-0.20099	-7.70053	-0.16722	0.021715	0.88662624
LOC441453	217551_at	Hs.523504.1.3	-0.70568	-7.69857	-0.17896	0.023246	0.87998426
DSC2	204751_x_at	Hs.95612.1.5	-0.10769	-7.64137	-0.21709	0.02841	0.8577078
CSK	202329_at	Hs.77793.1.24	-0.30118	-7.61093	-0.18184	0.023891	0.8796827
RFC5	203209_at	Hs.506989.1.23	-0.49426	-7.57119	-0.1573	0.020777	0.895878
SLC25A31	221037_s_at	Hs.149030.1.11	0.385952	-7.55374	-0.21498	0.028461	0.8537564
TPI1	210050_at	Hs.524219.1.15	-0.58015	-7.52426	-0.43598	0.057944	0.70364296
SRRT	201679_at	Hs.111801.1.18	0.763314	-7.52385	-0.26748	0.035551	0.82462054
CHD4	201184_s_at	Hs.162233.1.4	-0.58136	-7.49321	-0.19691	0.026278	0.8699236
WSB2	201760_s_at	Hs.506985.1.9	-0.49216	-7.45727	-0.13883	0.018617	0.90615374
HINFP	206495_s_at	Hs.504091.1.37	-0.59561	-7.44234	-0.14768	0.019843	0.9029614
TIAL1	202405_at	Hs.501203.1.2	-0.76238	-7.438	-0.15127	0.020338	0.89964294
ECHDC1	219974_x_at	Hs.486410.1.5	0.624277	-7.43264	-0.28818	0.038773	0.80548143
TOR1A	202349_at	Hs.534312.1.3	1.199056	-7.42471	-0.13402	0.01805	0.91053605
HISPPD1	203253_s_at	Hs.212046.1.54	0.465343	-7.42322	-0.14731	0.019844	0.9024756
HISPPD1	203253_s_at	Hs.212046.1.55	0.466299	-7.42322	-0.14731	0.019844	0.9024756
FBXO34	218539_at	Hs.525348.1.8	-0.40959	-7.41997	-0.11548	0.015563	0.92236185
NIF3L1	218133_s_at	Hs.145284.1.11	0.08156	-7.40834	-0.12288	0.016587	0.9170253
CTBP2	201220_x_at	Hs.501345.1.6	-0.74645	-7.3739	-0.17157	0.023267	0.88560325
MRPL28	204599_s_at	Hs.513230.1.3	-0.27624	-7.3734	-0.14176	0.019226	0.90566254
DDX28	203785_s_at	Hs.458313.1.2	-0.22749	-7.36349	-0.15135	0.020554	0.9002905
LPO	210682_at	Hs.234742.1.9	-0.14419	-7.33325	-0.24561	0.033493	0.8397367
NUP98	210793_s_at	Hs.524750.1.52	-0.72575	-7.3236	-0.17364	0.02371	0.88503975
BCL2	203685_at	Hs.150749.1.2	-0.09239	-7.30791	-0.12319	0.016856	0.91688675
CHP	207993_s_at	Hs.406234.1.6	-0.34724	-7.27595	-0.20211	0.027778	0.8637066
AMOT	209521_s_at	Hs.528051.1.2	1.60552	-7.26226	-0.45766	0.063018	0.70787066
EHD1	222221_x_at	Hs.523774.1.2	-0.66275	-7.24773	-0.15192	0.02096	0.89649016
EHD1	222221_x_at	Hs.523774.1.7	-0.66003	-7.24773	-0.15192	0.02096	0.89649016
EHD1	222221_x_at	Hs.523774.1.8	-0.65849	-7.24773	-0.15192	0.02096	0.89649016
EHD1	222221_x_at	Hs.523774.1.3	-0.66142	-7.24773	-0.15192	0.02096	0.89649016
CYCS	208905_at	Hs.437060.1.5	0.700034	-7.24468	-0.32802	0.045278	0.77130806
SUMO3	200740_s_at	Hs.474005.1.4	0.1841	-7.24266	-0.11985	0.016548	0.91945523
ETAA1	219216_at	Hs.353022.1.16	0.018232	-7.24189	-0.3649	0.050387	0.74694216
OAT	201599_at	Hs.523332.1.4	-0.74951	-7.23878	-0.17373	0.024	0.8867666
ANKRD12	216550_x_at	Hs.464585.1.21	-0.11363	-7.22982	-0.42622	0.058953	0.72194123
MAP7	202890_at	Hs.486548.1.2	0.640076	-7.22449	-0.13278	0.018379	0.9110237
EBI3	219424_at	Hs.501452.1.12	-0.06775	-7.15008	-0.17876	0.025002	0.8800343
WDHD1	204727_at	Hs.385998.1.1	-0.41172	-7.14933	-0.19073	0.026679	0.8688579
GLRA3	207928_s_at	Hs.413099.1.4	0.410532	-7.13113	-0.17059	0.023921	0.88541716
FBN2	215717_s_at	Hs.519294.1.49	0.478194	-7.10973	-0.17612	0.024772	0.8822563
REG1A	209752_at	Hs.49407.1.5	0.023487	-7.10934	-0.1746	0.024559	0.88450956
MTSS1L	213978_at	Hs.432387.1.3	-0.2249	-7.09869	-0.16559	0.023327	0.8892862
CCND1	208712_at	Hs.523852.1.22	-0.65146	-7.09582	-0.26724	0.037662	0.8161082

Table B.4 (continued)

Gene name	Probeset ID	Poly(A) site ID	Expected score	Observed score	Numerator	Denominator	Fold change (unlogged)
ATP7A	205198_s_at	Hs.496414.1.21	1.445468	-7.0921	-0.27796	0.039193	0.8105867
PTP4A1	200730_s_at	Hs.227777.1.13	0.580208	-7.08292	-0.13019	0.018381	0.91118026
ADM	202912_at	Hs.441047.1.9	-0.71153	-7.08074	-0.13763	0.019437	0.907549
ADM	202912_at	Hs.441047.1.8	-0.71276	-7.08074	-0.13763	0.019437	0.907549
C21orf7	221211_s_at	Hs.222802.1.15	0.163712	-7.05473	-0.13092	0.018558	0.9107128
PEBP1	211941_s_at	Hs.433863.1.5	-0.49109	-7.04149	-0.14927	0.021198	0.90199757
ZAK	218833_at	Hs.444451.1.34	0.065268	-7.04112	-0.42775	0.060751	0.72764003
SFRS18	212177_at	Hs.520287.1.23	0.604665	-6.94638	-0.12601	0.01814	0.91437536
DKC1	216212_s_at	Hs.4747.1.24	2.031887	-6.90973	-0.23144	0.033495	0.8424745
WDR26	218107_at	Hs.497873.1.14	-0.92031	-6.84512	-0.15478	0.022612	0.89706737
SCARB2	201647_s_at	Hs.349656.1.9	0.362117	-6.83823	-0.38602	0.05645	0.73126984
ZNF207	200829_x_at	Hs.500775.1.32	-0.18305	-6.81503	-0.2165	0.031768	0.857782
SMARCA2	206544_x_at	Hs.298990.1.54	0.987793	-6.79825	-0.17412	0.025612	0.8821816

Table B.5Significantly 3’UTR shortened genes in ERα-infected MDA-MB-231 cells compared to control.

Gene name	Probeset ID	Poly A site ID	Expected score	Observed score	Numerator	Denominator	Fold change (unlogged)
WNT16	224022_x_at	Hs.272375.1.3	1.01731	26.27276	1.822064	0.06935184	3.525134
SLC25A42	242274_x_at	Hs.303669.1.6	-0.09608	15.83061	1.145703	0.0723726	2.2171903

Table B.6Significantly 3’UTR shortened genes in ERE-binding defective ERα-infected MDA-MB-231 cells compared to ERα-infected MDA-MB-231 cells.

Gene name	Probeset ID	Poly A site ID	Expected score	Observed score	Numerator	Denominator	Fold change (unlogged)
WNT16	224022_x_at	Hs.272375.1.3	0.932211	15.7172	1.797023	0.114335	3.4408667
WDR60	219251_s_at	Hs.389945.1.28	0.968162	13.33901	0.63915	0.047916	1.5568092
CD44	204490_s_at	Hs.502328.1.28	-0.82894	12.50567	0.733537	0.058656	1.6633995
CD44	204490_s_at	Hs.502328.1.13	-0.83057	12.50567	0.733537	0.058656	1.6633995
CD44	204490_s_at	Hs.502328.1.11	-0.83203	12.50567	0.733537	0.058656	1.6633995
LOC440131	233804_at	Hs.132371.1.1	-0.59381	11.48993	0.653921	0.056913	1.5738183
ST3GAL2	229336_at	Hs.368611.1.3	-0.29968	11.14743	0.941434	0.084453	1.9181896
MAP7	202890_at	Hs.486548.1.2	0.770252	11.14631	0.642443	0.057637	1.5590887
SRP72	208801_at	Hs.237825.1.35	0.40386	11.10564	0.478901	0.043122	1.3934757
SLC25A42	242274_x_at	Hs.303669.1.6	-0.08883	10.94707	0.867113	0.07921	1.829169
FLJ31485	243351_at	Hs.288262.1.2	-0.6155	10.64791	0.762006	0.071564	1.6972654
SEMA3C	203788_s_at	Hs.269109.1.7	0.863068	9.986657	0.648018	0.064888	1.564895
BRD8	210352_at	Hs.519337.1.3	0.58305	9.877386	0.839934	0.085036	1.7854321
ASCC1	219336_s_at	Hs.500007.1.3	-0.96816	9.447157	0.40888	0.043281	1.3282617
TMEM38B	222735_at	Hs.411925.1.24	1.297696	9.10306	0.425222	0.046712	1.3437139
RPS25	200091_s_at	Hs.512676.1.15	-0.74271	9.077871	0.43495	0.047913	1.3520378
RPS25	200091_s_at	Hs.512676.1.10	-0.74433	9.077871	0.43495	0.047913	1.3520378
OGT	207564_x_at	Hs.405410.1.63	1.619891	8.928095	0.846037	0.094761	1.8014696
FRMD6	225481_at	Hs.434914.1.39	-0.50327	8.704706	0.587537	0.067496	1.5023122
UBXN7	212840_at	Hs.518524.1.3	0.360751	8.364058	0.566687	0.067753	1.4840872
BIRC6	233093_s_at	Hs.150107.1.118	-0.00715	8.142216	0.375975	0.046176	1.2976894
KLC1	212877_at	Hs.20107.1.27	-0.43884	7.723	0.381019	0.049336	1.3018478
WIPI2	214699_x_at	Hs.122363.1.30	0.815474	7.711061	0.690051	0.089488	1.6184205
DCP2	235258_at	Hs.443875.1.31	0.556913	7.622591	0.319925	0.041971	1.248329
ZMYM3	1554171_at	Hs.522684.1.41	1.603828	7.478566	0.932167	0.124645	1.9268546
COL9A2	213622_at	Hs.418012.1.5	-1.65592	7.419562	1.011618	0.136345	1.9878395
PTPN11	212610_at	Hs.506852.1.26	-0.62247	7.304267	0.664188	0.090932	1.5776857
ZFP90	242400_at	Hs.461074.1.7	-0.30205	7.142789	0.675272	0.094539	1.6059068
ARPP19	214553_s_at	Hs.512908.1.13	-0.39911	7.084623	0.734931	0.103736	1.657376
SCRN3	222849_s_at	Hs.470679.1.14	0.071343	7.078383	0.636846	0.089971	1.5566671
ATG12	213930_at	Hs.483127.1.2	0.564647	6.9883	0.450822	0.064511	1.3660768
SLC25A17	211754_s_at	Hs.474938.1.11	0.250557	6.938119	0.598275	0.08623	1.5108095
ARPC2	213513_x_at	Hs.529303.1.29	0.121732	6.89833	0.330204	0.047867	1.2574985
SLC25A16	214140_at	Hs.180408.1.3	-0.98906	6.889637	0.408503	0.059292	1.3279775
AGBL5	231857_s_at	Hs.138207.1.20	-0.01819	6.807596	0.460146	0.067593	1.3775839
SNAPC1	205443_at	Hs.179312.1.13	-0.48827	6.799215	0.553989	0.081478	1.4709727
GNL1	203307_at	Hs.83147.1.3	0.663074	6.47381	0.520713	0.080434	1.4353423
MED4	222438_at	Hs.181112.1.5	-0.56853	6.432151	0.449443	0.069874	1.3640698
LRRC57	229232_at	Hs.234681.1.2	-0.41837	6.410457	0.357806	0.055816	1.2819729
EIF5	208708_x_at	Hs.433702.1.28	-0.44004	6.383058	0.329089	0.051557	1.2550405
WDFY2	227490_at	Hs.208550.1.16	-0.56338	6.376346	0.38144	0.059821	1.3023478
RCSD1	225763_at	Hs.493867.1.26	-1.24361	6.30058	0.34383	0.054571	1.2683434
MYH11	207961_x_at	Hs.460109.1.41	-0.34448	6.218151	0.614067	0.098754	1.5245816
POT1	204354_at	Hs.31968.1.4	0.934136	6.130178	0.589788	0.096211	1.5004686
EIF2S3	224936_at	Hs.539684.1.31	1.442625	6.126629	0.36698	0.059899	1.2905408
STK3	204068_at	Hs.492333.1.3	1.09033	6.116352	0.309882	0.050665	1.2389835
UBE2D3	200668_s_at	Hs.518773.1.7	0.441266	6.071738	0.412315	0.067907	1.3299639
MAGEH1	218573_at	Hs.279819.1.3	1.551685	6.055461	0.315552	0.05211	1.2446123
SMARCC1	201075_s_at	Hs.476179.1.4	0.282402	6.041569	0.447788	0.074118	1.3627625
RPL3	211666_x_at	Hs.119598.1.45	0.243807	6.002161	0.359989	0.059977	1.2818357
TRAK1	202079_s_at	Hs.535711.1.41	0.27331	5.981543	0.484609	0.081017	1.4042693
ABI1	209028_s_at	Hs.508148.1.4	-1.04665	5.978069	0.433434	0.072504	1.3519118
IGL@	209138_x_at	Hs.449585.1.64	0.230271	5.917118	0.628661	0.106245	1.5494146
IGL@	209138_x_at	Hs.449585.1.63	0.229124	5.917118	0.628661	0.106245	1.5494146
IGL@	209138_x_at	Hs.449585.1.58	0.228014	5.917118	0.628661	0.106245	1.5494146
YPEL5	217783_s_at	Hs.515890.1.16	-0.00937	5.881544	0.3134	0.053285	1.2421267
PPIH	204228_at	Hs.256639.1.18	-1.63794	5.790391	0.378591	0.065383	1.3015186
TDGF1	206286_s_at	Hs.385870.1.17	0.281253	5.780181	0.699375	0.120995	1.6196091
C8orf76	225702_at	Hs.521800.1.2	1.132091	5.731122	0.381702	0.066602	1.3034843
SERPINB1	213572_s_at	Hs.381167.1.4	0.636444	5.70877	0.439996	0.077074	1.355842
SLC16A1	209900_s_at	Hs.75231.1.3	-1.35464	5.699339	0.47291	0.082976	1.3842754

Table B.6 (continued)

Gene name	Probeset ID	Poly A site ID	Expected score	Observed score	Numerator	Denominator	Fold change (unlogged)
RGL1	209568_s_at	Hs.497148.1.29	-1.17775	5.68917	0.469106	0.082456	1.3840044
CCT6A	201327_s_at	Hs.82916.1.21	0.859467	5.608499	0.369712	0.06592	1.2944047
NBAS	202926_at	Hs.467759.1.4	-0.03727	5.604252	0.30729	0.054832	1.2378598
TNIK	213109_at	Hs.34024.1.6	0.342191	5.565248	0.389971	0.070072	1.3128029
RPL13	212933_x_at	Hs.410817.1.11	-0.28461	5.526751	0.445698	0.080644	1.3591381
SLU7	231718_at	Hs.435342.1.10	0.615502	5.500828	0.463473	0.084255	1.3746166
SLU7	231718_at	Hs.435342.1.9	0.614113	5.500828	0.463473	0.084255	1.3746166
SPTLC1	202277_at	Hs.90458.1.3	1.2369	5.49175	0.368128	0.067033	1.2906532
ATP6V1G1	238765_at	Hs.388654.1.5	1.313413	5.489677	0.351348	0.064002	1.2754904
EIF2B3	218488_at	Hs.533549.1.10	-1.58013	5.449824	0.462532	0.084871	1.3802625
FAM92A1	228011_at	Hs.125038.1.23	1.07037	5.371749	0.339767	0.063251	1.2636511
KREMEN1	235370_at	Hs.229335.1.9	0.233622	5.366154	0.58057	0.108191	1.490515
PCYT1A	204210_s_at	Hs.435767.1.19	0.358393	5.352134	0.499839	0.093391	1.4180946
PCYT1A	204210_s_at	Hs.435767.1.20	0.359568	5.352134	0.499839	0.093391	1.4180946
KBTBD3	228777_at	Hs.101949.1.11	-0.77339	5.337503	0.280465	0.052546	1.2148762

Table B.7Significantly 3’UTR lengthened genes in ERE-binding defective ERα-infected MDA-MB-231 cells compared to ERα-infected MDA-MB-231 cells.

Gene name	Probeset ID	Poly A site ID	Expected score	Observed score	Numerator	Denominator	Fold change (unlogged)
PML	235508_at	Hs.526464.1.16	-0.37843	-15.6844	-0.91905	0.058596	0.52855873
RPS6KB1	204171_at	Hs.463642.1.34	-0.20793	-11.0468	-0.83455	0.075547	0.5595159
YME1L1	216304_x_at	Hs.499145.1.10	-1.04433	-10.1754	-1.84064	0.180891	0.28279927
PPP4R1L	233069_at	Hs.473179.1.10	0.186588	-9.47417	-0.75493	0.079683	0.5917893
CLDN22	222738_at	Hs.333179.1.37	0.487052	-9.10301	-0.85949	0.094419	0.554145
ADH5	208847_s_at	Hs.78989.1.10	0.435232	-8.98932	-0.763	0.084879	0.5897172
DCAF13	231784_s_at	Hs.532265.1.24	1.113245	-8.97864	-0.62355	0.069448	0.64876527
USP9X	201100_s_at	Hs.77578.1.84	1.481632	-8.73002	-0.77546	0.088827	0.58229667
PANK2	218809_at	Hs.516859.1.22	0.141801	-8.62692	-0.39771	0.046101	0.7594598
TBC1D9	212956_at	Hs.480819.1.3	0.458356	-7.3502	-0.8834	0.120187	0.53757834
FLII	212024_x_at	Hs.513984.1.1	-0.26647	-7.19178	-0.69093	0.096073	0.6178902
G3BP2	208841_s_at	Hs.303676.1.3	0.413379	-6.73972	-0.25918	0.038455	0.83563673
ZNF625	244406_at	Hs.512823.1.2	-0.10315	-6.55469	-0.72501	0.110609	0.6026725
ZNF625	244406_at	Hs.512823.1.1	-0.10424	-6.55469	-0.72501	0.110609	0.6026725
SNHG10	238691_at	Hs.448753.1.2	-0.44987	-6.35629	-0.95907	0.150885	0.51152927
ANKRD12	216550_x_at	Hs.464585.1.21	-0.16416	-6.35487	-0.4454	0.070088	0.735094
CLDN22	222738_at	Hs.333179.1.38	0.488273	-6.32179	-0.53525	0.084667	0.69293326
CBS	1553972_a_at	Hs.533013.1.7	0.205729	-6.26499	-0.24201	0.038629	0.84570926
LDHA	200650_s_at	Hs.2795.1.18	-0.8578	-6.25189	-0.23307	0.037281	0.8508745
RAP2B	238622_at	Hs.98643.1.4	0.332988	-6.1899	-0.48882	0.07897	0.7149284
HADHA	208630_at	Hs.516032.1.4	-0.02606	-5.99499	-0.84346	0.140694	0.5658113
HADHA	208630_at	Hs.516032.1.3	-0.02722	-5.99499	-0.84346	0.140694	0.5658113
AFF1	201924_at	Hs.480190.1.39	0.426729	-5.97131	-0.27733	0.046443	0.82535845
FERMT1	218796_at	Hs.472054.1.3	0.144093	-5.9439	-0.33576	0.056488	0.7924388