TRANSCRIPTION PATTERN COMPARISON OF TWO UBIQUITIN SPECIFIC PROTEASES (*USP6* AND *USP32*)

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

BY

SHIVA AKHAVAN TABASI

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

AUGUST 2007

Approval of the Thesis

TRANSCRIPTION PATTERN COMPARISON OF TWO UBIQUITIN SPECIFIC PROTEASES (*USP6* AND *USP32*)

submitted by **SHIVA AKHAVAN TABASI** in partial fulfillment of the requirements for the degree of **Master of Science in Biology Department, Middle East Technical University** by,

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

Prof. Dr. Zeki Kaya _____
Head of the Department, **Biology**

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

Examining Committee Members

Prof. Dr. İnci Togan _____
Biology Dept., METU

Assist. Prof. Dr. A. Elif Erson _____
Biology Dept., METU

Assoc. Prof. Dr. Sertaç Önde _____
Biology Dept., METU

Assist. Prof. Dr. Özlen Konu _____
Molecular Biology and Genetic Dept., Bilkent University

Dr. Sreeparna Banerjee _____
Biology Dept., METU

Date: 28/08/2007

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: Shiva Akhavan Tabasi

Signature:

ABSTRACT

TRANSCRIPTION PATTERN COMPARISON OF TWO UBIQUITIN SPECIFIC PROTEASES (*USP6* AND *USP32*)

Akhavan Tabasi, Shiva
M.Sc., Department of Biology
Supervisor: Assist. Prof. Dr. A. Elif Erson

August 2007, 93 pages

Breast cancer is the most common type of cancer among women worldwide. The incidence of breast cancer is 1 in 8 among women. Usually loss of tumor suppressor genes and overexpression of proto-oncogenes are known to be involved during mammary tumorigenesis. *USP32* (Ubiquitin Specific Protease 32) gene is located on chromosomal band 17q23, a region of amplification in breast cancer. Gene amplification is known to be a common mechanism in breast cancer cells, through which proto-oncogenes are overexpressed and contribute to tumor progression. Presence of multiple oncogene candidates on 17q23 requires individual characterization of these genes.

USPs (Ubiquitin Specific Protease), have various roles in protein degradation pathways (e.g; by editing the ubiquitin chains, recycling of ubiquitin,

deubiquitinating the target proteins and inhibiting their degradation by the proteasome). Deregulated expression of USPs is likely to interfere with the degradation of many key regulatory proteins in the cell. Therefore, *USP32* becomes an interesting oncogene candidate that may have roles in protein degradation pathways based on the fact that it is located on an amplicon region and that it is overexpressed in breast tumors.

On the other hand, *USP6* (Ubiquitin Specific Protease 6), a known oncogene on 17p13, is also a deubiquitinating enzyme, with conserved histidine and cysteine domains, which are also shared by *USP32*. Interestingly there is a 97% sequence similarity between bases 3,197 to 7,831 of *USP6* and 2,390 to 7,024 of *USP32* gene.

In this study, we aimed to investigate the expression patterns of *USP32* and *USP6* (including alternative transcripts) in breast tissue to avoid any possibility of overlapping functions of two enzymes due to their high sequence similarity.

In addition, we sub-cloned *USP32* gene into TOPO-TA vector, so that further functional studies (e.g; localization and overexpression) can be performed. Further characterizations of Ubiquitin Specific Protease 32, may help us understand its importance in the protein degradation pathway during breast tumorigenesis.

Key words: Oncogenes, Amplification, Overexpression, 17q23, Ubiquitin Specific Proteases (USPs).

ÖZ

İKİ UBİKUİTİN'E ÖZGÜ PROTEAZIN (*USP6* VE *USP32*) GEN İFADE ŞEKİLLERİNİN KARŞILAŞTIRILMASI

Akhavan Tabasi, Shiva
Yüksek Lisans, Biyoloji Bölümü
Tez Yöneticisi: Y. Doç. Dr. A. Elif Erson

Ağustos 2007, 93 sayfa

Meme kanseri kadınlar arasında en yaygın olan kanser türüdür ve her sekiz kadından birinde bu hastalığın çıkma olasılığı vardır. Tümör oluşumunda genellikle tumor baskılayıcı genlerin kaybı ve/veya proto-onkogenlerin fazla ifade edilmesi rol oynar. *USP32* (ubikuitine özel proteaz 32) geni 17q23 kromozomal bandında bulunur ve bu bölgenin meme kanserinde çoğaldığı gösterilmiştir. Gen çoğalması da meme kanseri hücrelerinde yaygın olarak görülen bir durumdur. Bu durumda proto-onkogenlerin fazla ifade edilip tümor oluşumuna yol açtığı düşünülür. 17q23 bölgesindeki çeşitli onkogen adayları da bu bölgenin karakterizasyonunun önemli olduğunu işaret eder.

USP32 geninin, 17q23 amplikonunda bulunur, 212 kb'lik yer kaplar ve bir ubikitine özgü proteaz kodladığı düşünülmektedir. *USP32*'nin de içinde yer aldığı

USP protein ailesi üyelerinin protein yıkımında çeşitli görevleri vardır (hedef proteinlerin deubikitinlenmesi–ubikitinden arındırılması- ve proteazomda gerçekleşecek olası yıkımın engellenmesi gibi). Böylelikle, USPlerin fazla ifadesi hücredeki önemli regülasyon rolleri olan proteinlerin düzeyini değiştirebilir. Çeşitli kanıtlar, meme kanseri hücrelerinde *USP32* geninin fazla ifade edildiğine işaret etmektedir. Bu da *USP32* geninin meme kanseri oluşumundaki olası rolünü akla getirir.

Öte yandan, başka bir ubikitine özgü proteaz olan ve bir onkogen olduğu bilinen *USP6*, *USP32*'de de bulunan korunmuş histidine ve sistein bölgelerini içeren bir deubikitinleyici enzimdir. Bunun yanında *USP6*'nin 3.197 ve 7.831 nükleotidler arasındaki dizi ile *USP32*'nin 2.390 ve 7.024üncü nükleotidleri arasında %97lik bir benzerlik gözlenir.

Bu çalışmada, meme dokusunda *USP6* ve *USP32* genleri arasındaki ifade şekillerini (alternatif yazılım ürünlerini de içerecek şekilde) karşılaştırmayı amaçladık. İki protein, yüksek oranda benzerlik içerdiğinden, ortak görevlerde yer alabileceklerini de göz önüne alarak karşılaştırmayı yaptık.

Bununla birlikte, *USP32* genini, ileride lokalizasyon ve fazla ifade çalışmalarında kullanmak üzere TOPO-TA vektörüne klonladık.*USP32* geninin ve proteininin ayrıntılı karakterizayonu meme kanseri oluşumunda protein yıkım mekanizmasının rolünün aydınlatılması açısından yararlı olacaktır.

Anahtar kelimeler: Onkogenler, Amplifikasyon, genin fazla ifade edilmesi (overexpression), 17q23, ubikitine özgü proteazlar (USP).

... TO AORKAMM

ACKNOWLEDGEMENT

I am deeply thankful to my supervisor Assist. Prof. Dr. A. Elif Erson for her endless help, continuous advice, patience and encouragement throughout this study.

I would like to thank Assist. Prof. Dr. Uygar Tazebay from Bilkent University for the MCF7 and MDA-MB- 231 breast cancer cell lines, Dr. Volkan Seyrantepe for the pEGFP-N1 vector and Turkish Scientific and Research Council (TUBITAK) for their partial support.

I would like to thank my friends in the Molecular Biology and Genetic Lab for their friendship and support. Also I would like to thank Bilgen Ekim for her previous work on a part of this project.

Finally, I would like to express my best regards to my family and my husband for their endless love and great support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBEREVIATIONS

a.a	amino acid
bp	base pair
ddH_2O	double distilled water
DUBS	Deubiquitinating enzymes
DMSO	Dimethyl Sulfoxide
EtBr	Ethidium Bromide
Kb	Kilo base
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
rpm	revolution per minutes
Tm	Melting Temperature
Ub	Ubiquitin
UBP	Ubiquitin Proteases
USPs	Ubiquitin Specific Proteases
UPS	Ubiquitin Proteasome System

CHAPTER1

INTRODUCTION

1.1. Cancer

Cancer is one of the main problems in biology and can be described as a process that leads to a rapid and abnormal growth of cells, in which a normal cell is transformed into a malignant cell. Cancer may develop as a result of disruption of the cell cycle and it begins when a cell starts to divide more frequently, or die less than it should, when compared to the normal cell type that the cell is derived from. Different factors from the inside and outside of the cells can disrupt the normal cell proliferation. Growth factors and signaling molecules as well as transcription factors are important regulators which control the events inside the cell.

Three type of genes; proto- oncogenes (genes that usually promote cell growth), DNA repair genes and tumor suppressors (genes which suppress cell growth), are potential genes that may go through alteration in cancer. Any disruption in the normal expression of these genes can change the mechanism of their actions and lead to cell proliferation. In addition, such important genes can be deregulated under some mechanisms such as gene amplification, over expression, and mutations in the cells.

1.2. The Etiology of Breast Cancer

Breast cancer is the most important cause of cancer death in women after lung cancer in developing countries. Many researches have been performed to find the risk factors of breast cancer and incidence of this disease was investigated among countries, inside the migrating populations and also within the countries over time (Parkin *et al*., 1992. Ziegler *et al*., 1993. Kliewer *et al*., 1995. Hermon *et al*., 1996. Ries *et al*., 2001).

There are some known risk factors for breast cancer such as genetic and familial causes as well as hormonal, lifestyle and environmental factors. Other factors including; height among postmenopausal women (Van den Brandt *et al*., 2000), benign breast disease, mammographically dense breasts, menopause age (less than 45 and more than 54), high endogenous estrogen levels, post menopausal hormone use, exposure to ionizing radiation in childhood, menarche age (less than 12 and more than 14), alcohol use, oral contraceptive use, high prolactin levels and high premenopausal insulin like growth factor levels are factors, which increase the risk of breast cancer. Physical activities, breast feeding and non steroidal anti- inflammatory drug use are known to decrease the risk of breast cancer (Hankinson *et al*., 2004). However, multiple unknown factors and mechanisms are also involved in human breast tumorigenesis.

Inherited breast cancer, in which germline mutations pass trough the generations, is the case of about 5 to 10% of the whole breast cancer cases (Foulkes *et al*., 1995) and among these cases, 20 to 30% are caused by mutations in *BRCA1* and *BRCA2* genes. These two well known genes are responsible for transcriptional regulation and DNA repair mechanisms. There are some important genes in breast cancer such as, laminin 3A, *MYC*, *cyclin D1, STAT1* and *JAK1*, which their expressions are regulated by *BRCA1* gene (Welcsh *et al*., 2002).

In addition, it is claimed that, there are 13 polymorphisms in 10 genes, related to breast cancer, which alter the risk of the disease in different individuals (De Jong *et al*., 2002). The gene which codes for androgen receptor protein and *SULT1A1* gene are examples of these genes (Giguere *et al*., 2002 and Zheng *et al*., 2001).

At the level of hormones, estrogen and progesterone have the potential of causing cell proliferation. Although these two hormones are essential for growth of mammary glands, they can also bind to special receptors in the cell and cause cancer (Henderson *et al*., 2000). For example, estradiol, the major estrogen found in women, can bind to the estrogen receptor alpha and induces DNA synthesis, cell division and also promotes the production of growth factors and progesterone receptor proteins (Lillian *et al*., 2002) Also estrogen is capable of getting metabolically activated and causing damage on the genome and the cells by producing DNA adducts (Lacroix *et al*., 1997 and Bianco *et al*., 2003). Any reason that causes a long exposure to endogenous and exogenous estrogen will be considered as a risk factor for breast cancer.

Differences in the incidence of breast cancer in different countries (with a higher rate in developing countries and a lower rate in Asia); reveal the importance of environment and lifestyle in the appearance of this disease.

1.3. Gene Amplification and Breast Cancer

Amplification alters the copy number of the genes and this kind of genetic change in the case of critical genes may lead to tumor progression. Based on the Comparative Genomic Hybridization (CGH) analysis, 24 independent amplicons were identified in breast cancer (0.2-12 Mb in size). These amplicons are found on chromosomes; 1q21, 17q12–q21.2, 17q22–q23, 20q13.1, 20q13.2, 9p13, and 17q21.3 (Hyman *et al*., 2002). In some tumor types, the amplification of these

chromosomal regions is associated with their overexpression, which is a critical condition for cancer causing genes.

1.4. 17q23 Amplicon and Candidate Genes on this Chromosomal Band

A number of chromosomal regions are detected in breast cancer in which amplification occurs frequently. These regions have the potential of carrying important genes, essential for tumor progression (Knuutila *et al*., 1998). 17q23 is also one of these amplification sites and requires more attention, as it is a gene-rich region.

Various studies for determining the limits of 17q23 amplicon have been performed. According to the results of Southern Blotting and Fluorescent *in situ* Hybridization (FISH) based studies, the size of 17q23 amplicon was predicted to be 4 Mbp and 50 genes were estimated to exist in that region (Colleen *et al*., 2003). Further studies (Barlund *et al*., 1997, 2000; Couch *et al*., 1999; Erson *et al*., 2001; Monni *et al*., 2001; Wu *et al*., 2001) and the information available in the Human Genome Databases (http://www.ncbi.nlm.nih.gov/mapview/ and http://www.ensembl. org/Homo _sapiens/index.html) suggested a 5Mbp region, involved in 17q23 amplification, containing 29 genes with known functions (J Parssinen *et al*., 2007).

Several target genes have been located on 17q23 amplicon such as; *APPBP2* (also known as *PAT1*), *RAD51C, TBX2, TRIM37* (*MUL*), *THRAP1* (*TRAP240*), *PPM1D,* and *BRIP1* (Barlund *et al*., 2000; Wu *et al*., 2000, 2001; Erson *et al*., 2001; Monni *et al*., 2001; Bulavin *et al*, 2002; Li *et al*., 2002). The list and location of the genes, existing in this area are shown in Figure 1.1 and Table 1.1.

Table 1.1. The list of 29 genes, resided on 17q23, inside the 5 Mbp amplicon (Taken from J Parssinen *et al*., 2007)

Gene	Description
SEPT4	Peanut-like protein 2
TEX14	Testis expressed sequence 14
RAD51C	DNA repair protein RAD51 homolog
PPM1E	Protein phosphatase 1E
TRIM37	Tripartite motif-containing 37 protein
FAM33A	Family with sequence similarity 33, member A
YPEL2	Yippee-like 2 protein
DHX40	DEAH (Asp-Glu-Ala-His) box polypeptide 40
CLTC	Clathrin heavy chain 1
PTRH2	Peptidyl-tRNA hydrolase 2
TMEM49	Transmembrane protein 49
TUBD1	Tubulin δ chain
RPS6KB1	Ribosomal protein S6 kinase
ABC1	Amplified in breast cancer
CA4	Carbonic anhydrase IV precursor
USP32	Ubiquitin C-terminal hydrolase 32
APPBP2	Amyloid β precursor protein binding protein 2
PPM1D	Protein phosphatase 2C δ isoform, magnesium-dependent
BCAS3	Breast carcinoma amplified sequence 3
TBX2	T-box transcription factor TBX2
TBX4	T-box transcription factor TBX4
BRIP1	BRCA1 interacting protein C-terminal helicase 1
THRAP1	Thyroid hormone receptor-associated protein
TLK2	Serine/threonine-protein kinase tousled-like 2
MRC2	Mannose receptor, C type 2
CYB561	Cytochrome b561
ACE	Angiotensin-converting enzyme, somatic isoform precursor
KCNH6	Potassium voltage-gated channel, subfamily H, member 6
WRD68	WD-repeat protein 68

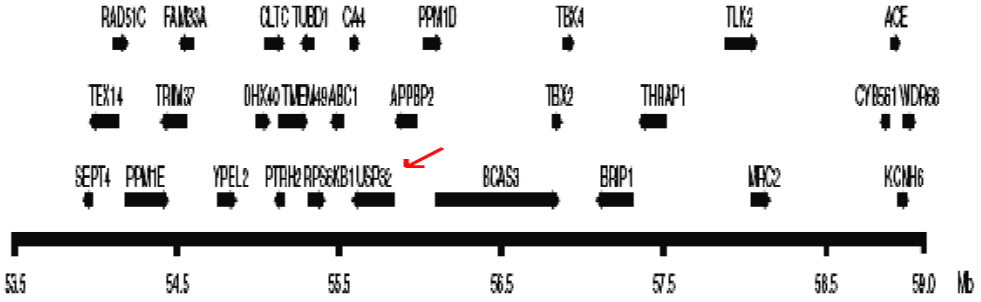


Figure1.1. The physical map of the 17q23 amplicon, which elucidates the location of 29 known genes in the area. The arrow shows the location of *USP32* gene in the amplification region (Figure was taken from; J Parssinen *et al*., 2007).

The amplification of 17q23 chromosomal band was first detected in breast cancer (Kallioniemi *et al*., 1994). The amplification of this region is not restricted just to breast cancer, it was also found in brain, lung, ovary, pancreas, bladder, testis, and liver tumors (Muleris *et al*., 1994; Ried *et al*., 1994; Korn *et al*., 1996; Solinas- Toldo *et al*., 1996; Brinkschmidt *et al*., 1997; Marchio *et al*., 1997; Richter *et al*., 1997; Schwendel *et al*., 1997; Sonoda *et al*., 1997; Weber *et al*., 1997; Vandesompele *et al*., 1998; Clark *et al*., 2002;.Willis *et al*., 2003). The highest amplification level of this region was detected in breast cancer (J Parssinen *et al*., 2007).

1.5. *USP32* and 17q23

USP32, which is the gene of our interest and codes for ubiquitin specific protease32 protein, is located on the 17q23.2 chromosomal band at the core region

(1.5 Mbp) of the 5 Mbp amplicon. Figure 1.2 shows the location of 17q23 band on the 17th chromosome.

Figure 1.2. The position of 17q23 band on the chromosome 17 (Taken from UCSC, genome browser).

The expression patterns of the 17q23 genes were investigated in 26 different breast cancer cell lines with high, moderate and no amplification of 17q23 amplicon by quantitative RT-PCR (Parssinen *et al*., 2007). Among these genes, eleven showed the significant expression difference between the high level and no amplification tumor types. *USP32* was one of these genes, which its amplification level was correlated with its expression level in the tumor cells.

1.6. Protein Degradation

1.6.1. Protein Degradation and Cancer

Degradation of unnecessary and abnormal proteins is one of the main responsibilities of the cells and any deregulation in this pathway can cause different problems for the cells. The cells can handle degradation, using two major mechanisms which are lysosomal and proteasomal degradation. Approximately

20% of unwanted proteins are degraded by lysosomal pathways and the rest 80% are degraded, using the proteasomal pathway (Rock *et al*,. 1999).

Many important pathways in the cells, such as signal transduction, transcription, cell cycle, differentiation, stress response, neuronal morphogenesis, cell surface receptor modulation, secretion, DNA repair, long-term memory, immune response, biogenesis of organelles and antigen processing (Goldberg *et al*,. 1995, Glickman and Ciechanover 2002) are very sensitive to protein degradation. As a result, many diseases like cancer can be the cause of deficiency in the proteolysis system.

1.6.2. Ubiquitination and Proteasomal Degradation System

Ubiquitin, a conserved regulatory protein is composed of 76 amino acids and is a main component of Ubiquitin- Proteasome System (UPS) pathway. The function of ubiquitin and the UPS system was first introduced in 1980's and for these studies, which were performed by Aaron Ciechanover, Avram Hershko and Irwin Rose, whom the 2004 noble prize in chemistry was awarded.

Ubiquitination is considered as a post translational modification, in which, ubiquitin is covalently attached to the target proteins and this is a reversible process which occurs in the cells, by the help of specific hydrolysis enzymes called Ubiquitin Specific Proteases. Normally deubiquitination happens, in the time and place that a protein should not be degraded and in this situation, the ubiquitin is removed from the protein. It can also occur in an abnormal manner in which a deubiquitinating enzyme, gets out of control and is expressed more.

A complex of enzymes is involved in the conjugating of ubiquitin to the target proteins. E1 (a ubiquitin activating enzyme), in an ATP- dependent manner,

binds to the C-terminus of the ubiquitin (Figure 1.3). In this reaction the glycine of ubiquitin is attached to thiol group of cysteine residues in the active site of E1 enzyme. The active site of cysteine residue of E2 enzyme (a ubiquitin conjugating enzyme), receives this ubiquitin. Finally, the E3 enzyme (a ubiquitin- protein ligase enzyme), catalyses the binding of ubiquitin to lysine residue of the substrate protein. Ubiquitins can also bind to each other in a way that C-terminus of one ubiquitin attaches to the amino group of the next ubiquitin (Hochstrasser, 2005). Ubiquitin chains, containing four or more ubiquitin residues, can bind to the target protein and introduce it to the 26S proteasome (Pickart, 2001, Thrower *et al*., 2000).

The 26S proteasome is composed of two 19S (regulatory) subunits and one 20S (core) subunit, which is responsible for proteolysis (Pickart and Cohen, 2004). Substrate proteins which arrive at the proteasome for proteolysis must get deubiquitinated. This is also necessary for keeping the free ubiquitin levels constant in the cell. In addition, a number of deubiquitinating enzymes such as JAMM protease POH1 (Rpn11 in yeast), UCH-L5, and USP14 (Ubp6 in yeast) have been detected in combination with the proteasome's 19S subunits which shows their relation in the proteasomal proteolysis pathway (Borodovsky *et al*., 2001; Park *et al*., 1997; Verma *et al*., 2002).

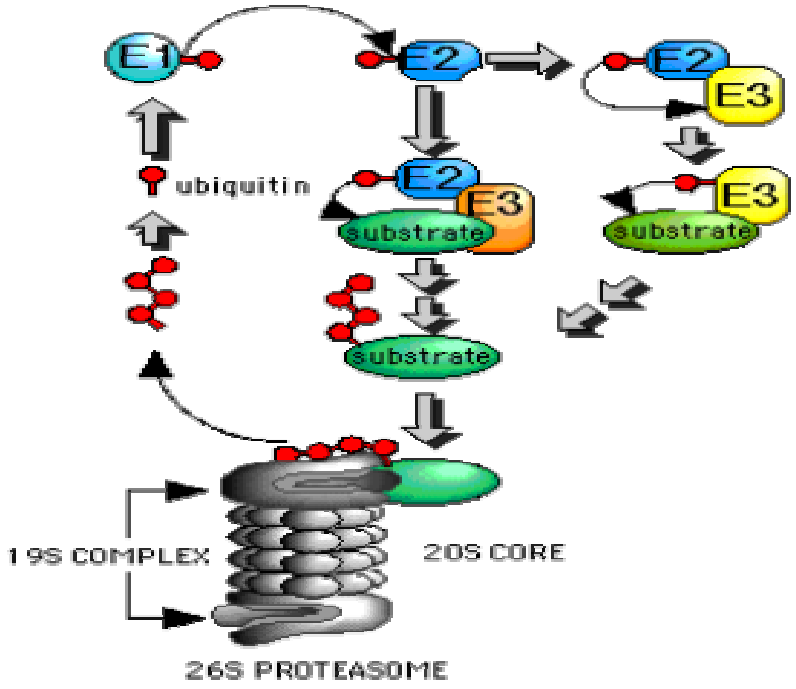


Figure 1.3. Protein degradation in the UPS system (Figure was taken from; http://www.uni-stuttgart.de/ibc/hilt/pstruc.html).

1.7. Deubiquitinating Enzymes (DUBs)

Proteases are a big superfamily with approximately 560 members in human (Puente and Lopez-Otin, 2004). Depending on their functions, proteases are divided in to 5 subclasses; aspartic, serine, threonine, metallo and cysteine proteases.

Most of the deubiquitinating enzymes belong to the cysteine protease group. This group of proteases also are divided into four subclasses, based on

their proteolytic domain; USP (Ubiquitin Specific Protease) also refers to UBP (Ubiquitin Protease), UCH (Ubiquitin C- terminal Hydrolase), OUT (Otubain Protease) and MJD (Joseph Disease Protease). The rest DUBs, belong to the metallo protease group of proteases. (Nijman *et al*., 2005). The conserved domains of UBP/USP enzymes are shown in Figure 1.4.

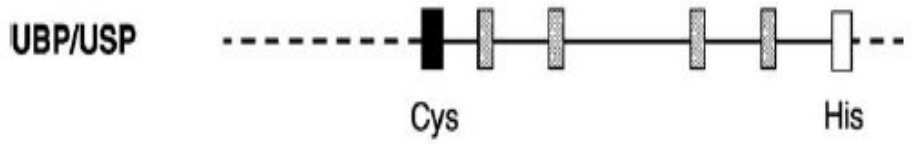


Figure 1.4. The conserved domains of UBP/USP enzymes (Wing, 2003).

The *USP32*, which is the gene of our interest, is thought to be a Ubiquitin Specific Protease (USP) with the conserved histidine and cysteine domains which are shown in the Figure 1.4.

Deubiquitinating enzymes remove ubiquitin from the substrate proteins, during the following mechanisms; processing, editing, recycling and reversal of ubiquitin conjugates. Deubiquitinating enzymes can also rescue ubiquitin from adducts with molecules such as glutathione or polyamines in the cell (wing, 2003). It has been claimed that overexpression of some USPs in cultured cells, causes a decrease in the level of ubiquitinated proteins (Yeh, 2000) Therefore, overexpression of USPs), maybe involved in many important regulatory

pathways. There are many proteins that are necessary in the regulation of the vital pathways in the cells (such as cell cycle regulator proteins) and are required just for a short period of time (after that, they could be harmful for the cells). Inhibition of degradation of these key regulatory proteins, which can be caused by the overexpression of USPs, is one of the main issues in the occurrence of many important disorders such as cancer. Figure 1.5 shows the catalytic functions of deubiquitinating enzymes.

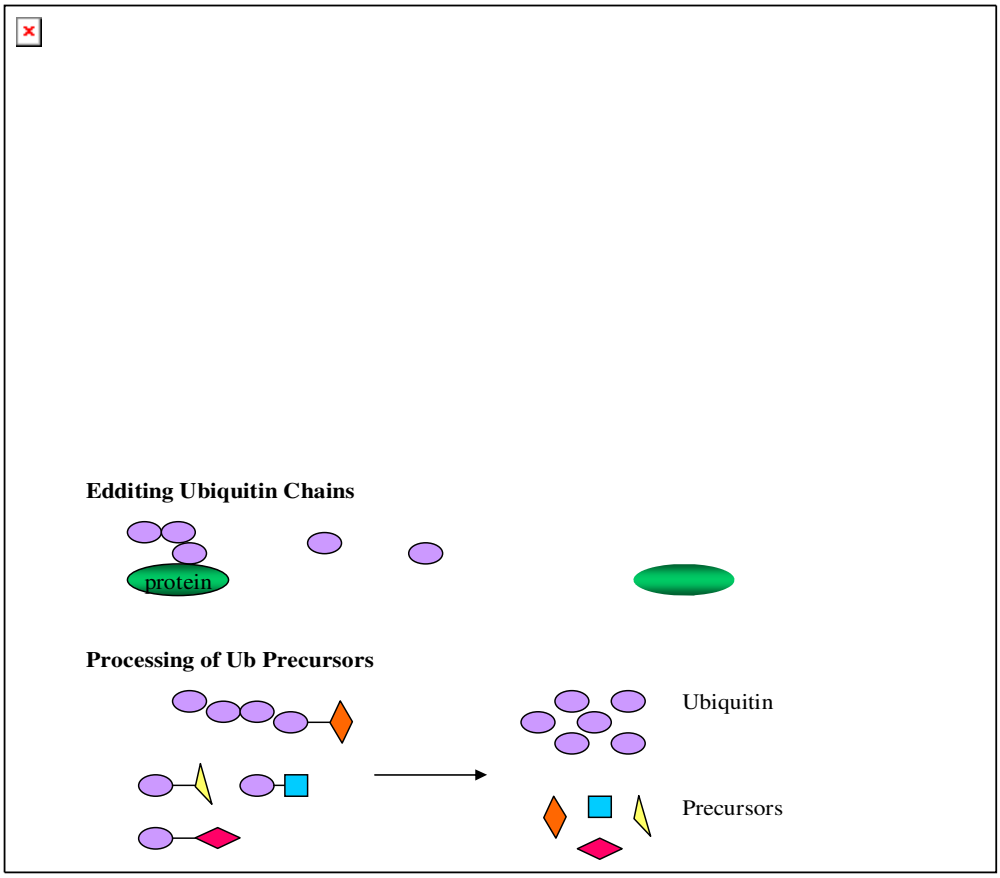


Figure 1.5. The catalytic functions of deubiquitinating enzymes.

1.8. *USP6* (*Tre-2 or Tre-17*) Gene and Protein Structure

USP6 (*Tre-17*, *Tre-2* oncogene), located on chromosomal band 17p13 is a Ubiquitin Specific Protease and a chimera gene derived from *USP32* (NY-REN-60, gene bank accession #; AF533230), which is an ancient and highly conserved gene, and *TBC1D3*, which is derived from a recent segmental duplication, that is new and also absent in a number of mammals. The *USP6* gene is a hominoid

specific gene which appeared approximately 21-33 million years ago, after the occurrence of *TBC1D3* segmental duplication in the primate lineage (Paulding *et al*., 2002). *USP32* and *TBC1DC* genes are expressed in a variety of human tissues but *USP6* was shown to be expressed in testis tissue (Paulding *et al*., 2002). The *in silico* analysis suggests presence of alternatively spliced forms of *USP6* whose expression profiles are not known.

Amino acids 1- 496 of *USP6*, which are derived from *TBC1DC3* gene, contain a TBC domain that is predicted to be involved in Rab GTPase signaling and vesicle trafficking (Neuwald, e*t al*., 1997). Amino acids 501- 1,406 of *USP6* are derived from *USP32* and contain a USP domain. The chimeric fusion, *USP6* gene, has lost two Ca-binding EF-hand domain and myristoylation site, which are normally present at the 5' end of *USP32* gene.

In silico analysis and early literature suggests that *USP6* has two different transcripts, clones 210 and 213 (which are suggested to be the result of alternative splicing), each having two open reading frames (Nakamura *et al*., 1992). Clone 210 ORF1 (2,348 base pair), clone 210 ORFII (1,750 bp), clone 213 ORF1 (1,130 bp) and clone 213 ORFII (3,269 bp).

1.9. *USP32* Gene and Protein Structure

USP32 (Accession #: NT 010783 and mRNA accession #: NM 032 582), an ubiquitin specific protease, is located on chromosomal band 17q23, which is found to contain an amplicon in breast cancer cells. *USP32* gene spans approximately 212 kb and has 34 exons as predicted by mRNA to genomic sequence alignment tools (Spidey, http://www.ncbi.nlm.nih.gov/spidey/). *USP32* gene codes for a 1,604 a.a long protein, which contains six domains, Ca^{2+}-binding protein domain (EF- Hand Super Family), from a.a 189 to 339, EF-hand, calcium

binding motif, from a.a 237 to 293, DUF1055 (a domain of unknown function), from a.a 548 to 712, Peptidase_C19E, from a.a 733 to 911, Peptidase_C19, from a.a 1,225 to 1,318 and Peptidase_C19R, from a.a 1,510 to 1,565 (Figure 1.6).

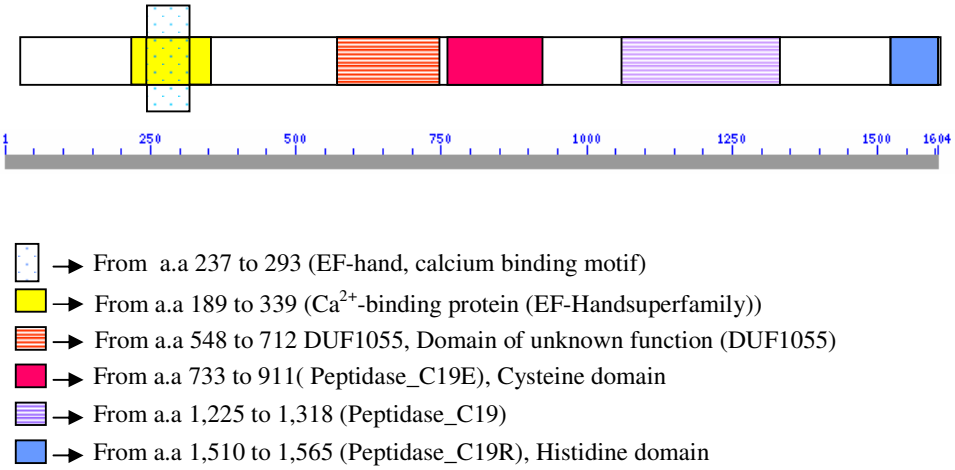


Figure 1.6. The protein domains of *USP32* gene. The numbers show the amino acid sequence of the gene. Based on NCBI conserved domain database (www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?INPUT_TYPE=precalc&SEQUENCE=22550104).

1.10. The Aim of the Study

In this study, the expression pattern of two Ubiquitin Specific Proteases (USPs), *USP32* and *USP6*, has been compared in different tissues and breast cancer cell lines. The *USP32* gene, resides on chromosomal band 17q23, on a region which is found to be amplified in a number of breast cancer cell lines and codes for a 1604 amino acid long protein, whereas *USP6,* a previously known oncogene, which its expression is claimed to be just in the testis tissue, is located on the chromosomal band 17p13 and codes for a protein containing 1,406 amino acids. Also two alternative possible transcripts are suggested for the *USP6* gene transcripts; clone 210 and clone 213, with each having two open reading frames. Although these genes are located on different chromosomal bands, they show a 97% sequence similarity in the C-terminus which two enzymes share. As *USP6* is known to be an oncogene, and *USP32* is located on an amplicon in breast cancer, which suggests an oncogenic role for this enzyme, we designed our experiments in a way that we can compare the expression pattern of these genes in different tissues and also in breast cancer cell lines. Based on the sequence similarity and any possibilities of functional overlap, expression profiles of these proteins needed to be investigated before any further functional analyses of *USP32.*

In addition, the full length and partial fragments of *USP32* gene, which contained different domains of the gene, were sub-cloned into the TOPO-TA vector and sequence analysis were performed for them. These constructs will be used for the further experiments such as localization and expression studies. Further characterization of *USP32* gene may reveal the importance of this gene in the breast cancer tumorigenesis processes.

CHAPTER 2

MATERIALS AND METHODS

2.1. Investigation of the Expression Pattern of *USP6*

2.1.1. Primers and Their Positions

For investigating the existence of possible transcripts of *USP6* gene in breast tissue, eight primers were designed which were specific for each transcript. Also as there is sequence similarity between *USP32* and *USP6* gene, we were careful to design primers without any hits to *USP32* gene in the case of investigating the *USP6* transcripts. Firstly the sequences of *USP6* full length mRNA (NM_032582), *USP6*- transcript 210 (X63546), *USP6*- transcript 213 (X63547) and *USP32* gene (NM_032582), which were available in NCBI (National Center of Biotechnology Information), were used to design primers, specific for each transcript and for this purpose, Primer-3 program was used. The melting temperature of primers was between 57°C - 60°C and the length of them was 18-23 bp.

Using blast program of NCBI, all sequences mentioned above were blasted with each other to find the sequence similarities and differences between them and also they were blasted with designed primers to make sure that they are specific for each transcript.

2.1.2. Cell Culture and Growing Conditions

MDA-MB-231 and MCF7 breast cancer cell lines were grown under the following conditions; a vial of each frozen cell was removed from the storage and quickly thawed in a 37°C water bath. After melting the culture, the cells were plated into a T-25 flask and prewarmed appropriate growth medium was added to the flask for each case. After one day, the cells were washed with Hanks buffer (to get rid of the DMSO (Applichem, Cat#; A1584) which was used for freezing of the cells) and treated with tripsin (PAA, Cat #: l11-002) and new medium was added to the flasks. Every 2-3 days, the old medium was sucked off from the flasks and new medium was added instead. After the cells got confluent, they were sub cultured as follows; the confluent cells were washed with Hanks buffer and treated with 2-4 ml tripsin, (which is a cell detaching agent) and were incubated for 2-3 minutes at 37°C. The cells were checked for their dissociation in tripsin. When they were separated, a few milliliters of culture medium containing serum were added to the flask to stop the activity of the enzyme. Then cells were divided into flasks and appropriate amount of culture medium was added to each flask in a way that covered the surface of the cultured cells. The cells were used for RNA isolation, when they were 70% confluent. The growth media used for MCF7 cells was Minimum Essential Medium with Earl's Salt and L-Glutamine (PAA, Cat #; E15-825), supplemented with 10% fetal calf serum (PAA, Cat#; A11640) and 1% penicillin-streptomycin (Biochrom AG, Cat #;A2212), whereas for MDA-MB-231 cells, the Dulbecco's Modified Eagle' Medium with L-Glutamine (from PAA, Cat #; E15-810), 10% fetal calf serum and 1% penicillin-streptomycin. The cells were grown at 37°C, in the existence of 5% carbon dioxide.

2.1.3. RNA Isolation from Cell Lines

Total RNA from MDA-MB-231 and MCF7 cells were isolated by using Trizol RNA isolation reagent. The RNA from these cells were isolated, when they were 70% confluent because they should actively do transcription. The cultured cells in the T75 flasks were treated with Trizol and incubated at room temperature to permit complete dissociation of nucleoprotein complexes. After 5 minutes, the cells were transferred to a 15 ml sterile blue cap tubes and 0.2 ml of chloroform per 1ml of Trizol (1.6 ml chloroform) was added to the tubes and the tubes were shaked vigorously for 15 seconds and incubated at room temperature for 2-3 minutes. The sample was centrifuged at 4700g for 20 minutes at 8ºC. The aqueous phase (volume was about 60% of the original Trizol volume and contains the RNA of the cell) was transferred to a fresh tube. The RNA from aqueous phase was precipitated by mixing with 4 ml isopropyl alcohol (0.5 ml per 1ml of Trizol) and incubating at room temperature for 10 minutes. The tube was centrifuged at 4700g for 20 minutes at 8ºC. The RNA pellet was washed with 8 ml 70% ethanol and then was centrifuged at 4700g for 7 minutes at 8ºC. At the end, RNase free water was used to dissolve the air dried RNA pellet.

2.1.4. cDNA Synthesis

The cDNA of MCF7, MDA-MB-231, normal breast, testis and ovary were synthesized using Revert Aid First Strand cDNA Synthesis Kit. The total RNA of breast, taken from a Caucasian female (Ambion, Cat #: 7952), stored in 1mM sodium citrate, pH: 6.4, testis from a gun shot wound of a Caucasian male (Ambion, Cat #: 7972), stored in 1mM sodium citrate, pH: 6.4 and ovary from a Caucasian female (Ambion, Cat #: 6974) tissue were received and we isolated the total RNA of MCF7 and MDA-MB-231 breast cancer cell lines (RNA isolation by Trizol method and concentration of the used RNA for cDNA synthesis was 1µg

per 20µl of prepared reaction). The Revert Aid First Strand cDNA Synthesis kit (Fermentas, Cat #; K1622) was used to prepare the cDNA strands.

The protocol that was used for each cDNA synthesis process is as follows; in a microcentrifuge tube, 1µ1of RNA (total 1-0.5 µg), 1µ1 of oligo (dT) primers (0.5 µg/µ1, provided by the kit) and 10µ1 DEPC-treated water were mixed gently and spin down for 3-5 seconds in a microcentrifuge. Then the tube was incubated at 70°C for 3 minutes and chilled on ice. The tube was briefly centrifuged and again placed on ice. 4µ1 of 5X reaction buffer, 1µ1 of Ribo lock Ribo nuclease inhibitor (20U/µ1) and 2µ1 of 10 mM dNTP mix was added and the tube was briefly mixed and centrifuged. After that the tube was incubated for 5 minutes at 37°C and then 1 µ1 of RevertAid M-MuLV Reverse Transcriptase (200U/µ1) was added to the mixture. The tube was incubated at 42°C for 60 minutes. Then the reaction of the enzyme was stopped by incubating the mixture at 70°C for 10 minutes. The tube was chilled on ice and the cDNA was stored at -20°C.

2.1.5. RT-PCRs and Duplex PCRs

The cDNA from MCF7, MDA-MB-231 cell lines and normal breast, testis and ovary tissue (prepared, using Revert Aid First Strand cDNA Synthesis Kit, Fermentas, K1622) were used in RT-PCR reactions. The forward and reverse primers which were specific for *GAPDH* gene were used in the PCR reactions as well as specific primers for *USP6* and its different transcripts. Each 30 µl PCR amplification reaction mixture contained 3µl of 10X *Taq* polymerase complete reaction buffer, 1µl of prepared cDNA from (part 2.1.4). A final concentration of 200µM of each deoxyribonucleoside triphosphates (dNTPs), and 0.5µM of each primer was also used. Sterile dH_2O (distilled dH_2O) was used to reach the volume to 30µl. 70µl of mineral oil was used to cover the mixture in the 0.5 ml volume test tube. The negative control PCRs also were prepared, in which all PCR

reagents were used except template DNA. 1.25 U of *Taq* polymerase enzyme (Applichem) was used in each reaction. The samples were incubated at 94ºC for initial denaturation and a thermal cycling for a total of 30-34 cycles were used: denaturation at 94ºC for 30 seconds, annealing at 58ºC for 30 seconds and elongation for 45 seconds at 72ºC. A final elongation for 10 minutes was performed at the end and the samples were analyzed for the correct size by gel electrophoresis using 1% agarose gel containing ethidium bromide at a final concentration of 0.5 µM/ml and appropriate size markers. For optimizing the duplex PCRs, the *GAPDH* primers were diluted and used for a final concentration of 0.025µM, 0.05µM and 0.075µM (different for each reaction).

2.2. Sub Cloning of *USP32* Gene

2.2.1. Primers for PCR Amplification of *USP32* Gene

The *USP32* gene was amplified by polymerase chain reaction (PCR). Specific primers (synthesized by Iontek, Alpha-DNA and IDT.DNA companies), which contained both *Xho1* and *Apa1* recognition and restriction sites, were designed and used to amplify the full length Coding Sequence (CDS) and partial fragments of *USP32* gene to be cloned into pEGFP-N1 vector (Clontech), considering the Open Reading Frame (ORF) of the vector in a way that the amplified PCR products didn't interrupt the open reading frame of the Pegfp-N1 vector. Figure 2.2 shows the map and multiple cloning site of pEGFP-N1 vector.

Extra nucleotides were added, after the restriction site of R1N and R2N reverse primers and the stop codon of the gene was removed by excluding the stop codon from the reverse primers, as the multiple cloning site of the vector is located before the EGFP (Enhanced Green Fluorescent Protein) gene and we want

the expression of the *USP32-GFP* tagged protein. For the purpose of further expressional investigations, the PCR products, using these primers were sub-cloned into TOPO- TA vector and were sequenced. Figure 2.3 shows the map of TOPO-TA vector. The sequence of the primers which were designed for each amplification of each fragment of the *USP32* gene are shown in the next page in which (*) shows the added nucleotides inside the sequence of primers and (- - -) shows the removed stop codon from the sequence.

R1N 5'-GGGCCC G - - - CTGTAACACACAGTACTTTTTGTAATCAG
(Apa1 site; * ; Sequence of *USP32*)

R2N 5'-GGGCCC T GGCTCCCTTTTCTGTGGGAAC
(Apa1 site; * ; Sequence of *USP32*)

F1N 5'-CCG CTCGAG ATG GGTGCCAAGGAGTCAC-3'
(Rec. site; Xho1 site; * * *; Sequence of *USP32*)

F2N 5'-CCG CTCGAG ATG GATAGACACAAGGTTCCCA
(Rec. site; Xho1 site; * * *; Sequence of *USP32*)

F3N 5'-CCG CTCGAG ATG CCAGGGGCTATTGATAATCA
(Rec. site; Xho1 site; * * *; Sequence of *USP32*)

The location of each primer on the *USP32* gene's full and partial fragments is also shown in Figure 2.1.

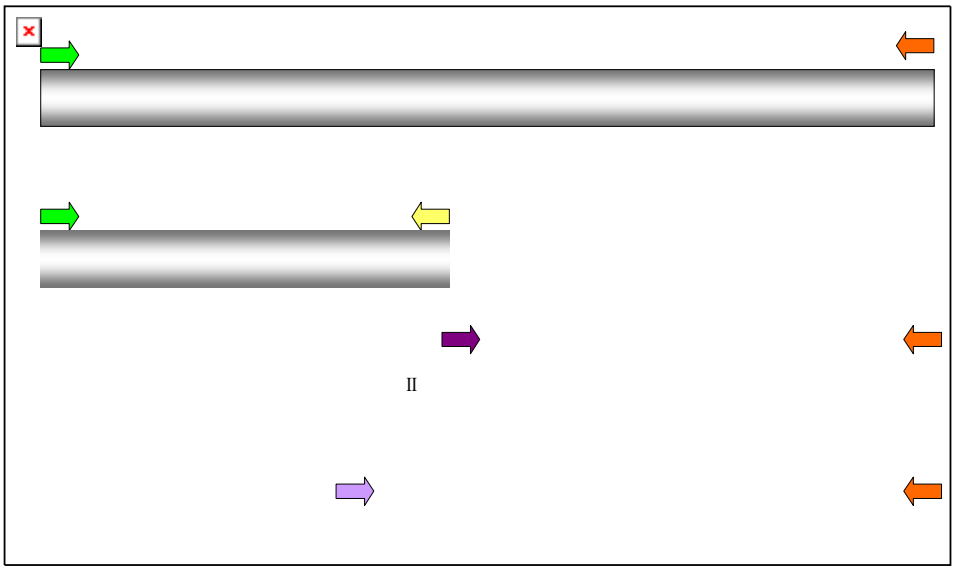


Figure 2.1. The location of primers on the full length (F) and three fragments of *USP32* gene (I, II, III).

2.2.2. The Map and Multiple Cloning Site of pEGFP-N1 Vector

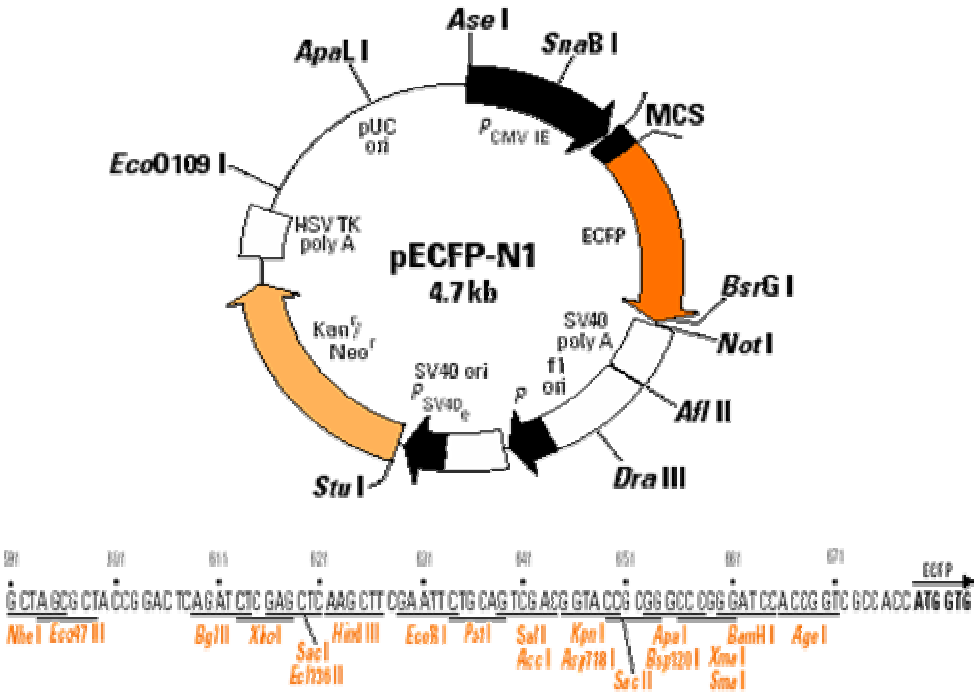


Figure 2.2. The map and multiple cloning sites of the pEGFP-N1 vector. *Xho1* and *Apa1* sites were used for primer designing which were available in the vector.

2.2.3. The Map of TOPO-TA Cloning Vector

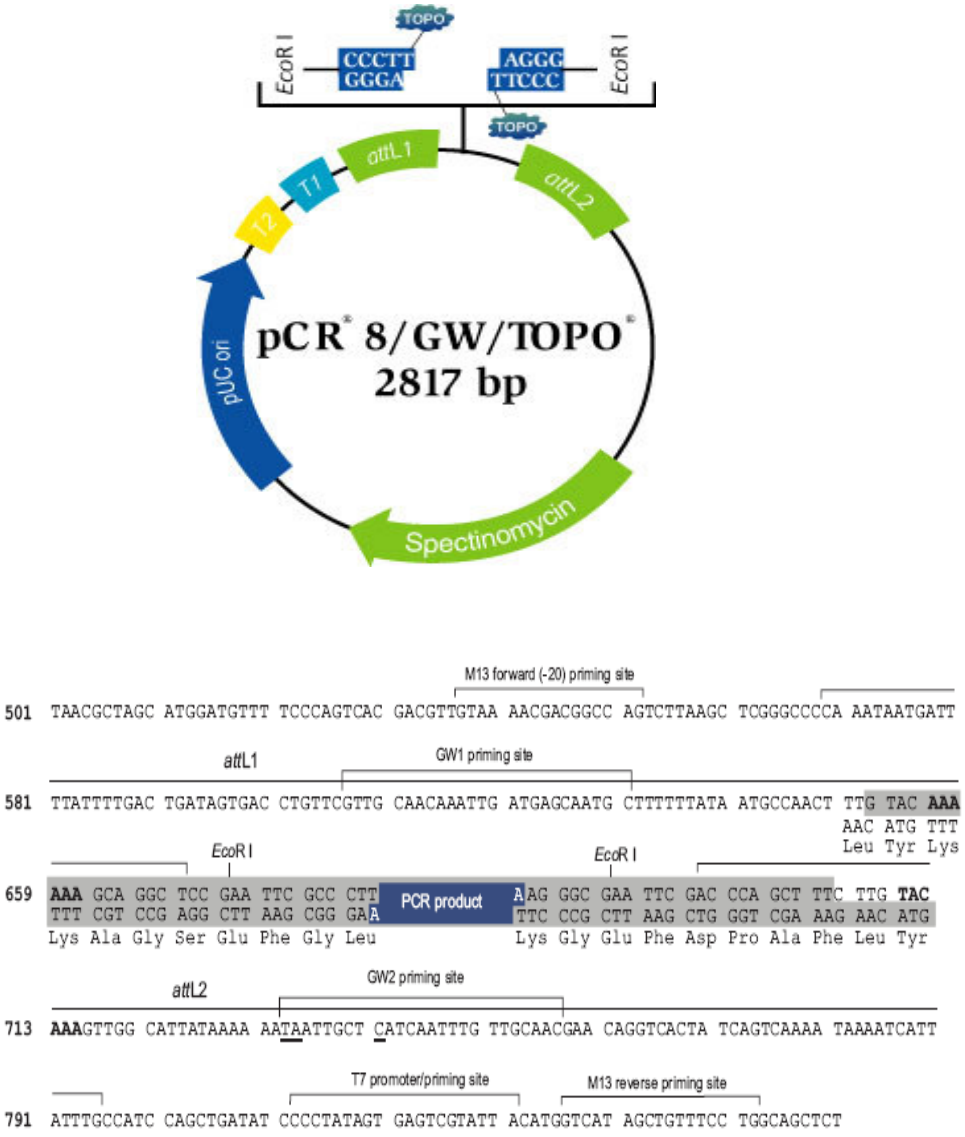


Figure 2.3. The map and cloning site of the TOPO-TA vector. PCR products containing two *Apa1* and *Xho1* restriction sites were cloned into this vector. Also *Apa1* (base 568) and *Xho1* (base 2,078) sites are available in the sequence of the vector.

2.2.4. PCR Optimizations for *USP32*

Specific primers were used for amplification of the different parts of *USP32* gene. The *PFU* DNA polymerase (Promega), Expand Long Read and Expand High Fidelity enzymes (Roche) were used for the optimization of the PCR products and finally the products were optimized, using Expand High Fidelity enzyme. The PCR amplification for NPA1 (2.2 kb, first fragment of the gene), NPA2 (2.6 kb, second fragment of the gene), NPA3 (a 3.2 kb long fragment of the gene) and full length cDNA of the *USP32* gene, was performed using appropriate primers and template DNA for each case. The PCR conditions were as follow; Each 50 µl PCR amplification reaction mixture contained 10µl of Expand High Fidelity, 5X reaction buffer 2, (1.5 mM $MgCl_{2)}$, 10 ng of template DNA (pTARGET plasmid containing the *USP32* gene), 200 µM of each deoxy ribonucleoside triphosphates (dNTPs), and 0.2- 0.6 µM of each primer. Sterile dH_2O (distilled dH_2O) was used to reach the volume to 50µl. 70µl (approximately 2 drops) of mineral oil was used to overlay the mixture in the 0.5 ml volume test tube. The negative control PCRs also were prepared, in which all PCR reagents were used except template DNA. 1.25U of Expand High Fidelity Enzyme (Roche, Catalog #: 3300226) was used in each reaction. The best temperatures and PCR program which was optimized for the primers were as follows; The samples were incubated at 94ºC for initial denaturation and a thermal cycling as follows for a total of 25-30 cycles was used: denaturation at 94ºC for 15 second, annealing at 73ºC for 30 second and elongation for 2-5 minutes (1 minute per kb, was calculated for each case) at 68ºC -72ºC (68ºC for fragments larger than 3 kb and 72ºC for fragments less than 3 kb). A final elongation for 7 minutes was performed at the end of the cycles (68-70ºC) and the samples were analyzed for the correct size, by gel electrophoresis, using 1% agarose gel containing ethidium bromide as the stain and appropriate size markers. The samples were used for inserting into TOPO-TA cloning. The correct cloned constructs which contained

restriction sites for both *Xho1* and *Apa1* enzymes, were sent for sequencing (Tables; 3.6, 3.7 and 3.8)

2.2.4.1. Agarose Gel Electrophoresis

Agarose was used to prepare the agarose gels, using TBE buffer. EtBr (at final concentration of 0.5µg/ml) was added to the agarose gel after cooling and before its solidification. 1µl of 10X loading dye was added per 10 µl of DNA samples before loading. The appropriate DNA marker was used to determine the size of the DNA in each experiment. Electrophoresis was carried out at 75 volts. Occasionally, 0.8%, 1% and 2% agarose gels were prepared and used in the experiments depending on the size of DNA.

2.2.4.2. DNA Extraction from Agarose Gels

The PCR products from part 2.2.4, which contained DNA fragments of full length (4.8 kb), and partial fragments of the *USP32* gene (2.2 kb, 2.6 kb and 3.2 kb) were run on the agarose gel containing 0.5µg/ml EtBr. After confirming the size of the products, the parts of the gel containing DNA were cut and the DNA was extracted by using the gel extraction kit (Roche, Catalog #: 11696505001), in which the DNA was eluted in dH_2O at the end of the extraction.

2.2.5. Cloning of PCR Products into the PCR® 8/GW/TOPO®TA Vector

The PCR® 8/GW/TOPO®TA Cloning Kit (Invitrogen, catalog #: 1375493), was used for cloning of *Taq*-polymerase amplified PCR products, into the TOPO-TA plasmid. The desired PCR products from *USP32* part1 (2.2 kb), part2 (2.6 kb) and part3 (3.2 kb), and also full length cDNA, were amplified and their sizes were confirmed by using gel electrophoresis.

The PCR products were generated by a DNA polymerase capable of adding adenine at the end of each strand. For this purpose *Taq* DNA polymerase was the best enzyme. Expand High Fidelity was the DNA polymerase that we used in the PCR reactions and this enzyme was supplied with *Taq* DNA polymerase. The desired, PCR amplified DNA, was run and extracted from the agarose gel. The extracted DNA was used to be cloned into TOPO-TA cloning vector as follows; 7µl of gel extracted DNA, 1µl of 10X *Taq* polymerase complete buffers, 1µl of 2mM dATP and 1 µl (5 unit) of *Taq* DNA polymerase to add adenine to the 3' end of DNA strands, were mixed gently in a 0.5 ml tube and was incubated at 70°C for 10 minutes. After 10 minutes the tube was placed on ice and used to be cloned into PCR® 8/GW/TOPO® vector. The procedure for cloning, provided by the manufacture was as follows: 1-4 µl of the previous mixture and 1µl of salt solution was mixed gently in a 0.5 ml reaction tube and water was added to reach the volume to 6µl. The mixture was incubated at 22°C for 5-30 minutes, depending on the size of the insert (more time is needed for larger DNA fragments to be inserted into the vector).

The PCR products that were cloned into PCR® 8/GW/TOPO® vector were as follows: 1-2µl of fresh PCR product (1µl for partial fragments and 2µl for the full length gene), 1µl of salt solution, 1µl of TOPO vector and 2-3µl of water (to reach the volume to 6µl. The mixture was mixed gently in a 0.5 ml reaction tube (final volume of 6µl) and was immediately incubated for 15-20 minutes at 22°C. Then the reaction was placed on ice and 2 µl of it was used in transformation of competent *E.coli* cells.

2.2.5.1. Preparation of Competent *E. coli* Cells

50 ml of LB medium was incubated in a 250 ml flask with a single or several *E. coli* colonies, at 37°C for 12- 16 hours (overnight). 300 µl from overnight growth culture was put in 50 ml incubated LB medium and grown till

OD600 reaches to 0.6 (about 2-3 hours). After the culture medium reached to the desired bacterial growth density, it was divided into 2 sterile prechilled centrifuge tubes and kept 10 minutes on ice, then centrifuged at 4000 rpm for 10 minutes. The pellet from this step was resuspend in 5 ml ice cold $CaCl_2$ (10 mM), on ice and centrifuged at 3000 rpm and 4°C for 10 minutes. Again the pellet inside each tube was dissolved in 1ml of 75 mM $CaCl_2$ and 200 µl of ice-cold glycerol was added to it. After making aliquots, the cells were frozen in liquid nitrogen and then stored at -80°C.

2.2.5.2. Transformation of the Competent *E. coli* Cells

The TOP10 and DH5-alpha chemically competent *E. coli* cells (stored in glycerol, at -80°C) were thawed on ice and 50-100µl were used in each transformation reaction tube. 2 µl of the plasmid (from each TOPO reaction) were added to the tubes (1pg– 10ng) and the tubes were incubated on ice again. After 30 minutes the mixture was incubated at 42°C for 45 second and immediately placed on ice. 300µl S.O.C medium (prewarmed at room temperature) was added to each vial after 5 minutes incubating on ice. The vials were incubated at 37°C and shaked at 200 rpm for 45 minutes. 50-100 µl of the transformed cells were cultured in LB-Agar plates containing spectinomycin for a final concentration of 100µg per ml. A successful transformation resulted in a number of bacterial colonies in the antibiotic containing LB-plates and some of these colonies were inoculated into LB-broth medium (containing spectinomycin for a final concentration of 100µg/µl) and stored for further use.

2.2.5.3. Storage of Bacterial Cultures

The transformed *E. coli* cells were stored at -80°C, using the following procedure; 0.75 ml of bacterial culture and 0.25 ml of sterile 60% glycerol were mixed and gently vortexed in a 1.5 ml storage tube equipped with a screw cap.

The culture was frozen in liquid nitrogen and then transferred immediately to - 80°C for long term storage. For the next usage, the frozen surface of the culture was scraped; using a sterile loop and the frozen culture on the loop was inoculated in new LB-antibiotic medium. The rest of the frozen culture which was not yet thawed was restored again at -80°C for the future usages.

2.2.5.4. Isolating and Confirming the Size and Sequence of the TOPO-Cloned Constructs

A number of colonies from plates containing 20ml LB and spectinomycin for a final concentration of 100µg/ml, were selected and isolated using Gene Mark (cat #: Dp01-50) and FAVORGEN (Cat #: FAPDE001-1) plasmid isolation kits, following the procedure provided by manufacture. The isolated plasmids were digested with *EcoR1, Apa1* and *Xho1* restriction enzymes to confirm their sizes. The correct colonies were selected and sent for sequencing to Iontek Company (Istanbul, Turkey) and McLab (San Francisco, USA). Among the sent constructs, those which contained both *Xho1* (at 5'end of the gene) and *Apa1* (at the 3' end) and also did not contain any none sense or missense mutations were chosen to be cloned into pEGFP-N1 vector.

CHAPTER 3

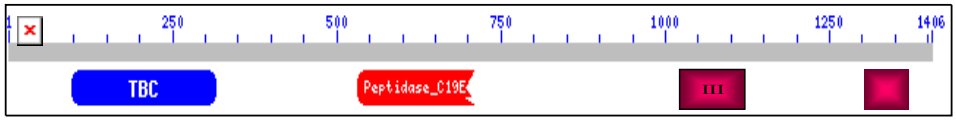
RESULTS AND DISCUSSIONS

3.1. Sequence Similarity between *USP6* and *USP32*

USP6 (Ubiquitin Specific Protease 6) gene, codes for a 1,406 a.a long protein and shows 89% nucleotide similarity to *TBC1D3* gene and 97% sequence identity to *USP32* gene (the similarity refers to sequences between 3,197th to 7,831st bases of *USP6* and 2,390th to 7,024th bases of *USP32* gene). *USP6* gene codes for a deubiquitinating enzyme, which contains two conserved histidine (from amino acids 530-714) and cysteine (from amino acids 1,303-1,366) domains that are necessary for the deubiquitinating function of the enzyme (Figure 3.1.A). Nucleotides 1-3,197 of *USP6* gene show similarity with *TBC1DC* gene. Amino acids 75-330 contain a TBC domain, whereas amino acids 189-339 of *USP32* gene contain an EF-hand calcium binding motif. Also a domain of unknown function exists between amino acids 548-712 of USP32 protein (Figure 3.1.B). The shared sequences between *USP32* and *USP6* genes which contain both conserved cysteine and histidine domains are critical for the enzymatic function of these proteins as these conserved domains were detected in all Ubiquitin Specific Proteases. Two different transcripts have been suggested for *USP6* gene; clones 210 and 213 (which are suggested to be the result of alternative splicing) with each having two open reading frames (Nakamura *et al.*, 1992). Clone 210 ORFI (2,348 bp), clone 210 ORFII (1,750 bp), clone 213 ORFI (1,130 bp) and clone 213 ORFII (3,269 bp) (Figure 3.2). Due to the high similarity between two USP6 and USP32 proteins, it is likely to have similar or

overlapping functions of these proteins in the cell. For revealing the role of *USP32* in breast cancer, we investigated the expression pattern of these two genes and their related transcripts in normal tissues or breast cancer cell lines by RT-PCR.

A. The domains of USP6



B. The domains of USP32

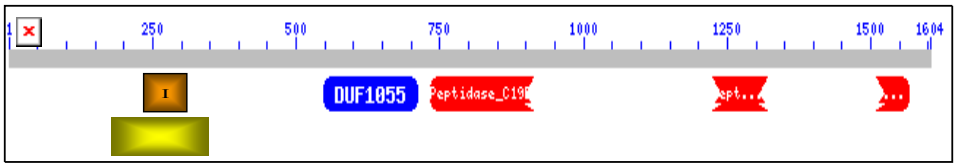


Figure 3.1.A. The domains of USP6. USP6 protein consists of 1,406 amino acids (1-1,406) and the locations of conserved domains are indicated along the amino acid sequence. I) TBC domain with suggested GTP-activator activity on Rab-like GTPases. II) Cd02674_C19E peptidase. They remove ubiquitin molecules from the polyubiquitinated molecules by cleavage of isopeptide bonds (also known as cysteine domain), III) Cd02674_C19R domain, which cleaves the isopeptide bonds. IV) Cd02674_C19R with ubiquitinyl hydrolysis activity (the conserved histidine domain). Data was taken from; NCBI- conserved domain databases).

B. The domains of USP32. I) Cd00051, EFh, EF-hand, calcium binding motif. Binding of $Ca^{2+,}$ causes a conformational change in this motif, leading to activation or inactivation of the target protein. II) COG5126, FRQ1, Ca^{2+}-binding protein which is predicted to have general functions. III) Pfam06337, DUF 1055, found in Ubiquitin Specific Proteases with an unknown function. VI) Cd02661, peptidase _C19E, which cleaves the isopeptide bonds and deubiquitinating the polyubiquitinated molecules. V) Cd02674_C19R, which cleaves the isopeptide bonds. VI) Cd02674_C19R with ubiquitinyl hydrolysis activity. These proteolysis domains are important in editing and recycling of ubiquitin (Taken from NCBI conserved domain databases).

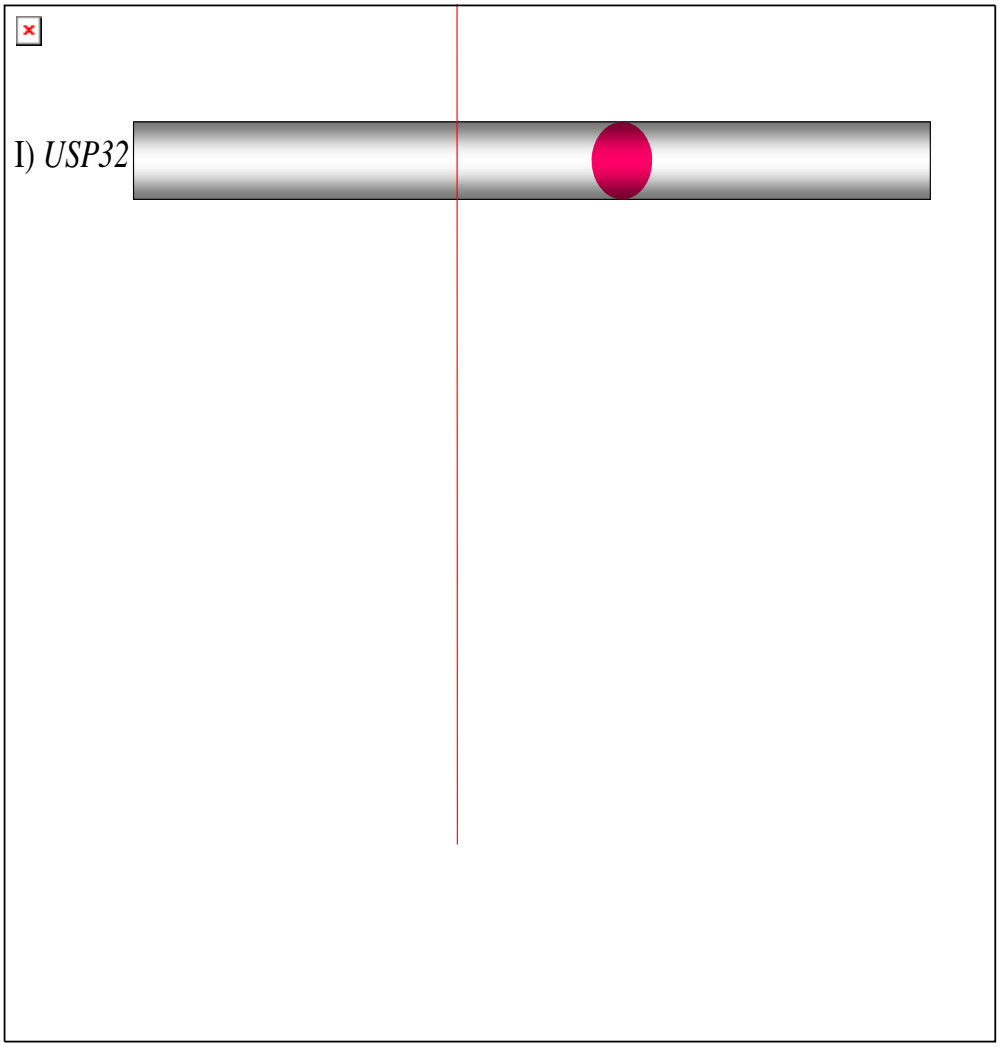


Figure 3.2. The transcript sequence similarities between *USP32* and *USP6* (and its different transcripts). I) *USP32* mRNA, II) *USP6* mRNA, III) Transcript 213 of *USP6* gene with its two open reading frames, IV) Transcript 210 of *USP6* gene with its two open reading frames. The sequence similarity between these two genes is between 3,197th to 7,831st bases of *USP6* and 2,390th to 7,024th bases of *USP32* gene

3.2. The Expression Pattern of *USP6* Gene's Different Transcripts

There is 97% sequence similarity between *USP6* and *USP32* genes. The conserved histidine and cysteine domains of *USP6* gene which are responsible for the deubiquitinating activity are also found in the *USP32* gene. Investigating the expression pattern of different transcripts of *USP6* gene (NM_004505, X63547 and X63546) and also their possible ORFs (210-I, 210-II, 213-I or 213-II), can help us understand more about these highly similar genes and may lead us to perform further functional assays to find the exact role of these genes in cancer cells.

The fact that *USP6* and *USP32* are highly similar genes (although on different chromosomes), we thought that these two proteins may have overlapping functions. Therefore, before focusing on *USP32* as an overexpressed oncogene candidate, we wanted to differentiate the expression patterns of these two transcripts.

To understand which transcripts of *USP6* gene are transcribed, we designed specific primers which were capable of binding to each transcript specifically but not to *USP32*. By using the NCBI blast program, different transcripts of *USP6* gene and the position of primers are shown in Figure 3.3.

The expected sizes from PCR products, using different primers on different *USP6* transcripts, are shown on (Table 3.1).

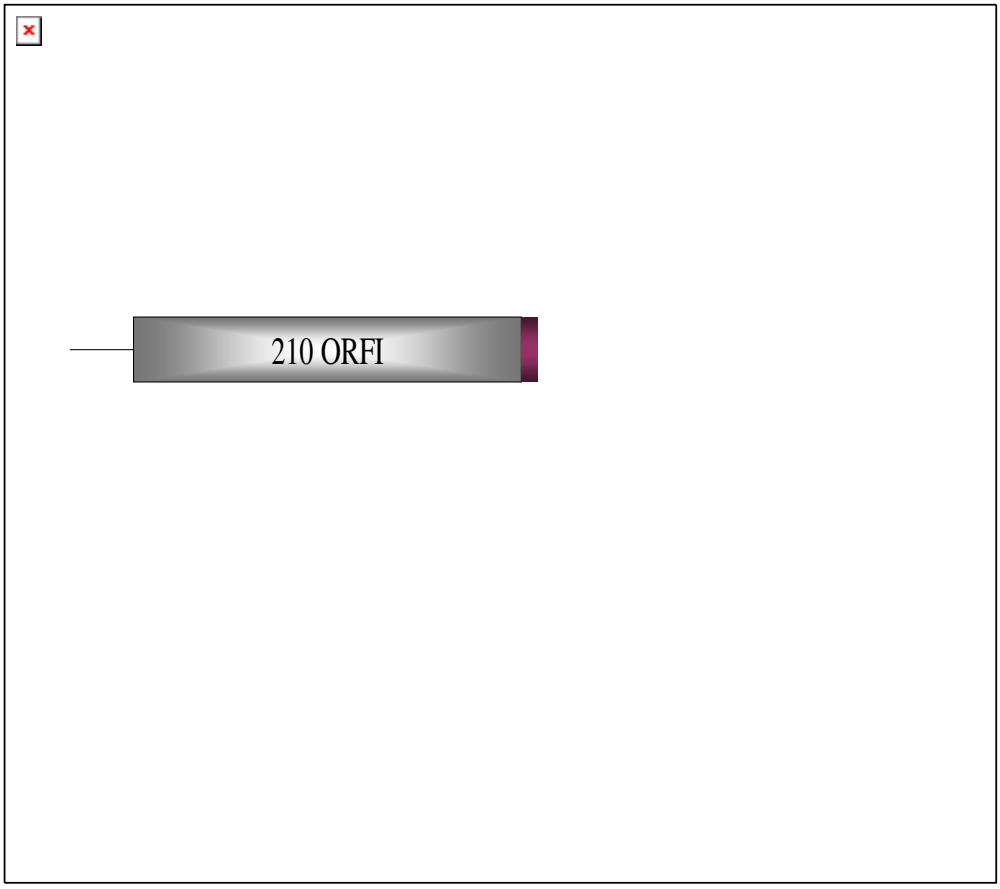
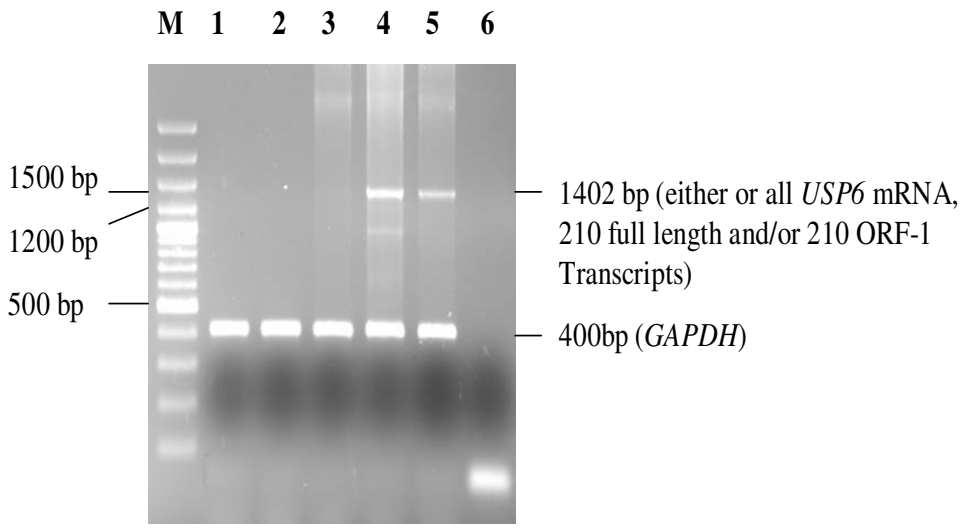


Figure 3.3. The different possible transcripts of *USP6* gene and the location of primers on them. The sequence of transcripts was taken from NCBI and the primers were designed using primer-3 program. Blast program of NCBI was also used to check if primers were specific enough for each transcript.

Table 3.1. The expected sizes from PCR products of different transcripts of *USP6* gene, when using different primers.

	213-F and R(S)	**2F-CDS & 2R-CDS**	**2F(B) & 2R(B)**	**1F(B) & R(S)**	**1F(B) & 1R(B)**
210-full	-	-	692 bp	1402 bp	-
210-ORF-I	-	-	-	1402 bp	-
210-ORF-II	-	-	-	-	-
***USP6*mRNA**	-	-	804 bp	1402 bp	-
***USP32*-mRNA**	-	600 bp	804 bp	-	-
213-full	704 bp	-	804 bp	1613 bp	790 bp
213-ORF- I	-	-	-	-	790 bp
213-ORF-II	704 bp	-	804 bp	-	-

The first primer set, 1F (B) & R (s), was used to show the expression of *USP6* mRNA in MCF7 and MDA-MB-231 breast cancer cell lines, and normal breast, testis and ovary tissue. These primers amplified a 1,402 bp long fragment of transcript belonging to either or all of the 210-full length, 210-ORFI and *USP6*-full mRNA and a 1,613 bp long region of transcript 213-full length. So in the case that we observed a 1,402 bp long band, we concluded the expression of either one of the 210-full length, 210-ORFI and/or *USP6* transcripts or all of them. While the existence of the 1,613 bp band would show the expression of the transcript 213, absence of this band suggested the lack of this transcript in the MCF7and MDA-MB-231 breast cancer cell lines and normal breast, testis and ovary cells. A set of primers for amplification of a 400 bp long fragment of the *GAPDH* gene was also used in the same PCR reaction (Figure 3.4.).

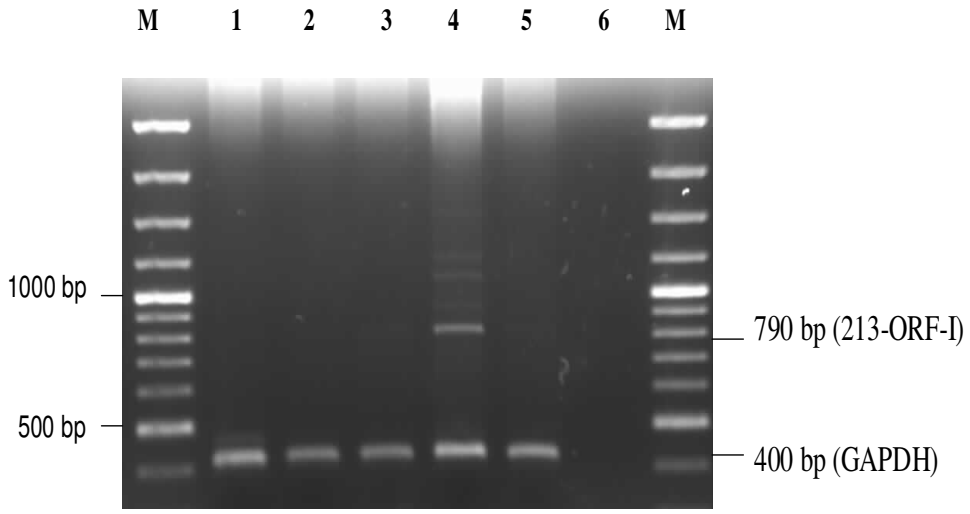


Expected sizes; 1,402 bp band for *USP6* and/or transcript 210 (210 full length and/or 210 ORF-I). 1,613 bp band for transcript 213 if it existed.

Figure 3.4. Duplex-PCR using 1F (B) & R (s) and *GAPDH* primers, which showed the expression of *USP6* mRNA, 210 full length and/or 210 ORF-1 transcripts. M): 100 bp ladder was used as the DNA marker. 1) MCF7 cDNA, 2) MDA-MB-231 cDNA, 3) Normal Breast cDNA, 4) Normal Testis cDNA, 5) Normal Ovary cDNA, 6) Blank (no cDNA).

This gel photo showed the amplification of *GAPDH* gene and another band which was 1,402 bp and indicated the existence of either one or more of the 210-full length, 210-ORFI and/or *USP6*- full mRNA transcripts in both testis and ovary tissues. As there was no band of 1,613 bp, it was assumed that there was no expression of 213-full length transcript in the tested cell lines and tissues. Therefore, we concluded that 213-full length transcript didn't exist in MCF7, MDA-MB-231 breast cancer cell lines and breast, testis and ovary tissues.

The second primer set, (1F (B) and 1R (B)), was used in an RT-PCR reaction to show the expression of transcripts; 213-full length or 213-ORFI (Figure 3.5). The expected area to be amplified by these primers was 790 bp. However, based on the Figure 3.4, the transcript 213-full length was absent in the tested cell lines and tissues. In this RT-PCR reaction also the cDNA from MCF7 and MDA-MB-231 cell lines and normal breast, testis and ovary tissue were used. The *GAPDH* primers were used and a region of 400 bp from GAPDH cDNA was amplified, using these primers.

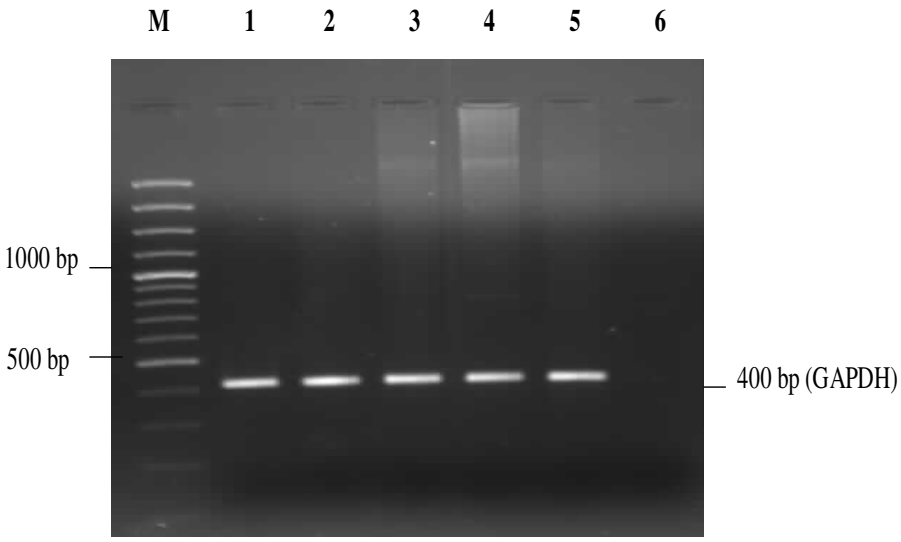


Expected size; 790 bp for 213-ORFI transcript, 400 bp for *GAPDH.*

Figure 3.5. RT-PCR which showed the expression of *USP6*-clone 213-ORFI, using 1F &1R and *GAPDH* primers. M) 100 bp DNA marker. 1) MCF7 cDNA, 2) MDA-MB-231 cDNA, 3) Normal Breast cDNA, 4) Normal Testis cDNA, 5) Normal Ovary cDNA, 6) Blank (no cDNA) and M) 100bp DNA marker. The *GAPDH* band was also amplified (400 bp).

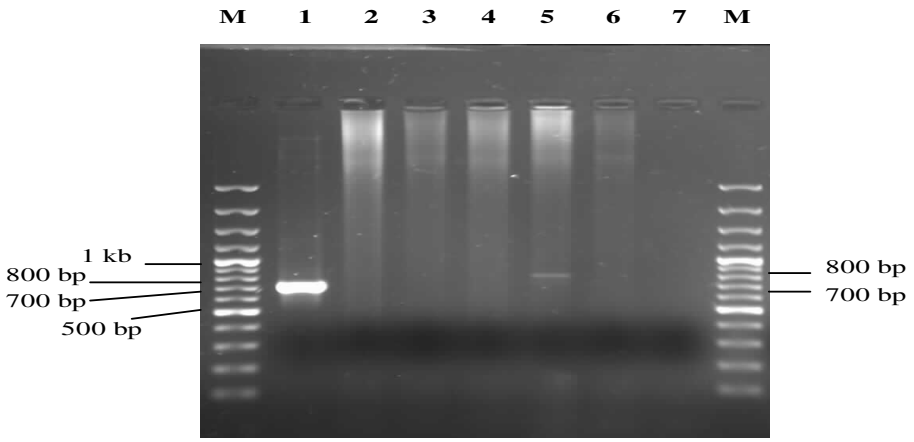
According to the result of this reaction, the expression of 213-ORFI transcript was found in the testis tissue as the 790 bp band was just detected in the testis tissue because the presence of 213-full transcript was excluded based on the previous PCR reaction in Figure 3.4.

In order to investigate the expression pattern of 213-ORFII transcript, two primers were designed which were specific just for transcript-213. As there is a base insertion area specific for this transcript, the forward primer (213-F) was designed inside the insertion area and also inside the second open reading frame of the 213 transcript. These primers also gave a hit for the full-213 transcript but the absence of this transcript was found from the result of the Figure 3.4. The result of the duplex PCR in which primers for the *GAPDH* gene were also used, indicated the expression pattern of transcript 213-ORFII in the MCF7 and MDA-MB-231 breast cancer cell lines and normal breast, testis and ovary tissues. No expression of this transcript was detected from this set up (Figure 3.6). We also performed another PCR set up using a control plasmid which contained the cDNA of 213-full transcript to make sure that our primers worked. Also this PCR was carried out in the absence of *GAPDH* primers and both 213-F and R(S) primers were used. The result of this PCR is shown in Figure 3.7. The results of these PCRs showed no expression of 213-ORFII transcript in the MCF7 and MDA-MB-231 breast cancer cell lines and normal breast, testis and ovary tissues as we didn't detect the 704 bp expected band in none of the used cell lines and tissues. In addition an 800 bp band was detected in testis tissue which may show the existence of another isoform of the gene (Figure 3.7).



Expected size; 704 bp for 213- ORFII transcript, 400 bp for *GAPDH*

Figure 3.6. RT-PCR detected no expression of *USP6*-clone 213-ORFII, using 213-F & R(S) and *GAPDH* primers. M) 100 bp DNA marker, 1) MCF7 cDNA, 2) MDA-MB-231 cDNA, 3) Breast cDNA, 4) Testis cDNA, 5) Ovary cDNA, 6) Blank (no cDNA). The *GAPDH* band was also amplified (400 bp).



Expected size; 704 bp for 213- ORFII transcript

Figure 3.7. RT-PCR detected no expression of *USP6*-clone 213-ORFII, using 213-F & R(S) primers. M) 100 bp DNA marker, 1) Control plasmid containing the 213-full transcript, 2)MCF7 cDNA, 3) MDA-MB-231 cDNA, 4) Breast cDNA, 5) Testis cDNA, 6) Ovary cDNA, 7) Blank (no cDNA).

The summary of PCR results, using different primers, on different cell lines and tissues are shown in Table 3.2. The results showed the expression of different transcripts in the mentioned cell types. These results suggested that, *USP6*-mRNA, 213-full and 213-ORF-II were not expressed in MCF7, MDA- 231 cell lines and breast tissue. 213 ORF-II also was not expressed in testis and ovary tissue. 213 ORF-I, was detected to be expressed just in testis. A 1,400 bp band was also detected in both testis and ovary, which suggested the expression of 210 or *USP6*-mRNA transcripts in these tissues. As there was no way to design specific primers to separate 210- full transcript from 210-ORFI and 210-ORF-II, it was not possible to distinguish these two transcripts from each other.

Table 3.2. The expression of *USP32* and also different transcripts of *USP6* gene in different cell lines and tissues. N.D; Not Detected, -; No expression and +; expression.

	Ovary	**Testis**	**Breast**	**MDA-231**	**MCF7**	**Primer pair**
210-full	N.D	N.D	N.D	N.D	N.D	2F(B) 2R(B)
210-ORF- I	N.D	N.D	N.D	N.D	N.D	-
210-ORF-II	N.D	N.D	N.D	N.D	N.D	-
***USP6*-mRNA**	N.D	N.D	-	-	-	1F(B) and R(S)
***USP32*- mRNA**	+	+	+	+	+	2F-CDS 2R-CDS
213-full	-	-	-	-	-	1F(B) R(S)
213-ORF-I	-	+	-	-	-	1F(B) and1R(B)
213-ORF-II	-	-	-	-	-	213-F and R(S)

3.3. *USP32* Gene and its Expression Pattern

USP32 gene, which is approximately 212 kb long, is predicted to have 34 exons according to mRNA to genomic sequence alignment tools (NM-32582 vs.NT-010783). By using Spidey program, the location and length of the exons were found and the sizes of the introns were also calculated. We also confirmed the introns and exon boundaries of this gene, by PCR and sequencing. These data is shown on Table 3.3.

Table 3.3. The length of the exons and introns of *USP32* gene and also the position of the exons on genomic DNA and mRNA.

Exon #	Exon size (bp)	Genomic coordinates	mRNA coordinates	Intron #	Intron size (bp)
1	344	17122304-17122647	1-344	1	46274
2	128	17075903-17076030	345-472	2	43777
3	106	17032021-17032126	473-578	3	6798
4	119	17025105-17025223	579-697	4	6001
5	160	17018945-17019104	698-857	5	17042
6	132	17001772-17001903	858-989	6	1793
7	108	16999872-16999979	990-1097	7	3359
8	116	16996398-16996513	1098-1213	8	502
9	63	16995834-16995896	1214-1276	9	10154

Table 3.3. The length of the exons and introns of *USP32* gene and also the position of the exons on genomic DNA and mRNA (cont'd).

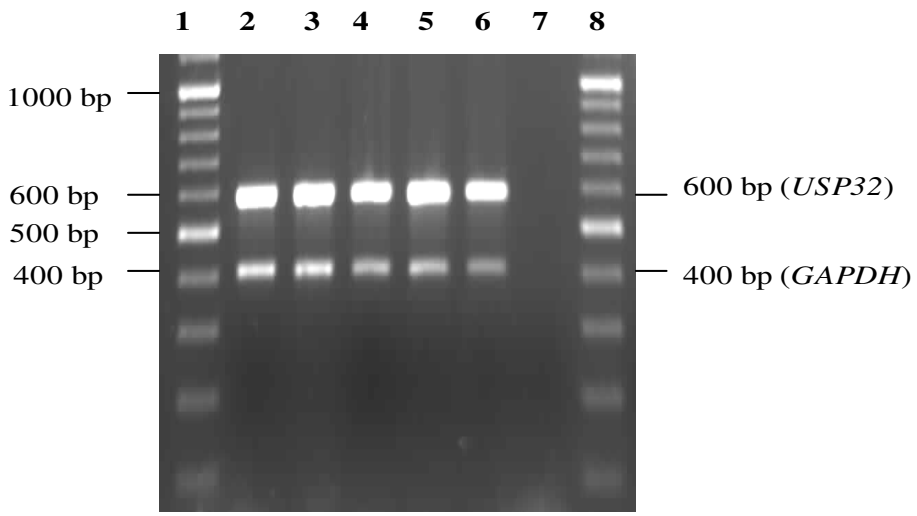
10	84	16985597-16985680	1277-1360	10	2733
11	62	16982803-16982864	1361-1422	11	16141
12	103	16966560-16966662	1423-1525	12	9907
13	193	16956461-16956653	1526-1718	13	2453
14	176	16953833-16954008	1719-1894	14	862
15	141	16952831-16952971	1895-2035	15	2622
16	118	16950092-16950209	2036-2153	16	4896
17	155	16945042-16945196	2154-2308	17	958
18	82	16944003-16944084	2309-2390	18	1483
19	75	16942446-16942520	2391-2465	19	510
20	139	16941798-16941936	2466-2604	20	258
21	106	16941435-16941540	2605-2710	21	1470
22	174	16939792-16939965	2711-2884	22	542
23	175	16939076-16939250	2885-3059	23	444
24	152	16938481-16938632	3060-3211	24	927
25	112	16937443-16937554	3212-3323	25	1363
26	212	16935869-16936080	3324-3535	26	7003
27	185	16928682-16928866	3536-3720	27	5330
28	87	16923266-16923352	3721-3807	28	2161

Table 3.3. The length of the exons and introns of *USP32* gene and also the position of the exons on genomic DNA and mRNA (cont'd).

29	121	16920985-16921105	3808-3928	29	4912
30	192	16915882-16916073	3929-4120	30	2007
31	289	16913587-16913875	4121-4409	31	1417
32	425	16911746-16912170	4410-4834	32	687
33	93	16910967-16911059	4835-4927	33	1117
34	2099	16907752-16909850	4928-7026		

USP32 gene, located on chromosomal band 17q23, within a known area of amplification, found in breast cancer, has also been shown to be overexpressed in a number of breast cancer cells. Because the protein sequence of this gene suggests a deubiquitinating function for it and deubiquitinating enzymes play important roles in the reversal pathways of protein degradation (especially in the case of key regulatory proteins), the expression pattern of this gene can reveal information about its involvement in the breast tumorigenesis.

To confirm the expression of *USP32,* we performed RT-PCRs with MCF7 and MDA-MB- 231 breast cancer cell lines and normal breast (Ambion), testis (Ambion), and ovary (Ambion) tissues. Figure 3.8, shows the duplex PCR, indicating the expression of *USP32* gene in these cells. In this reaction, 2F-CDS and 2R-CDS primers in the combination with *GAPDH* primers were used for amplifying a 600 bp long fragment of the *USP32* gene and a 400 bp long fragment from the *GAPDH* gene. All the used cell lines and tissues, used in this experiment showed the expression of *USP32* gene.



Expected sizes; 600 bp for *USP32* gene and 400 bp for the *GAPDH* gene.

Figure 3.8. The expression of *USP32* in; MCF7 cDNA (lane 2), MDA-MB-231 cDNA (lane 3), Normal Breast cDNA (lane 4), Normal Testis cDNA (lane 5) and Normal Ovary cDNA (lane 6). Lane 7 is the negative control without DNA. 100 bp DNA marker was loaded on lanes 1 and 8. 2F-CDS and 2R-CDS primers were used for the amplification of the *USP32* gene.

3.4. Cloning of *USP32* gene

The *USP32* gene which codes for a 1,604 a.a long protein is the gene of our interest. To make it easier to find the functional roles of this gene inside the cell, we cloned the gene into the TOPO-TA cloning vector. Because the size of this gene was big, we divided the gene into 3 parts by designing specific PCR primers. According to the location of different domains of the gene, different fragments which have to be cloned into the vector were separated by PCR

amplification The pEGFP-N1, which has an enhanced GFP gene, followed by the multiple cloning site (in which our gene is pasted), is the suitable vector for further localization investigations of the TOPO-TA cloned constructs. This vector causes a green fluorescent signal which shows the location of the GFP-tagged protein in the cell. In addition, for appropriate expression of the gene, the primers were designed in a way that doesn't alter the open reading frame of the pEGFPN1 vector. For this purpose extra nucleotides before the start codon of primers were added if necessary. Also the stop codon of the *USP32* gene was removed by excluding this sequence from the reverse primers because we desired the co-expression of the inserted gene (inside the multiple cloning site of the vector) with the *GFP* gene (inside the pEGFP-N1 vector) which came after the multiple cloning site of the vector.

Using these specific primers, the desired fragments of the gene were PCR amplified. We digested the PCR amplified fragments, which contained both *Apa1* and *Xho1* recognition and restriction sites, with *Apa1* and *Xho1* restriction enzymes. Each time that we tried to clone these DNA fragments into the double digested pEGFP-N1 vector (digested with the same restriction enzymes used for the insert DNA), we got no result and the cloning was not successful. For this reason we decided to sub clone this PCR products into the TOPO-TA cloning vector. There were some advantages, using TOPO-TA cloning vector such as the possibility of sequencing the PCR products, before cloning into the pEGFP-N1 vector to avoid any PCR derived mutation. Also digestion of the end of PCR products which contained recognition and restriction sites for *Xho1* and *Apa1* restriction enzymes, using these two enzymes, was not efficient without cloning the products into the TOPO-TA cloning vector. Also the coding sequence of the *USP32* gene was 4.8 kb and the vector was 4.7 kb long in size and cloning of the full length gene caused difficulties. For this purpose, a number of additional primers were designed to divide the gene into three parts to be separately cloned into the pEGFP-N1 vector. In addition to the full length coding sequence of the

USP32 gene, these partial fragments of the gene, which were 2.2 kb, 2.6 kb and 3.2 kb long, were also cloned into the TOPO-TA vector. The location of these primers on different fragments of the *USP32* gene is shown in Figure 3.9.

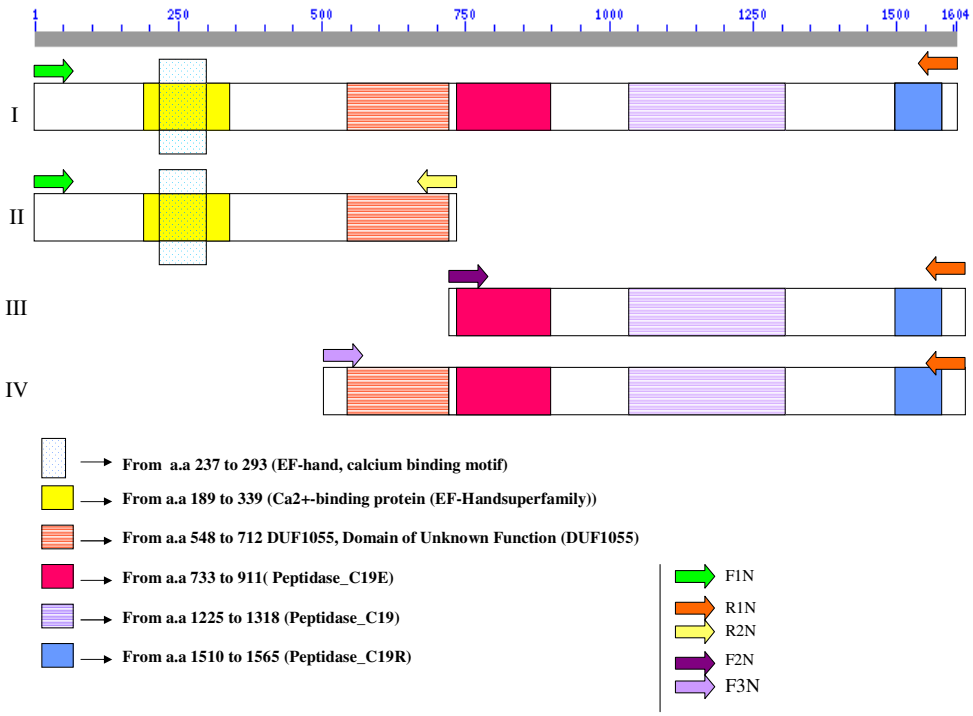
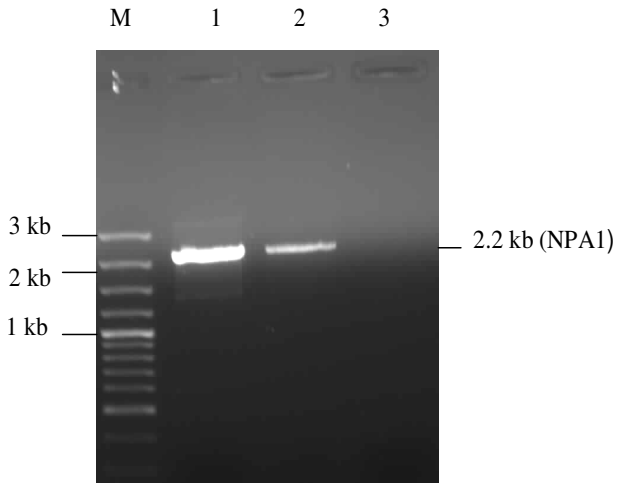


Figure 3.9. Map of the position of primers on *USP32*. The full length cDNA (I), *USP32*'s first partial fragment; NPA1, from nucleotide 287 to 2,484 of the cDNA(II), second partial fragment; NPA2 from nucleotide 2,470 to 5,098 (III) and third partial fragment; NPA3 from nucleotide 1,866 to 5,098 of the cDNA (IV). The numbers from 1 to 1,604 indicate the amino acid sequence of the USP32 protein. The locations of the domains are taken from NCBI- conserved domains information (www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?INPUT_TYPE= p recalc&SEQUENCE=22550104). Numbers show the amino acid sequence.

3.4.1. PCR Products of *USP32* Gene, Used for TOPO-Cloning Purpose

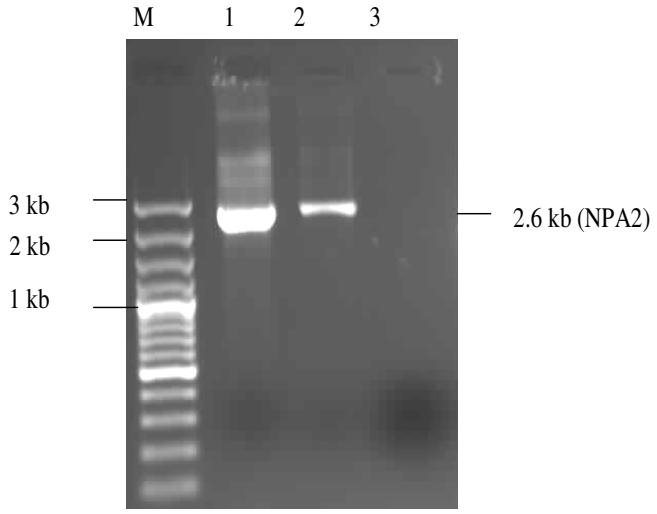
The first part of the *USP32* gene (NPA1), from the nucleotide 287 to 2,484 of the cDNA, was PCR amplified from the template DNA, using Expand High Fidelity Enzyme and the size of the amplified fragment was 2.2 kb. The primers, used in this PCR reaction (F1N and R2N), contained the *Apa1* and *Xho1* restriction sites. The cDNA of *USP32* gene which was cloned earlier into pTARGET vector was used as template DNA in this experiment (Figure 3.10). This PCR product was ligated into TOPO-TA cloning vector.



Expected size; 2.2 kb.

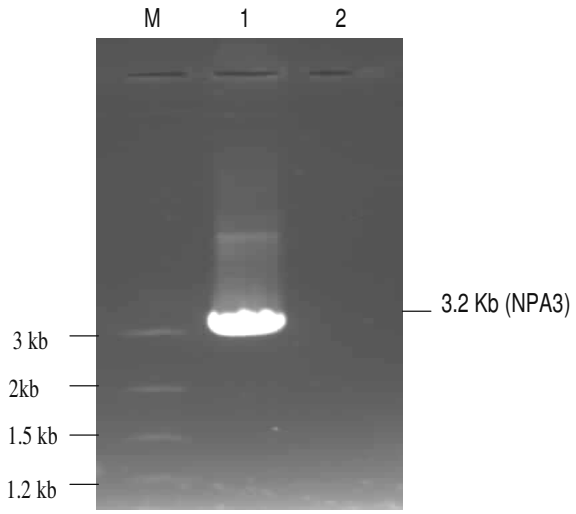
Figure 3.10. The PCR product (NPA1), Using F1N and R2N primers. M) 100 bp marker, 1) 10 µl of PCR product to check the PCR quality, 2) 2 µl of PCR product to confirm the exact size, 3) the negative PCR control without template DNA.

The second part of the *USP32* gene (NPA2) from nucleotide 2,470 to 5,098 of the cDNA, which was 2.6 kb long and an extra 3.2 kb long fragment (NPA3), from nucleotide 1,866 to 5,098 of the cDNA (shown in Figure 3.8) were also PCR amplified, using specific primers, containing *Apa1* (in reverse primer) and *Xho1* (in forward primer) restriction sites (Figure 3.11 and Figure 3.12). Expand High Fidelity Enzyme was used as the polymerase and the cDNA of *USP32* in pTARGET vector was used as the PCR template in these PCR reactions.



Expected size; 2.6 kb

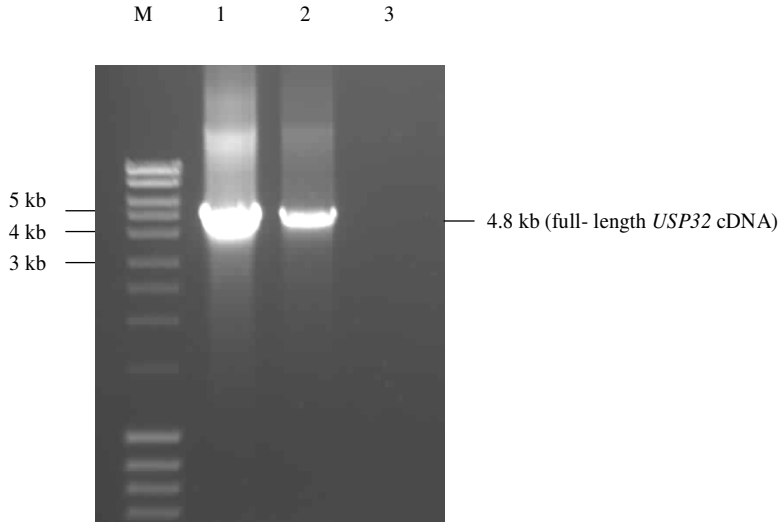
Figure 3.11. The 2.6 kb long PCR product, for the second part of the gene (NPA2), using F2N and R1N primers. M) 100 bp DNA marker, 1) 10 µl of PCR product to check the PCR quality, 2) 2 µl of PCR product to confirm the exact size of the product, 3) the negative PCR control without template DNA.



Expected size; 3.2 kb.

Figure 3.12. The 3.2 kb long PCR product, Using F3N and R1N primers. M) 100 bp DNA marker, 1) 10 µl of PCR product, 2) negative control with no template DNA.

The full length cDNA of *USP32* gene also was PCR amplified using primers which contain *Apa1* and *Xho1* recognition and restriction sites, Expand High Fidelity enzyme and *USP32* cDNA inside the pTARGET vector. The size of the amplified fragment was 4.8 kb (Figure 3.13).



Expected size; 4.8 kb.

Figure 3.13. The PCR product of the full length *USP32* cDNA which is 4.8 kb long. The F1N and R1N primers were used in this PCR. M) Mass ruler DNA marker, 1) 10 µl of PCR product, 2) 2 µl of PCR product for the size confirmation, 3) negative control without DNA.

3.4.2. Confirmation of the TOPO-TA- Cloned Constructs with Restriction Enzymes

The PCR products which had to be cloned into the TOPO-TA cloning vector were run on the gel for the confirmation of their sizes and also to check the PCR quality. *Taq* DNA polymerase was added to the correct PCR products for the purpose of adding adenine to the 3' end of the products. These PCR products were cloned into the TOPO-TA and the cloned constructs were transformed into the chemically competent *E. coli* cells. The transformed cells were grown on the LB-

Agar medium, containing spectinomycin for the selection of the transformed bacterial cells. A number of colonies, grown on LB- spectinomycin plates, were inoculated into the LB-broth medium (with spectinomycin) and the plasmids of these cells were purified. The plasmids were digested with *Apa1*, *Xho1* and *EcoR1* restriction enzymes. *Apa1* and *Xho1* restriction enzymes also digest the TOPO-TA vector. The expected sizes from restriction digestion of the cloned constructs are shown in Table 3.4.

Table 3.4. The expected results for restriction digestion of *USP32*- TOPO cloned constructs. NPA1 (the first fragment of the *USP32* gene, 2.2 kb, which was cloned into TOPO vector), NPA2 (the second fragment of the *USP32* gene, 2.6 kb, which is cloned into TOPO vector), NPA3 (a 3.2 kb long fragment of the *USP32* gene inside TOPO vector) and the full length cDNA of the *USP32* gene. All these fragments were cloned into TOPO-TA cloning vector.

	EcoR1	***Xho1***	***Apa1***	***Apa1* and *Xho1***
NPA1	2.2 kb+2.8 kb	3.6 kb + 1.4 kb	2.3 kb +2.7 kb	2.2 kb + 1388 bp+ 104 bp +1307bp
NPA2	2.4 kb+2.8 kb	4 kb + 1414 bp	2.7 kb+2.7 kb	2.6 kb + 1388 bp + 104 bp +1307bp
NPA3	3 kb+ 2.8 kb	4588 bp+ 1414 bp	3.6 kb+ 2.7 kb	3.2 kb + 1388 bp + 104 bp +1307bp
FULL-LENGTH	4.6kb+ 2.8 kb	1.4 kb+ 6.2 kb	4.9 kb+ 2.7 kb	4.8kb+1388 bp+ 104 bp +1307 bp

3.4.2.1. Restriction Digestion of *USP32* First Fragment (2.2 kb) inside TOPO-TA Vector (NPA1).

This fragment of *USP32* cDNA, which was 2.2 kb long, was sub cloned into TOPO-TA cloning vector and a number of constructs were digested by *Apa1*, *Xho1* and *EcoR1* restriction enzymes. Among these constructs, NPA1- 102, NPA1- 111, NPA1- 114 and NPA1- 118 were chosen to be sequenced. The expected sizes from restriction digestions are mentioned in Table 3.4.

Apa1 and *Xho1* digest the insert cDNA just at the 3' and 5'ends of the insert respectively. These two enzymes also digest the TOPO-TA vector as follows; *Apa1* at nucleotide 568 and *Xho1* at nucleotide 2,078. *EcoR1* enzyme digests the TOPO-TA vector at nucleotides; 672 and 690 and there is no restriction site inside the insert sequence for *EcoR1*. Figure 3.14 shows the expected sizes for NPA1 construct, when digesting with these three enzymes. The correct colonies with the same digestion results were sent for sequencing.

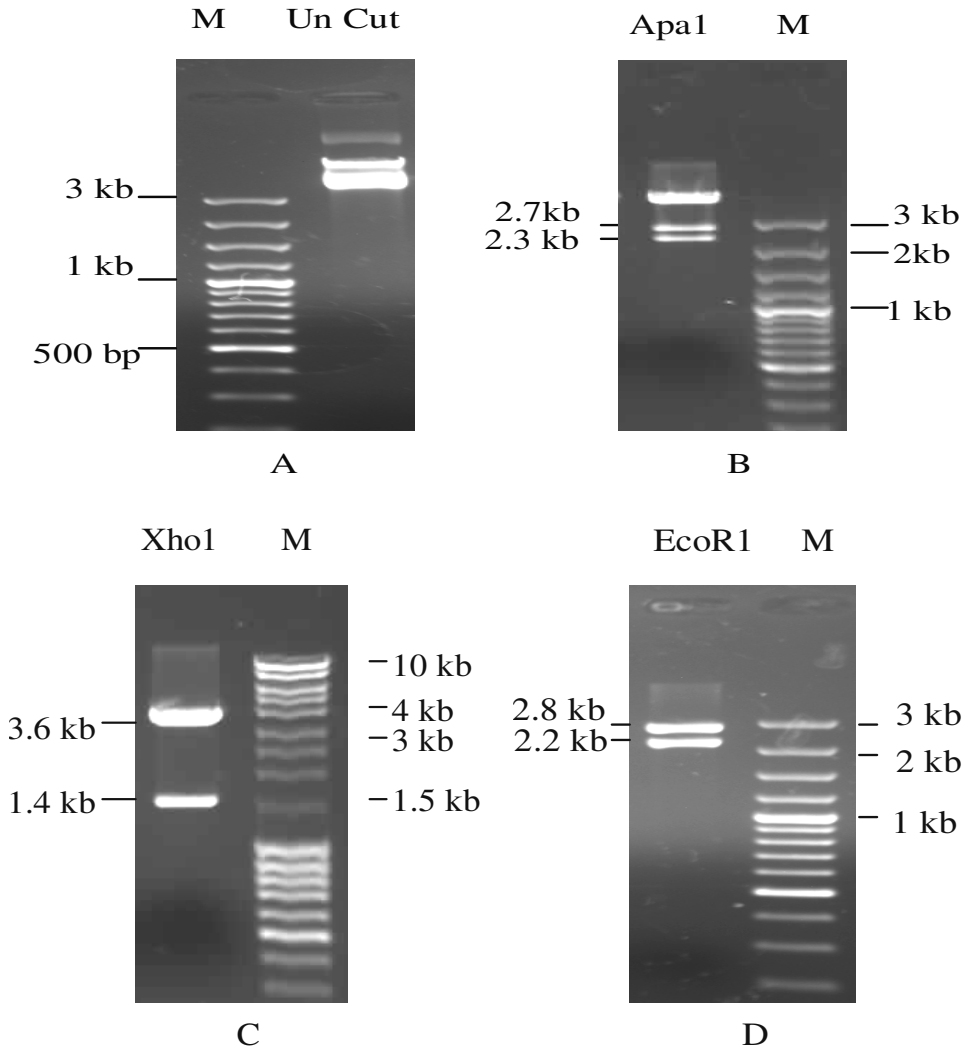


Figure 3.14. Expected results from restriction digestion of NPA1 (the first fragment of *USP32* gene, 2.2 kb) with *Apa1*, *Xho1* and *EcoR1* enzymes. A) The uncut NPA1 (2.2 kb insert inside the TOPO-TA vector). M; 100 bp DNA marker. B) The *Apa1* digested NPA1 in which two 2.3 and 2.7 kb bands are expected. A portion of uncut vector is also remained in this digestion (the upper band, larger than 3 kb). M; 100 bp DNA marker. C) *Xho1* digestion of NPA1 with expected 3.6 and 1.4 kb bands. M; Mass ruler marker. D) *EcoR1* digestion of NPA1 resulting in two 2.2kb (insert) and 2.8 kb (vector) bands. M; 100 bp DNA marker.

3.4.2.2. Restriction Digestion of *USP32* Second Fragment (2.6 kb) inside TOPO-TA Vector (NPA2)

The NPA2 fragment which consisted of a 2.6 kb *USP32* cDNA inside TOPO-TA vector was digested with *Apa1*, *Xho1* and *EcoR1* restriction enzymes. *Apa1* and *Xho1* digest the 2.6 kb insert once at 3' and 5' ends respectively. *Apa1* digests the TOPO-TA vector at nucleotide 568 and *Xho1* digests the vector at nucleotide 2,078. *EcoR1* enzyme digests the insert at nucleotide 4,559 of full length cDNA and also digests the TOPO vector at positions; 672 and 690 (nucleotide). The expected sizes from the digestion of these three enzymes are mentioned in Table 3.4 and the desired result of restriction digest products are shown in Figure 3.15.

From digested constructs, those which had both *Apa1* and *Xho1* restriction sites and showed the same digested results as in Figure 3.15, were the candidates to be sequenced. Among the correct candidate constructs, NPA2- G, NPA2- B, NPA2- 203, NPA2- 206 and NPA2- 212 were sent for sequencing.

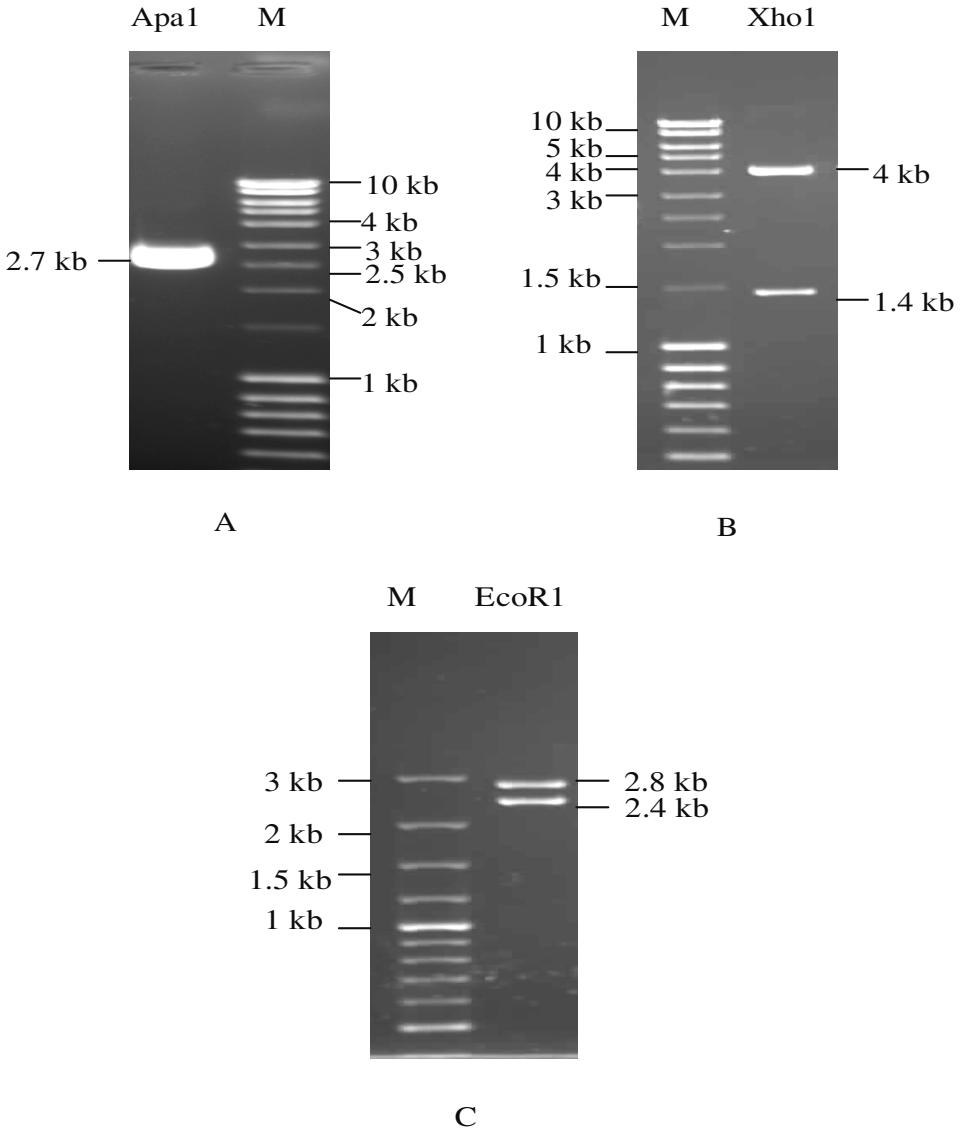


Figure 3.15. The desired results from restriction digestion of NPA2 construct (2.6 kb insert inside the TOPO vector). A) *Apa1* digestion of NPA2 which resulted in two 2.7 kb bands. M; Mass Ruler marker. B) *Xho1* digestion of NPA2 construct in which two 4 kb and 1.4 kb bands were expected. M; Mass Ruler marker. C) *EcoR1* digestion of NPA2 with expected 2.4 and 2.8 kb bands. M; 100 bp marker.

3.4.2.3. Restriction Digestion of *USP32* Second Fragment (3.3 kb) inside TOPO-TA Vector (NPA3)

NPA3, the third fragment of *USP32* gene, which was also cloned into TOPO vector, was digested with three restriction enzymes; *Apa1, Xho1* and *EcoR1*. *EcoR1* enzyme digests the insert (at nucleotide 4,559 of gene's full length cDNA) and the vector at nucleotides 672 and 690. *Apa1* enzyme digests the insert at the 3' end and the vector at nucleotide 568, while *Xho1* digests the insert at the 5' end and the vector at nucleotide 2,078. The expected sizes from restriction digestions are in Table 3.4. The NPA3 restriction results are shown in Figure 3.16.

From the correct constructs which contained both *Apa1* and *Xho1* restriction sites and had the same restriction digestion results were chosen and among them, NPA3-7, NPA3-10 and NPA3-14 were sent for sequencing.

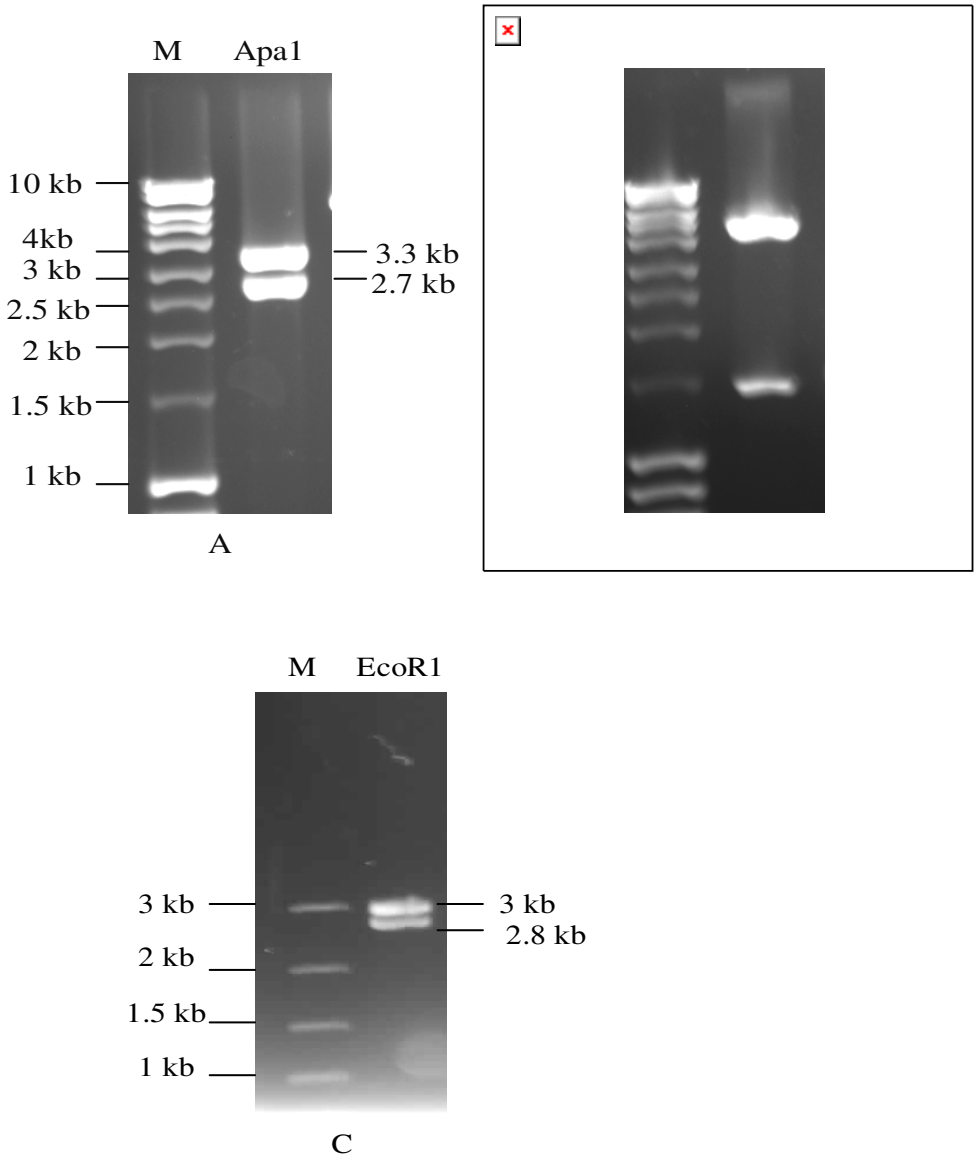


Figure 3.16. The expected results from restriction digestion of NPA3 (3.2 kb insert inside the TOPO vector). A) *Apa1* digestion of NPA3, resulting in two 3.3 kb and 2.7 kb bands. M; Mass Ruler marker. B) *Xho1* digestion of NPA3 constructs which results in two 4.6 kb and 1.4 kb bands. M; Mass Ruler marker. C) *EcoR1* digestion of NPA3 with expected 3 and 2.8 kb bands. M; 100 bp marker.

3.4.2.4. Restriction Digestion of *USP32* Full Length (4.8 kb) inside TOPO- TA Vector (N- Full)

The full length *USP32* gene (4.8 kb) was also cloned into the TOPO-TA vector and digested with restriction enzymes *Apa1* and *Xho1*. These two enzymes digested the insert at the 3' and 5' of the strand and digested the vector as follows; *Apa1* digests the TOPO-TA vector at nucleotide 568 and *Xho1* at nucleotide 2,078. The sizes which we got from these restriction reactions are available in Table 3.4.

Among the correct N- Full constructs, which contain both *Apa1* and *Xho1* restriction sites, and indicated the same restriction digestion results as in Figure 3.16, (N- Full-4, N- Full-13, N- Full-17 and N- Full-18 constructs) were sent to McLab (San Francisco, USA) to be sequenced. Figure 3.17, shows the size of the digested N- Full constructs, using *Apa1* and *Xho1* enzymes.

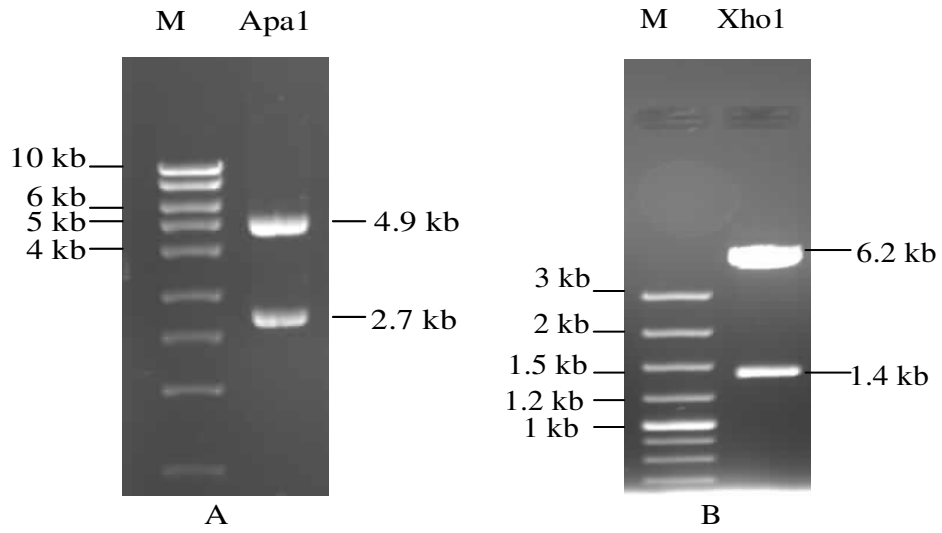


Figure 3.17. The expected results from *Apa1* and *Xho1* digestions of N- Full construct (*USP32* full length cDNA inside TOPO-TA vector). A) *Apa1* digestion of N-Full, resulting in two 4.9 and 2.7 kb bands. M; Mass Ruler DNA marker. B) *Xho1* digestion of N- Full construct with expecting two 6.2 and 1.4 kb bands. M; 100 bp marker.

3.4.3. Sequencing Results of *USP32* Gene inside TOPO-TA Cloning Vector

The cloned constructs (*USP32* gene into TOPO-TA vector) which their sizes were confirmed, using restriction enzymes, were sent for sequencing (Iontek Company; Istanbul, Turkey and McLab, San Francisco, USA) to avoid using constructs containing PCR- created mutations which could be missense or nonsense. The PCR products were cloned into PCR® 8/GW/TOPO vector as mentioned in part 3.4. The correct cloned constructs which contained both *Xho1* and *Apa1* restriction sites were sent for sequencing. The primers by which the constructs were sequenced are shown in Table 3.5.The sequence results are shown in Tables 3.6, 3.7 and 3.8.

Table 3.5. The primers, used for sequencing of TOPO- cloned constructs. * shows the primers which were used for sequencing of each construct.

	Primers													
Samples ▼	GW1	1F	2F	3F	4F	5F	5R	6F	6R	7F	8F	9F	9R	GW2
NPA1	*	*	*	*	*									*
NPA2	*							*	*	*	*	*	*	*
NPA3	*				*		*	*		*	*	*	*	*
N-FULL	*	*	*	*	*	*			*	*	*	*	*	*

Table 3.6. The sequence results of NPA1 (the first fragment of the *USP32* gene, from nucleotide 287 to 2,484 of the cDNA, which is 2.2 kb long), NPA2 (the second partial fragment of the gene, from nucleotide 2,470 to 5,098 of the cDNA, which is 2.6 kb in size) and NPA3 which is from nucleotide 1,866 to 5,098 of the cDNA and 3.2 kilo base in size.

NPA2-G (2.6 kb)	POSITION	CODON CHANGE	AMINO ACID CHANGE
MUTATION # 1	2777	CTG to CTA	Leu to Leu
MUTATION # 2	3572	AAT to GAT	Asn to Asp
NPA2-B (2.6 kb)			
MUTATION # 1	5019	ACT to GCT	Thr to Ala
NPA3-7 (3.2 kb)			
MUTATION # 1	2428	TCT to TCG	Ser to Ser
MUTATION # 2	2608	ACA to ACG	Thr to Thr
MUTATION # 3	4725	GGC to GTC	Gly to Val

Table 3.6. The sequence results (cont'd).

MUTATION # 4	4915	GAC to GAT	Asp to Asp
MUTATION # 5	4926	AAG to AGG	Lys to Arg
MUTATION # 6	3471	GGT to GAT	Gly to Asp
NPA3-10 (3.2 kb)			
MUTATION # 1	2472	ACA to ACG	Thr to Thr
MUTATION # 2	2430	CCA to CCG	Pro to Pro
MUTATION # 3	3570	AAG to AGG	Lys to Arg
MUTATION # 4	3758	CTA to CTT	Leu to Leu
MUTATION # 5	4135	AAG to AAA	Lys to Lys
MUTATION # 6	4316	AAA to GAA	Lys to Glu
MUTATION # 7	4798	ACT to ACC	Thr to Thr
NPA3-3 (3.2 kb)			
MUTATION # 1	3940	GAG to GGG	Glu to Gly
MUTATION # 2	5317	AAG to GAG	Lys to Glx
NPA3-14 (3.2 kb) (Just 2 primer)			
MUTATION # 1	4342	GGG to GAG	Gly to Glu
MUTATION # 2	4739	AAT to GAT	Asn to Asp
NPA1-102			
MUTATION # 1	1100	TTT to TTC	Phe to Phe
MUTATION # 2	1267	GAC to GGC	Asp to Gly
MUTATION # 3	1316	AGT to AGC	Ser to Ser

After receiving the previous sequence results, which harbored more mutations than we expected to have when using Expand High Fidelity DNA polymerase, we decided to change some parameters in PCR conditions such as dNTP and template DNA concentrations. In the new PCR set ups 100µM dNTP was used instead of 200µM (for final concentration) and 150 ng of template plasmid was used instead of 10 ng which was used in the previous set ups for amplifying the second partial fragment of the gene (NPA2). The PCR product from this step was sub cloned into TOPO-TA cloning vector. Also the correct clones from this cloning reaction, which contained both *Xho1* and *Apa1* digestion

sites on their 5' and 3' ends were selected (by restriction digestion) to be sequenced (by McLab). The results of sequencing are shown in Table 3.7.

Table 3.7. The PCR generated mutation list of sequenced clones of the NPA2 (the second partial fragment of the gene, from nucleotide 2,470 to 5,098 of the cDNA, which was 2.6 kb in size and was cloned into TOPO-TA cloning vector).

	POSITION	**CODON CHANGE**	**AMINO ACID CHANGE**
NPA2-203 (2.6 kb)			
MUTATION # 1	2452	ATA to ATG	Ile to Met
MUTATION # 2	2767	CAA to CAG	Gln to Gln
MUTATION # 3	2805	GAT to GGT	Asp to Gly
MUTATION # 4	2988	GTC to GCC	Val to Ala
MUTATION # 5	3282	GTC to GCC	Val to Ala
NPA2-206 (2.6 kb)			
MUTATION # 1	2452	ATA to ATG	Ile to Met
MUTATION # 2	3948	AGT to GGT	Ser to Gly
MUTATION # 3	3842	GAA to GAG	Glu to Glu
NPA2-212 (2.6 kb)			
MUTATION # 1	2452	ATA to ATG	Ile to Met
MUTATION # 2	3085	CCT to CCA	Pro to Pro
MUTATION # 3	3104	AGA to AGC	Arg to Ser
MUTATION # 4	3548	GAA to GAG	Glu to Glu
MUTATION # 5	3851	GCT to GCC	Ala to Ala
MUTATION # 6	4347	GAG to GGG	Glu to Gly

After increasing the amount of template DNA from 10ng to 150ng and decreasing the final concentration of dNTP from 200µM to 100µM, the number of mutations, produced by the enzyme during PCR was not decreased. So we

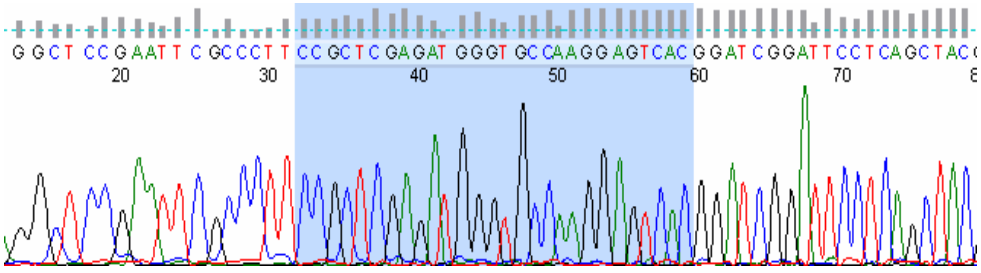
decided to change the elongation time and annealing temperature. The elongation time was decreased as much as possible and the annealing temperature was increased up to 77ºC, (when the melting temperature of the primers was 75ºC-76ºC). The PCR products after this alteration were cloned again into TOPO-TA cloning vector. The correct cloned constructs which were digested by *Xho1* and *Apa1* restriction enzymes, and contain both restriction sites, were selected to be sent for sequencing. The sequencing was done by McLab. The sequence results are shown in Table 3.8.

Table 3.8. The PCR generated mutation list of selected clones of *USP32* full and partial fragments which were cloned into TOPO-TA cloning vector after changing some PCR conditions. NPA1; the first fragment of the gene from nucleotide 287 to 2,484 of the cDNA, which was 2.2 kb long) and N-FULL; the full length cDNA of *USP32* gene, which was 4.8 kb in size and from nucleotide 287 to 5,098 of the cDNA.

FULL-N-4 (4.8 kb)	POSITION	CODON CHANGE	AMINO ACID CHANGE
MUTATION # 1	4019	AGT to AGC	Ser to Ser
MUTATION # 2	837	GCT to ACT	Thr to Ala
NPA1-118 (2.2kb)			
MUTATION # 1	764	TCT to CCT	Ser to Pro
MUTATION # 2	2247	TAT to TGT	Tyr to Cys
NPA1-111 (2.2kb)			
MUTATION # 1	358	GAG to GAA	Glu to Glu
N-FULL-17 (4.8 kb)			
MUTATION # 1	2085	CGC to TGC	Arg to Cys
N-FULL-18 (4.8 kb)			
MUTATION # 1	1269	ACC to GCC	Thr to Ala
N-FULL-13 (4.8 kb)			
MUTATION # 1	715	TAT to TAC	Tyr to Tyr

From these constructs, NPA1-111 (the first 2.2 kb fragment of the gene inside TOPO-TA vector) and N-FULL-13 (the full length *USP32* cDNA) constructs are suitable for cloning into pEGFP-N1 vector. There was one silent mutation in each of these constructs which didn't change the amino acid sequence of the gene, so wouldn't change the protein, coded by this gene. The sequences of primers inside the NPA1-111 and N-FULL-13 constructs are shown in Figures 3.18 and 3.19. These primers contain both recognition and restriction sites of *Apa1* and *Xho1* enzymes.

A



B

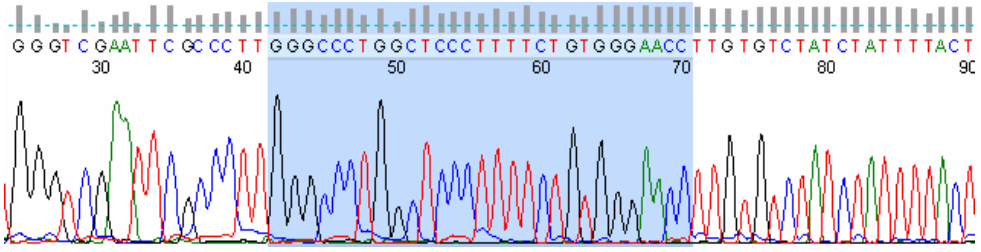
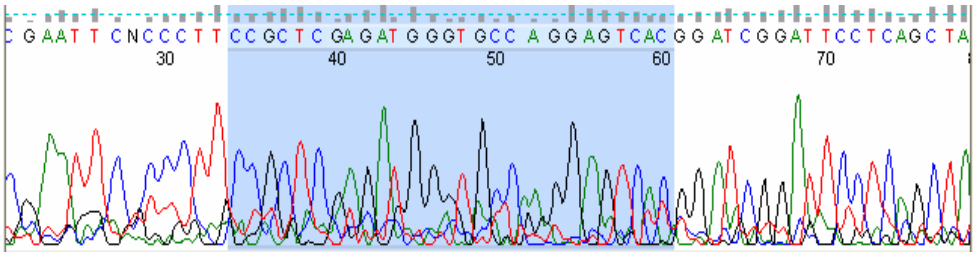


Figure 3.18. The sequence of the primers inside the NPA1-111 construct (the first 2.2 kb fragment of the gene inside TOPO-TA vector) A. The sequence of forward primer (F1N), containing both recognition and restriction sites for *Xho1* enzyme, sequenced by GW1 primer. B. The sequence of reverse primer (R2N), containing restriction site for *Apa1* enzyme, sequenced by GW2 primer.

A



B

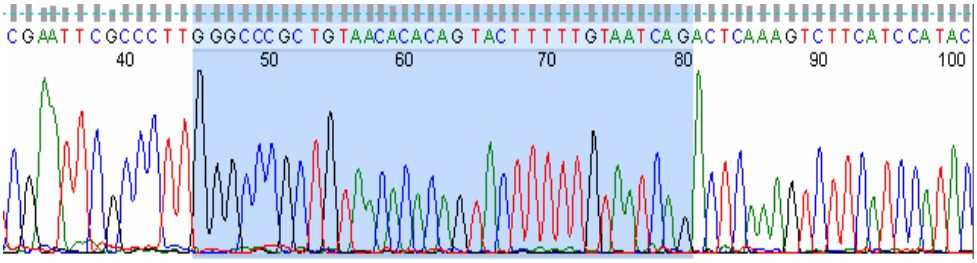


Figure 3.19. The sequences of the primers inside the N-FULL-13 construct (the full length *USP32* cDNA inside TOPO-TA vector). A. The forward primer (F1N), containing recognition and restriction sites for *Xho1* enzyme, sequenced by GW1 primer. B. The Reverse primer (R1N), containing restriction site for *Apa1* enzyme, sequenced by GW2 primer.

The full and partial fragments of the *USP32* gene, which were cloned into TOPO-TA cloning vector and found to have correct DNA sequences, are ready for further functional studies. These constructs also were designed in a way that is suitable for cloning into the pEGFP-N1 vector without interrupting the vector's open reading frame as this vector is used for expression purposes in mammalian cells.

CAPTER 4

CONCLUSION

Ubiquitin Specific Proteases (USPs), which are a sub class of proteases, have been found to play important roles in proteasomal degradation system and they can interfere with many important regulatory pathways as they have the ability of deubiquitinating the target proteins and inhibiting their degradation. *USP6* and *USP32* are two members of the USP family.

USP32 gene maps to 17q23, which is a region that undergoes amplification and overexpression in a number of breast cancer cells**.**

USP6 is a novel fusion gene with high similarity to *TBC1D3* (89%) and *USP32* (97%) genes. *USP6* is a hominoid specific gene which is located on chromosomal band 17p13 and is known to be an oncogene. The expression of this gene has been shown to be restricted to the testis. As *USP32* gene is amplified in breast cancer cells and it shows high similarity to *USP6* gene, we tried to compare the expression pattern of these two genes in normal tissues and in 2 breast cancer cell lines. In this study, the expression of *USP6* gene has been shown in both testis and ovary tissues whereas *USP32* gene was expressed in normal breast, testis and ovary cells and also in MCF7 and MDA-MB-231 breast cancer cell lines.

No expression of *USP6* was detected in neither in MDA-MB-231, MCF7 breast cancer cell lines nor in normal breast tissue.

It has been suggested that the *USP6* gene has two possible transcripts (transcript 210 and 213) with each having two open reading frames.

The expression of *USP6* full length mRNA and/or transcript 210 was detected in both testis and ovary tissues.

The transcript 213 (full length) of the *USP6* gene was not detected in none of the normal breast tissue, breast cancer cell lines (MDA-MB-231 and MCF7), testis or ovary tissue.

The first open reading frame of the transcript 213 was detected to be expressed only in testis tissue.

The second open reading frame of the transcript 213 was not detected in breast tissue and breast cancer cell lines and testis and ovary tissues.

The existence of other isoforms of *USP6* gene is also possible. This isoforms are the products of alternative splicing in which different splice sites cause alternative splicing of the mRNA and lead to the appearance of different transcripts of the gene inside the cells.

The expression comparison of *USP32* and *USP6* (with its two possible transcripts) enables us to do further functional characterization studies to reveal the exact role of these proteins.

To do further characterization studies (such as expression and localization) and also for the purpose of revealing the role of different domains of the *USP32* gene, the full length cDNA, and the partial fragments of this gene were sub-cloned into the TOPO-TA cloning vector and confirmed by sequencing.

Based on our initial data, *USP32* overexpression can now be studied in breast cancer cells as we eliminated the potential risk of a functional overlap with a similar and known oncogene *(USP6),* in breast cells.

Understanding how protein degradation mechanisms may be involved in breast cancer will help us to elucidate the molecular mechanism behind mammary tumorigenesis.

To understand the function of the *USP32* gene inside the cell, we are going to perform localization studies and according to the localization results, we may predict some functions for this gene. In addition we are going to use PPT technique (Protein Protease Technique) by which the exact location of the interested protein (which can be cytoplasm, nucleus or cell membranes) will be detected.

In addition, further functional assays can be performed by overexpressing *USP32* in non-tumorigenic mammary cell lines to observe any transformation characteristics of the cells. Understanding the role of *USP32* in breast tumorigenesis will help us reveal the importance of protein degradation pathways in cancer.

REFERENCES

Barlund, M., Monni, O., Kononen, J., Cornelison, R., Torhorst, J., Sauter, G., Kallioniemi, O., Kallioniemi, A. (2000). Multiple genes at 17q23 undergo amplification and overexpression in breast cancer. *Cancer Res* 60: 5340–5344.

Barlund, M., Tirkkonen, M., Forozan, F., Tanner, MM., Kallioniemi, O., Kallioniemi, A. (1997). Increased copy number at 17q22–q24 by CGH inbreast cancer is due to high-level amplification of two separate regions. *Genes Chromosomes Cancer* 20: 372–376.

Brinkschmidt, C., Christiansen, H., Terpe, HJ., Simon, R., Boecker, W., Lampert, F., Stoerkel, S. (1997). Comparative genomic hybridization (CGH) analysis of neuroblastomas–an important methodological approach in paediatric tumour pathology. *J Pathol* 181: 394–400.

Borodovsky, A., Kessler, B.M., Casagrande, R., Overkleeft, H.S., Wilkinson K.D., and Ploegh, H.L. (2001). A novel active site-directed probe specific for deubiquitylating enzymes reveals proteasome association of USP14. *EMBO J* 20, 5187–5196

Bulavin, DV., Demidov, ON., Saito, S., Kauraniemi, P., Phillips, C., Amundson, SA., Ambrosino, C., Sauter, G., Nebreda, AR., Anderson, CW., Kallioniemi, A., Fornace, Jr, AJ., Appella, E. (2002). Amplification of PPM1D in human tumours abrogates p53 tumour-suppressor activity. *Nature Genetics* 31(2):210-5.

Charles, A., Paulding, Maryellen, Ruvolo and Daniel, A. Haber. (2003). The Tre2 (USP6) oncogene is a hominoid-specific gene. *PNAS* vol. 100, no. 5, 2507–2511.

Clark, J., Edwards, S. John, M., Flohr, P., Gordon, T., Maillard, K., Giddings, I., Brown, C., Bagherzadeh, A., Campbell, C., Shipley, J., Wooster, R., Cooper, CS. (2002). Identification of amplified and expressed genes in breast cancer by comparative hybridization onto microarrays of randomly selected cDNA clones. *Genes Chromosomes Cancer* 34: 104–114.

C.M, Pickart. (2001). Mechanisms underlying ubiquitination, Annu. Rev. *Biochem.*70 503–533.

Colleen, S, Sinclair., Matthew, Rowley., Ali, Naderi., and Fergus J. Couch. (2003). The 17q23 amplicon and breast cancer. *Breast cancer research and treatment* 78: 313-322.

Couch, FJ., Wang, XY., Wu, GJ., Qian, J., Jenkins, RB., James, CD. (1999). Localization of PS6K to chromosomal region 17q23 and determination of its amplification in breast cancer. *Cancer Res* 59: 1408–1411.

De, Jong, MM., Nolte, IM., te, Meerman, GJ., van, der,. Graaf, WT., Oosterwijk, JC., Kleibeuker, JH., Schaapveld, M., de, Vries. EG. (2002). Genes other than BRCA1 and BRCA2 involved in breast cancer susceptibility. *J Med Genet* 39:225-42.

Elizabeth, Hyman., Päivikki, Kauraniemi., Sampsa, Hautaniemi., Maija, Wolf., Spyro, Mousses., Ester, Rozenblum., Markus, Ringnér., Guido, Sauter., Outi, Monni., Abdel, Elkahloun., Olli-P, Kallioniemi and Anne,

Kallioniemi. (2002). Impact of DNA Amplification on Gene Expression Patterns in Breast Cancer. *Cancer Research* 62, 6240-624.

Erson, AE., Niell, BL., DeMers, SK., Rouillard, JM., Hanash, SM., Petty, EM. (2001). Overexpressed genes/ESTs and characterization of distinct amplicons on 17q23 in breast cancer cells. *Neoplasia* 3: 521– 526.

Foulkes, WD., Narod, SA. (1995). Hereditary breast and ovarian cancer: epidemiology, genetics, screening and predictive testing. *Clin Invest Med* 18:473-83.

Giguere, Y., Dewailly, E., Brisson, J., Ayotte, P., Laflamme, N., Demers, A., Forest, VI., Dodin, S., Robert, J., Rousseau, F. (2001). Short polyglutamine tracts in the androgen receptor are protective against breast cancer in the general population. *Cancer Res* 61:5869-74.

Glickman, M, H., and Ciechanover, A. (2002). The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. *Physiol. Rev.* 82, 373-428.

Goldberg, AL., Stein, R., Adams, J. (1995). New insights into proteasome function: from archaebacteria to drug development. *Chem Biol* 503–508.

Gong, L., Kamitani, T., Millas, S., and Yeh, E.T (2000) Identification of a novel isopeptidase with dual specificity for ubiquitin- and NEDD8-conjugated proteins. *J. Biol. Chem* 275, 14212–14216.

Henderson, BE., Feigelson, HS. (2000) Hormonal carcinogenesis. *Carcinogenesis* 21: 427-433.

Hermon, C., Beral, V. (1996). Breast cancer mortality rates are leveling off or beginning to decline in many western countries: analysis of time trends, age-cohort and age-period models of breast cancer mortality in 20 countries. *Br J Cancer* 73955- 960.

J, Parssinen., T, Kuukasjarvi., R, Karhu., A, Kallioniemi. (2007). High level amplification at 17q23 leads to coordinated overexpression of multiple adjacent genes in breast cancer. *British journal of cancer* 96, 1258-1264.

Kallioniemi, A., Kallioniemi, OP., Piper, J., Tanner, M., Stokke, T., Chen, L., Smith, HS., Pinkel, D., Gray, JW., Waldman, FM. (1994). Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc Natl Acad Sci USA* 91: 2156–2160.

Kliewer, E., Smith, K. Breast cancer mortality among immigrants in Australia and Canada. *J Natl Cancer Inst* 1995, 87:1154- 1161.

Knuutila, S., Bjorkqvist, AM., Autio, K., Tarkkanen, M., Wolf, M., Monni, O., Szymanska, J., Larramendy, ML., Tapper, J., Pere, H., El-Rifai, W., Hemmer, S., Wasenius, VM., Vidgren, V., Zhu Y. (1998). DNA copy number amplifications in human neoplasms: review of comparative genomic hybridization studies. *Am J Pathol* 152: 1107–1123.

Korn, WM., Oide, Weghuis, DE., Suijkerbuijk, RF., Schmidt, U., Otto, T., du, Manoir, S., Geurts, van, Kessel, A., Harstrick, A., Seeber, S., Becher, R. (1996). Detection of chromosomal DNA gains and losses in testicular germ cell

tumours by comparative genomic hybridization. *Genes Chromosomes Cancer* 17: 78–87.

Lacroix, AZ., Burrke, W. (1997). Breast cancer and hormone replacement therapy. *Lancet* 350:1042-43.

Li, J., Yang, Y., Peng, Y., Austin, RJ., van, Eyndhoven, WG., Nguyen, KC., Gabriele, T., McCurrach, ME., Marks, JR., Hoey, T., Lowe, SW., Powers, S. (2002). Oncogenic properties of PPM1D located within a breast cancer amplification epicenter at 17q23. *Nat Genet* 31: 133–134.

Lillian, S., DeBruin and P, David, Josephy. (2002). Perspectives on the Chemical Etiology of Breast Cancer. *Environmental Health Perspectives* volume 110.

Marchio, A., Meddeb, M., Pineau, P., Danglot, G., Tiollais, P., Bernheim, A., Dejean, A. (1997). Recurrent chromosomal abnormalities in hepatocellular carcinoma detected by comparative genomic hybridization. *Genes Chromosomes Cancer* 18: 59–65.

Mark, Hochstrasser. (2005). Lingering Mysteries of Ubiquitin-Chain Assembly, *j.cell*.12.025.

Monni, O., Barlund, M., Mousses, S., Kononen, J., Sauter, G., Heiskanen, M., Paavola, P., Avela, K., Chen, Y., Bittner, ML., Kallioniemi, A. (2001). Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. *Proc Natl Acad Sci USA* 98: 5711– 5716.

Muleris, M., Almeida, A., Gerbault-Seureau, M., Malfoy, B., Dutrillaux, B. (1994). Detection of DNA amplification in 17 primary breast carcinomas with

homogeneously staining regions by a modified comparative genomic hybridization technique. *Genes Chromosomes Cancer* 10: 160–170.

Nakamura, T., Hillova, J., Mariage-Samson, R., Onno, M., Huebner, K., Cannizzaro, L. A., Boghosian-Sell, L., Croce, C. M. & Hill, M. (1992). Novel transcriptional unit of the *tre* oncogene widely expressed in human cancer cells. *Oncogene* **7,** 733–741.

Nicole.R. Bianco., George, Perry., Mark, A., Smith, Dennis, J.Templeton., and Monica, M. Montano. (2003). Functional Implications of Antiestrogen Induction of Quinone Reductase: Inhibition of Estrogen-InducedDeoxyribonucleic Acid Damage. *Molecular Endocrinology* 17(7):1344–1355.

Parkin, DM., Muir, CS., Whelan, SL., Gao, YT., Ferlay, J., Powell, J (Eds). (1992). Cancer Incidence in Five Continents, vol. VI [International Agency for Research on Cancer Scientific Publications, No. 120]. Lyon: International Agency for Research on Cancer Scientific.

Park, K.C., Woo, S.K., Yoo, Y.J., Wyndham, A.M., Baker, R.T., and Chung, C.H. (1997). Purification and characterization of UBP6, a new ubiquitin-specific protease in Saccharomyces cerevisiae. *Arch. Biochem. Biophys* 347, 78–84.

Puente, X.S., and Lopez-Otin, C. (2004). A genomic analysis of rat proteases and protease inhibitors. *Genome Res* 14, 609–622.

Ried T, Petersen I, Holtgreve-Grez H, Speicher MR, Schrock E, du Manoir S, Cremer T. (1994). Mapping of multiple DNA gains and losses in primary small cell lung carcinomas by comparative genomic hybridization. *Cancer Res* 54: 1801–1806

Ries, LAG., Eisner, MP., Kosary, CL., Hankey, BF., Miller, BA., Clegg, L., Mariotto, A., Feuer, EJ., Edwards, BK (Eds). (1975-2001). *SEER Cancer Statistics Review*.

Richter, J., Jiang, F., Gorog, JP., Sartorius, G., Egenter, C., Gasser, TC., Moch, H., Mihatsch, MJ., Sauter, G. (1997) Marked genetic differences between stage pTa and stage pT1 papillary bladder cancer detected by comparative genomic hybridization. *Cancer Res* 57: 2860–2864.

Rock, KL., Goldberg, AL. (1999). Degradation of cell proteins and the generation of MHC class1-presented peptides. *Ann Rev Immunol* 17, 139–179.

Schwendel, A., Langreck, H., Reichel, M., Schrock, E., Ried, T., Dietel, M., Petersen, I. (1997). Primary small-cell lung carcinomas and their metastases are characterized by a recurrent pattern of genetic alterations. *Int J Cancer* 74: 86–93

Sebastian M.B. Nijman, Mark P.A. Luna-Vargas, Arno Velds, Thijn R. Brummelkamp, Annette M.G. Dirac,Titia K. Sixma and René Bernards. A (2005) Genomic and Functional Inventory of Deubiquitinating Enzymes.cell.

Simon, S. Wing. (2003). Deubiquitinating enzymes — the importance of driving in reverse along the ubiquitin–proteasome pathway. *The International Journal of Biochemistry & Cell Biology* 35 590–605 genomic insatiability

Solinas-Toldo, S., Wallrapp, C., Muller-Pillasch, F., Bentz, M., Gress, T., Lichter, P. (1996). Mapping of chromosomal imbalances in pancreatic carcinoma by comparative genomic hybridization. *Cancer Res* 56: 3803–3807

Sonoda, G., Palazzo, J., du, Manoir, S., Godwin, AK., Feder, M., Yakushiji, M., Testa, JR. (1997). Comparative genomic hybridization detects frequent

overrepresentation of chromosomal material from 3q26, 8q24, and 20q13 in human ovarian carcinomas. *Genes Chromosomes Cancer* 20:320–328

Susan, E Hankinson1., Graham, A, Colditz., and Walter, C, Willett. (2004). Towards an integrated model for breast cancer etiology, The lifelong interplay of genes, lifestyle, and hormones. *Breast Cancer Res* 6:213-218.

Thrower, J.S., Hoffman, L., Rechsteiner, M., and Pickart, C.M. (2000). Recognition of the polyubiquitin proteolytic signal. *EMBO J* 19, 94–102.

Van, den, Brandt, PA., Spiegelman, D., Yaun, SS., Adami, HO., Beeson, L., Folsom, AR., Fraser, G., Goldbohm, RA., Graham, S., Kushi, L., Marshall, JR., Miller, AB and others. (2000). Pooled analysis of prospective cohort studies on height, weight, and breast cancer risk. *Am J Epidemiol* 152: 514-27.

Vandesompele, J., Van Roy, N., Van, Gele, M., Laureys, G., Ambros, P., Heimann, P., Devalck, C., Schuuring, E., Brock, P., Otten, J., Gyselinck. J., De, Paepe, A., Speleman, F. (1998). Genetic heterogeneity of neuroblastoma studied by comparative genomic hybridization. *Genes Chromosomes Cancer* 23: 141–152.

Verma, R., Aravind, L., Oania, R., McDonald, W.H., Yates, J.R., 3rd, Koonin, E.V., and Deshaies, R.J. (2002). Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science* 298, 611–615.

Weber, RG., Bostrom, J., Wolter, M., Baudis, M., Collins, VP., Reifenberger, G., Lichter, P. (1997). Analysis of genomic alterations in benign, atypical, and anaplastic meningiomas: toward a genetic model of meningioma progression. *Proc Natl Acad Sci USA* 94: 14719–14724.

Welcsh, PL., Lee, MK., Gonzalez-Hernandez, RM., Black, DJ., Mahadevappa, M., Swisher, EM., Warrington, JA., King, MC. (2002). BRCA1 transcriptionally regulates genes involved in breast tumorigenesis. *Proc Natl Acad Sci USA* 99:7560-5.

Willis, S., Hutchins, AM., Hammet, F., Ciciulla, J., Soo, WK., White, D., van, der, Spek, P., Henderson, MA., Gish, K., Venter, DJ., Armes, JE. (2003). Detailed gene copy number and RNA expression analysis of the 17q12–23 region in primary breast cancers. *Genes Chromosomes Cancer* 36: 382–392.

Wu, G., Sinclair, C., Hinson, S., Ingle, JN., Roche, PC., Couch, FJ. (2001). Structural analysis of the 17q22–23 amplicon identifies several independent targets of amplification in breast cancer cell lines and tumours. *Cancer Res* 61: 4951– 4955.

Wu, GJ., Sinclair, CS., Paape, J., Ingle, JN., Roche, PC., James, CD., Couch FJ. (2000). 17q23 amplifications in breast cancer involve the PAT1, RAD51C, PS6K, and SIGma1B genes. *Cancer Res* 60: 5371–5375.

Zheng, W., Xie, D., Cerhan, JR., Sellers, TA., Wen, W., Folsom, AR., Sulfotransferase, 1A1. (2001). polymorphism, endogenous estrogen exposure, well-done meat intake, and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 10:89-94.

Ziegler, RG., Hoover, RN., Pike, MC., Hildesheim, A., Nomura, AM., West , DW., Wu- Williams, AH., Kolonel, LN., Horn-Ross, PL., Rosenthal, JF. (1993). Migration patterns and breast cancer risk in Asian- American women. *J Natl Cancer Inst* 85:1819-1827.

APPENDIX A

PRIMERS

Primers for Amplification

Primer name	Sequence	Tm
GAPDH-F	5'- GGGAGCCAAAAGGGTCATCA-3'	60°C
GAPDH-R	5'- TTTCTAGACGGCAGGTCAGGT-3'	60°C
F1N	5'- CCGCTCGAGATGGGTGCCAAGGAGTCAC-3'	76°C
F2N	5'- CCGCTCGAGATGGATAGACACAAGGTTCCCA-3'	75°C
F3N	5'- CCGCTCGAGATGCCAGGGGCTATTGATAATCA-3'	75°C
R1N	5'-GGGCCCGCTGTAACACACAGTACTTTTTGTAATCAG-3'	75°C
R2N	5'- GGGCCCTGGCTCCCTTTTCTGTGGGAAC-3'	76°C
213- F	5'- CGTCAGTGTCAGACCACAGG-3'	60°C
1F (B)	5'- GTGTGGTCAGTCCTCCTGAAC-3'	60°C
1R (B)	5'- CTTCCTGCCACACACCTTC-3'	60°C
2F (B)	5'- GGACAAATCCCATTGGTATGA-3'	60°C
2R (B)	5'- CCATTGGGGATGAATATTGG -3'	60°C
3R (B)	5'- GTCCACAGAGATCCCTCAGC -3'	60°C
R (S)	5'-CCAATGGGATTTGTCCTGTT-3'	60°C
2F (CDS)	5'- GACACCACAAAGATGGGTCA -3'	60°C
2R (CDS)	5'- TCAATAGCCCCTGGTTTCTG -3'	60°C

Primers for Sequencing

Primer name	Sequence	Tm
CDS 1F	5'-TCTCGATGGCTTCTATCTGGA -3'	60°C
CDS 1R	5'-CACCTGGAAAAGGAGGTTCA -3'	60°C
CDS 2F	5'- GACACCACAAAGATGGGTCA-3'	60°C
CDS 2R	5'-TCAATAGCCCCTGGTTTCTG -3'	60°C
CDS 3F	5'-CTGAAGCCTCAGAAACTGCT -3'	60°C
CDS 3R	5'-CTCTTTAATGCGCAGCCTTT -3'	60°C
CDS 4F	5'-CCCCGCTATCTTCTCTTCCT -3'	60°C
CDS 4R	5'-TTTAGCCATATGCCCCTTCA -3'	60°C
CDS 5F	5'-TGAACTCAACAGGACAAATCC -3'	60°C
CDS 5R	5'-CCACAGAGATCACTCAGCTGTT -3'	60°C
CDS 6F	5'-ATGGGTTTCAGCAACAGGAC -3'	60°C
CDS 6R	5'-TGGCGAAGAGGAGAAATCTG -3'	60°C
CDS 7F	5'-CCTGTCCCTGTGTCTCCAAT -3'	60°C
CDS 7R	5'- AAGCTCTGTCTTCCCCACAA-3'	60°C
CDS 8F	5'- GATGCGGTTTGGATTCAAGT-3'	60°C
CDS 8R	5'- AGAGCCGGGTCTCTTGGTA-3'	60°C
CDS 9F	5'- CAAAGAAGCTGGATCTCTGGA-3'	60°C
CDS 9R	5'- GGTCCTGAGGAGTGACCAAC-3'	60°C
SN1F	5'- CACCAAAATCAACGGGACTT-3'	60°C
SN1R	5'- GCTGAACTTGTGGCCGTTTA-3'	60°C
GW1	5'- GTTGCAACAAATTGATGAGCAATGC-3'	60°C
GW2	5'- GTTGCAACAAATTGATGAGCAATTA-3'	60°C

APPENDIX B

DNA MARKERS

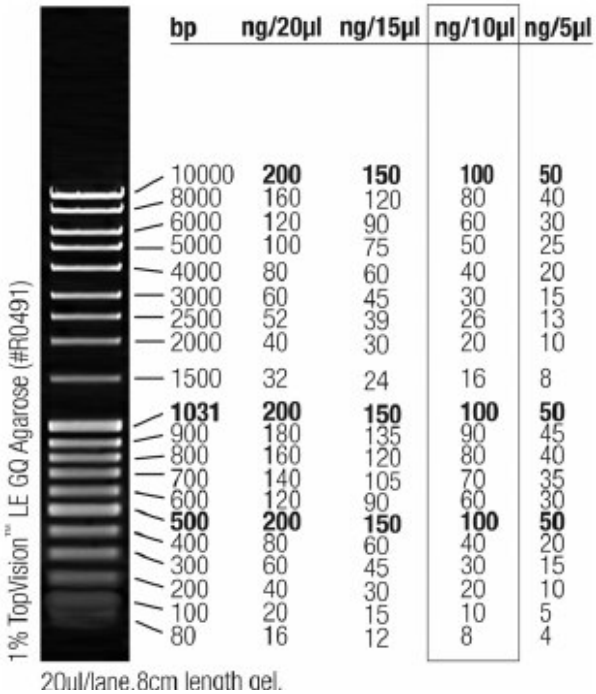


Figure B.1.. Mass Ruler DNA Ladder Mix.

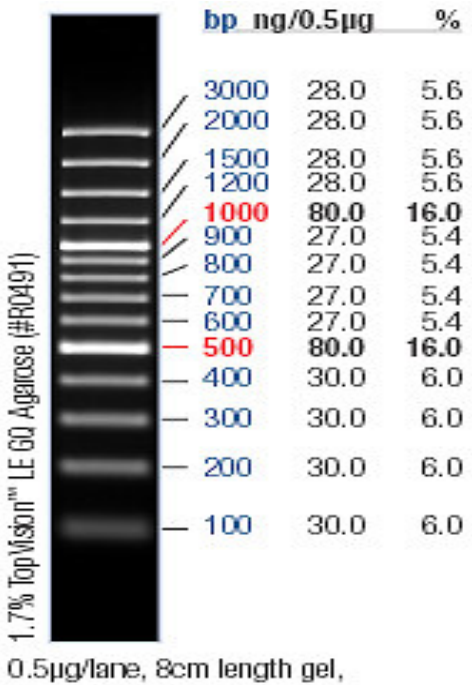


Figure B.2. Gene Ruller 100bp DNA Ladder

APPENDIX C

BUFFERS, SOLUTIONS AND GROWTH MEDIA

TBE Buffer (5X)/L

Tris base	54 g
Boric Acid	27.59 g
0.5 M EDTA (Ph 8.0)	20ml

Stored at room temperature.

TE Buffer (10X)

100 mM Tris-Cl (pH 8.0)

10 mM EDTA (pH 8.0)

pH was set to 8.0 and sterilized for 20 minutes at 15 Psi. Stored at room temperature.

DNA Loading Dye (10X)

Sucrose	2.5 g
Brom. Blue	0.5g
1M Tris (pH 8.0)	0.5 ml
0.5M EDTA (pH 8.0)	1 ml
Water	48.5 ml

Stored at -4ºC

Bacterial Growth Media

LB-Medium (Luria-Bertani Medium)

Component:

Tryptone	10 g
Yeast Extract	5 g
NaCl	10 g

The volume was reached to 1 liter and the medium was sterilized by autoclaving for 20 minutes at 15 Psi.

LB-Agar

Component:

Bacto Agar	15 g
Tryptone	10 g
Yeast Extract	5 g
NaCl	10 g

The volume was adjusted to 1 liter. The medium was sterilized by autoclaving for 20 minutes at 15 Psi. The pH was adjusted to 7 with using NaOH (5 N).

S.O.C Medium

Component:

Tryptone	20 g
Yeast Extract	5 g
NaCl	0.5 g
250 mM KCl	10 ml

The volume was adjusted to 1 liter. The pH was adjusted to 7, using NaOH (5 N). The medium was sterilized by autoclaving for 20 minutes at 15 Psi. 20 ml of sterile 1M solution of glucose was added to the cooled medium. Just before use, 5 ml of a sterile solution of $MgCl_2$ (2M) was added to the medium.

Tissue Culture Media

Minimum Essential Medium

Supplemented with: Earl's Salt and L-Glutamine supplemented with 10% Fetal Calf Serum (FCS) and 1% penicillin-streptomycin

Dulbecco's Modified Eagle' Medium

Supplemented with: L-Glutamine, 10% Fetal Calf Serum (FCS) and 1% Penicillin- streptomycin.

APPENDIX D

ENZYMES, DNA MARKERS AND KITS

Enzymes

Apa1	Roche
Expand Long Template DNA Polymerase	Roche
Expand High Fidelity DNA Polymerase	Roche
EcoR1	MIB Fermentas
Pfu DNA Polymerase	Promega
Phosphatase Alkaline Shrimp	Roche
RevertAid™ M-MuLV Reverse Transcriptase	Invitrogen
Taq DNA Polymerase	Applichem
Trypsin	PAA
T4 DNA Ligase	Roche
Xho1	MIB Fermentas

DNA Markers

Mass Ruler DNA Ladder Mix.	MIB Fermentas
Gene Ruller 100bp DNA Ladder	MIB Fermentas

Kits

Gel Extraction Kit	Roche
pCR ® 8/GW/TOPO®TA Cloning ® Kit	Invitrogen
Plasmid DNA extraction Mini Kit	GeneMark
Plasmid DNA extraction Mini Kit	Favorgen

Plasmid DNA Extraction Midi Kit	Roche
RevertAid™ First Strand cDNA Synthesis Kit	Invitrogen

APPENDIX E

CHEMICALLS AND THEIR SUPPLIERS

Chemicals	**Supplier**
Acetic acid	Merck
Agar Bacteriological	AppliChem
Agarose	Merck
Boric Acid	.. AppliChem
Bromophenl Blue	AppliChem
$CaCl_2$	AppliChem
Chloroform	AppliChem
DMSO	AppliChem
dNTP	MBI Fermentas
EDTA	AppliChem
Ethanol	AppliChem
Ethidium bromide	AppliChem
Glucose	AppliChem
Glycerol	AppliChem
HCl	AppliChem
Isoamyl alcohol	AppliChem
KCl	AppliChem
Luria Broth	Sigma
$MgCl_2$	AppliChem
NaCl	AppliChem
NaOH	AppliChem
Phenol-chloroform-isoamylalcohol	AppliChem

Sucrose	AppliChem
Tris Base	AppliChem
Trizol	Invitrogen
Tryptone	AppliChem
Yeast Extract	AppliChem
Isopropyl alcohol	AppliChem