CLONING AND INITIAL PROTEIN CHARACTERIZATION OF AN ESTROGEN RESPONSIVE GENE: *YPEL2*

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

BY

GİZEM GÜPÜR

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

AUGUST 2014

Approval of the Thesis:

CLONING AND INITIAL PROTEIN CHARACTERIZATION OF AN ESTROGEN RESPONSIVE GENE: *YPEL2*

submitted by **GİZEM GÜPÜR** 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. Orhan Adalı
Head of Department**, Biology** _____

Assoc. Prof. Dr. Mesut Muyan
Supervisor, **Biology Dept., METU** _____

Examining Committee Members:

Assoc. Prof. Dr. A. Elif Erson-Bensan
Biology Dept., METU _____

Assoc. Prof. Dr. Mesut Muyan
Biology Dept., METU _____

Assoc. Prof. Dr. Sreeparna Banerjee
Biology Dept., METU _____

Assist. Prof. Dr. Mehmet Somel
Biology Dept., METU _____

Assist. Prof. Dr. Özgür Şahin
Molecular Biology and Genetics Dept., Bilkent University _____

Date: 14.08. 2014

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Gizem Güpür

Signature :

ABSTRACT

CLONING AND INITIAL PROTEIN CHARACTERIZATION OF AN ESTROGEN RESPONSIVE GENE: *YPEL2*

Güpür, Gizem
M.S., Department of Biology
Supervisor: Assoc. Prof. Dr. Mesut Muyan
August 2014, 104 pages

17β-estradiol (E2), the main circulating estrogen in the body, is involved in physiological regulation of many tissue and organ functions, including mammary tissue. E2 is also involved in target tissue malignancies. E2 regulates cellular proliferation, differentiation and death in target tissues. The lasting effects of E2 on cells are mediated by estrogen receptor α and β that are the products of distinct genes and act as transcription factors. Upon binding to E2, the activated ER regulates the expression of E2 target genes through ERE (estrogen response element)-dependent and ERE-independent signaling pathways. The ERE-dependent signaling pathway refers to transcription events initiated by the interaction of E2-ER with ERE sequences. The transcription regulation involving the functional interactions of E2-ER with other transcription factors bound to their cognate response elements on DNA is called as the ERE-independent signaling pathway. In a microarray study conducted in our laboratory to identify genes involved in ERE-dependent and ERE-independent signaling pathways, *YPEL2*, a member of the highly conserved Yippee-like (*YPEL*) gene family, is suggested to be an E2 responsive gene regulated through the ERE-dependent signaling pathway. The *YPEL* gene family, named after Drosophila Yippee protein, has 100 members which share an extremely high amino-acid sequence identity in 68 species ranging from yeast, *C.elegans*, flies, plants to mammals. The members of the human YPEL genes, *YPEL1-5*, encode putative zinc binding small proteins with molecular

masses ranging from 13,500 to 17,500 Da. Although structures and functions of Ypel proteins are yet unclear, a limited number of studies suggests the involvement of Ypel proteins in development, cell cycle progression and mitosis, as well as cellular senescence and death. Our analyses using various bioinformatics tools suggest that Ypel proteins share a high degree of structural and functional properties that might be important for basic cellular processes. Our bioinformatics analyses also suggest that each YPEL gene is spatiotemporally regulated by different repertoire of transcription factors which may be activated by distinct signaling pathway in response to different internal and external clues.

To analyze the synthesis and intracellular localization of Ypel2, we initially cloned cDNAs of all five members of the human *YPEL* family, using a cDNA library from ERα-positive MCF7 cell line derived from a breast adenocarcinoma, for comparisons. We then showed that the un-liganded ERα regulates basal mRNA levels of *YPEL2*. Moreover, the expression of *YPEL2*, as well as *YPEL3*, is repressed by E2. These findings are consistent with our prediction that *YPEL2* and *YPEL3* are E2 and ERα responsive genes. We found that Ypel1, 2 and/or 3 are synthesized in COS7, derived from transformed African green monkey kidney fibroblast-like cells, and localized to a region just outside of the nucleus, however we could not detect any endogenous Ypel protein in MCF7 cells. On the other hand, we observed that over-expressions of *YPEL1-5* lead to the leakage of DNA from the nucleus into the cytoplasm in a pattern that overlaps with the localization of each Ypel protein in COS7 and MCF7 cells, in the latter the over-expression of Ypel1-5 is associated with a gross deterioration of the nuclear lamina integrity.

Future studies will address the regulation of *YPEL2* expression as well as the functions of Ypel2 in cell models.

Keywords: estrogen, estrogen receptor, *YPEL2*, Yippee

ÖZ

ÖSTROJEN YANIT GENİ *YPEL2*' NİN
KLONLANMASI VE PROTEİNİN İLKİN KARAKTERİZASYONU

Güpür, Gizem
Yüksek Lisans, Biyoloji Bölümü
Tez Yöneticisi: Doç. Dr. Mesut Muyan
Ağustos 2014, 104 sayfa

17β-östradiyol (E2), dolaşımdaki ana östrojen hormonu olup meme dokusu dahil birçok dokunun fonksiyonlarını düzenleyici role sahiptir. E2 nin bu dokulardaki habis (malignant) tümör oluşumunda ve gelişiminde rol oynadığı da bilinmektedir. E2, hedef dokularda hücre çoğalması, farklılaşması ve ölümünü düzenler. E2'nin hücre içindeki kalıcı etkileri, farklı gen ürünleri olan ve transkripsiyon faktörü olarak işlev gören öströjen reseptör α ve β tarafından düzenlenir. E2'nin bağlanmasıyla aktif hale gelen ER, E2 hedef genlerinin ifadelerini öströjen yanıt elemenları (estrogen response element, ERE)-bağımlı ve ERE-bağımsız sinyal yolaklarıyla düzenler. ERE-bağımlı sinyal yolağı, E2-ER kompleksinin ERE'lerle ilişkilenip başlattığı transkripsiyon olaylarını tanımlar. ERE-bağımsız sinyal yolağı ise E2-ER'nin kendi yanıt elemanlarına bağlanmış diğer transkripsiyon faktörleriyle girdiği işlevsel etkileşimler sonucu düzenlenen transkripsiyon olaylarını kapsar. Daha önce laboratuvarımızda yapılan ERE-bağımlı ve ERE-bağımsız sinyal yolaklarının düzenlediği genlerin tanısına yönelik mikrodizin çalışmalarında yüksek derecede korunmuş Yippee-like (*YPEL*) gen ailesinin bir üyesi olan *YPEL2* geninin ERE-bağımlı sinyal yolağıyla düzenlenen bir E2 yanıt geni olduğu öne sürülmüştür. Drosophila Yippee proteininin ardından adlandırılan *YPEL* gen ailesinin, maya, *C.elegans*, sinekler ve bitkilerden memelilere kadar 68 türde, yüksek derecede aminoasit dizi benzerliği gösteren 100 üyesi vardır. İnsan YPEL

gen ailesi üyeleri, *YPEL1-5*, çinko-bağlanabildiği öngörülen ve moleküler kütleleri 13,500 ile 17,500 Da arasında değişen küçük proteinleri kodlarlar. Ypel proteinlerin yapı ve işleri henüz bilinmemesine rağmen, sınırlı sayıdaki çalışma Ypel proteinlerinin gelişim, hücre döngüsünün ilerlemesi, mitoz, hücre yaşlanması ve ölümü gibi süreçlerde yer aldığını öne sürmektedir. Biyoinformatik analizlerimiz, Ypel proteinlerinin yüksek derecede korunmuş yapısal ve işlevsel özelliklerinin temel hücresel işlevler için önemli olabileceğini göstermiştir. Biyoinformatik analizlerimiz ayrıca her *YPEL* geninin yere ve zamana bağlı bir şekilde, iç ve dış sinyallere yanıt olarak gelişen sinyal yolaklarının aktif hale getirdiği farklı transkripsiyon faktörleri tarafından düzenlendiğini de öne sürmektedir.

Ypel2'nin sentezini ve hücre içi lokalizasyonu karşılaştırmalı olarak incelemek amacıyla öncelikle insan YPEL ailesinin beş üyesinin cDNA'larını meme adenokarsinomasından türetilmiş ER-pozitif MCF7 hücre hattının cDNA kütüphanesini kullanarak klonladık. Sonrasında, ligand bağlanmamış ER'nin *YPEL2*'nin bazal mRNA seviyelerini düzenlediğini gösterdik. Ayrıca, *YPEL2* ve *YPEL3*'ün ifadelerinin E2 ile baskılandığını bulguladık. Bu bulgular *YPEL2* ve *YPEL3*'ün E2 ve ER yanıtlı genler olduğuna dair öngörümüzü güçlendirmiştir. Ypel1, 2 ve/veya 3 ün Afrika yeşil maymunu böbrek fibroblast benzeri hücrelerden türetilmiş COS7 hücrelerinde sentezlendiğini ve hücre çekirdeğinin hemen dışında bir bölgeye lokalize olduğunu bulguladık. Ancak, MCF7 hücrelerinde endojen Ypel proteini tespit edemedik. Öte yandan, hem COS7 hem MCF7 hücrelerinde *YPEL1-5*'in fazla ifade edilmesinin (over-expression) DNA'nın hücre çekirdeğinden her bir Ypel proteinin konumuyla birebir örtüşen bir şekilde sitoplazmaya sızmasına neden olduğunu gözlemledik. MCF7 hücrelerinde ayrıca, *YPEL1-5* fazla ifadesinin nükleer lamina bütünlüğünde bozulmalara neden olduğunu bulguladık.

İleriki çalışmalarımız *YPEL2* ifadesinin düzenlenmesi ve Ypel2'nin hücre modellerindeki işlevleri üzerine yoğunlaşacaktır.

Anahtar kelimeler: öströjen, öströjen reseptörü, *YPEL2,* Yippee

To the mysterious world of Ypels

ACKNOWLEDGEMENTS

I would like to thank my supervisor Assoc. Prof. Dr. Mesut Muyan for holding my hand as I took my first steps in the scientific world and leading me in this journey. He has provided me with many learning opportunities, gaining of which I will hold on to for the rest of my life. I am grateful to him for believing in me and supporting me all the way through.

I would like to thank all my thesis committee members, Assoc. Prof. Dr. A. Elif Erson-Bensan, Assoc. Prof. Dr. Sreeparna Banerjee, Assist. Prof. Dr. Mehmet Somel and Assist. Prof. Dr. Özgür Şahin.

I would like to thank Assoc. Prof. Dr. A. Elif Erson-Bensan, in particular, for welcoming me to her laboratory as if I was one of her own students. Without her generous support, completion of this study within this time frame would not be possible. I also thank Assist. Prof. Dr. Can Özen for his support with cell culture studies, Assoc. Prof. Dr. Çağdaş Devrim Son for his help in fluorescence microscopy and Prof. Dr. Ahmet Koç and his laboratory for their collaboration.

I cannot express enough thanks to the members of ER lab for their help in experiments, support in tough times, kindness and friendship. My heartfelt thanks go to Sırma Damla User, for complementing me for what I am missing, for taking care of me when I most needed it and for making me laugh. I also appreciate her help in construction of oligo tags used in this study. Pelin Yaşar, my best lab partner ever, your helping hand touched my experiments as well as my heart. I thank Gamze Ayaz for her smiling face and one-of-a-kind comments that make my day. I thank Hasan Hüseyin Kazan for our delighted discussions on the roads. I also thank our special student Batuhan Çağrı Yapan for his help in experiments and for our cheerful little talks.

I would like to thank the previous and current members of Erson lab; Cansaran Saygılı, Begüm Akman, Merve Öyken and Nihan Özdemirler, for everything they have taught me and Tuna Çınkıllı and Esra Yavuz for their friendship. I also thank the members of Banerjee lab for their help and for sharing resources.

My dearest friends, Aylin Kömez, Beliz Bediz and Neşe Erdem, it is my privilege to have you in my life. I also thank Armağan Önder Hergüner for always being there for me.

Finally, I thank my family for raising me to be the person I am today, for supporting me in all of my decisions and for their endless love. I will continue to do my best to make you proud.

I would like to thank TÜBITAK for supporting me through National Scholarship Program for MSc Students.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES

LIST OF TABLES

TABLES

CHAPTER 1

INTRODUCTION

1.1 Estrogen signaling

17β-estradiol (E2) as the main circulating estrogen hormone in the blood has critical roles in the regulation of many tissue and organ functions [1]. Breast tissue is one of the target tissues, whose physiological and pathophysiological functions are regulated by circulating E2 levels.

Breast cancer is a complex disease resulting from deregulations of integrated signaling pathways that control cell proliferation, differentiation and death [2]. Despite the contribution of both inherited and environmental factors, perturbations in E2 signaling is considered to be the major factor for the uncontrolled cell growth in the initiation and progression of breast cancer [2], [3].

The effects of E2 in cells are primarily mediated by estrogen receptors α and β (ERα and β) that act as transcription factors. Although ERα and β are distinct gene products and are expressed in different tissues as well as in the same tissue at various levels, they exhibit similar structural and functional characteristics. ERs as other nuclear receptors are modular proteins in that isolated structural domains display subsets of the functional activities of the intact receptor. These domains are referred to as: A/B, C, D, E and F (Figure 1). The A/B domain is involved in protein-protein interactions and ligand-independent transactivation through the so-called the activation function-1 (AF-1). The C region or the DNA binding domain (DBD) consists of two zinc-finger like sub-domains that are primarily responsible for binding of the receptor to DNA. The D or hinge region contains the nuclear

localization signal. The multi-functional E/F region, or the ligand binding domain (LBD), is responsible for the ligand binding, dimerization and ligand-dependent transactivation function, the so called the activation function-2 (AF2) [4].

Figure 1. Schematic diagram of ERs.

ERs immediately after synthesis dimerize and primarily translocate to the nucleus independent of E2, while a fraction of newly synthesized ERs (about 5%) remains in the cytoplasm and/or is transported to the plasma membrane [5], [6]. Upon binding of E2, the LBD of ERs undergoes major structural reorganizations which generate surfaces critical for protein-protein interactions. Thus, E2 binding converts the inactive ERs to the functionally active form [6] that initiates genomic and non-genomic events resulting in target tissue-specific responses.

Genomic E2-ER signaling pathway involves the expression of target genes, leading to new gene expressions and protein synthesis that result in phenotypic changes in cells. E2-ER regulates target gene expressions by the estrogen response element (ERE)-dependent and ERE-independent signaling pathways. The ERE-dependent signaling pathway starts with the binding of ERs to EREs [6]. An ERE sequence consists of two palindromic half sites separated by three non-specific nucleotides. The consensus ERE sequence is defined as: 5'-GGTCAnnnTGACC-3'. However, deviations from the consensus by up to three nucleotides have been demonstrated to be functional [7]. Although the un-liganded ER interacts with ERE, the binding of E2 to ER enhances the stability of ER-ERE interactions [8]. The binding of E2 also results in the structural reorganizations primarily in the LBD of ER to generate protein interaction surfaces for the recruitment of co-regulatory proteins and the subsequent establishment of a transcription complex [4], [8].

In the ERE-independent signaling pathway, the E2-ER complex regulates gene expressions by interacting with and modulating activities of other transcription factors like AP-1 (activator protein 1) and SP-1 (stimulatory protein 1) bound to their cognate response elements [6].

1.2. Yippee-like proteins

The focus of the current study is on the Yippee-like 2 (*YPEL2*) gene, a member of the highly conserved Yippee-like (YPEL) gene family. However, because of the absence of studies on the *YPEL2* gene or its protein to date, the following sections will summarize studies conducted on other Yippee-like proteins to provide sufficient background.

1.2.1. Discovery of Yippee protein

Yippee protein is discovered in 2001 in a yeast two hybrid screen as an interacting protein with *Hyalophora cecropia* (giant silk moth) hemolin. Hemolin is a member of the immunoglobulin super family that binds to lipopolysaccharide (LPS), enhances phagocytosis and activates protein kinase C. In a search to identify hemolin interacting proteins in *Drosophila melanogaster*, hemolin was used as bait in a yeast two hybrid screen and 'Yippee' was identified as a protein that specifically interacts with hemolin. After obtaining the full length Yippee cDNA, the analysis of the deduced amino-acid sequence revealed two cysteine pairs that are 52 amino-acids apart (Cys-X_2-Cys-X_{52}-Cys-X_2-Cys) and predicted to form a zinc binding pocket. Identification of a human partially expressed sequence tag (EST) with homology to the Drososphila Yippee protein led to the cloning of the then-called human Yippee protein, which is now named as Ypel5. Furthermore, other Yippee-related proteins were identified in cellular slime mould, plant, fungi, *Caenorhabditis elegans*, insects, zebrafish and rodents that showed remarkable similarity among species. Cysteine motifs in Yippee were found to be similar to fish zinc finger protein, rat transcription factor KID-1 and to RING (really interesting new gene) finger proteins [9]. In a study where yippee was used as bait in a yeast interaction trap screen of a Drosophila disc cDNA-library, nine proteins were identified to be Yippee interacting proteins (Yips) [10]. NCBI accession numbers of identified Yips and their homologous proteins found in different species together with their functions are listed in Table 1.

Table 1. List of yippee-interacting proteins (Yips) and their homologous proteins.

Yip (Accession number)	Homologous protein	Species	Function
Yip1 (AF195186)	Split ends (spen)	Drosophila	Glia cell migration, axon guidance, RNA binding region
Yip2 (AF195187)	Acetyl-Coenzyme A acetyltransferase 2	Human	Thiolase, degradation of fatty acids
Yip3 (AF195188)	20S proteasome β2 subunit	Drosophila	Subunit of 20S proteasome
Yip4 (AF195185)	No match	-	
Yip5 (AF195189)	Rpn11 (proteasome endopeptidase)	Drosophila	Subunit of regulatory complexes in 26S proteasoemes
Yip6 (AF195190)	Ribosomal protein L5	*B.mori*	Binds to 5S rRNA, interacts with MDM2
Yip7 (AF195191)	Chymotrypsin	*L.cuprina*	Serine-type endopeptidase, degradation of intestinal proteins
Yip8(-)	Elongation factor 1α	Drosophila	Targeting tRNAs to ribosomes
Yip9 (-)	Mitochondrial 16s rRNA	Drosophila	Component of 30S ribosomal subunit

Adapted from [10]

Although the interacting partners of Yippee do not point out to a common function, the presence of a potential zinc binding motif and the interaction with rRNA and homologous RNA binding proteins raise the possibility that Yippee interacts with nucleotides directly or indirectly. Another possibility is that Yippee has proteasome related functions as the interaction of it with proteasomal protein homologues suggests.

1.2.2. Identification of Yippee-like 1 in mouse

The craniofacial development is dependent on the integrated processes that govern the concerted and coordinated growth and fusion of many tissues. Microdeletions in 22q11.2 in human cause craniofacial defects and syndromes including Di George syndrome, Velocardiofacial syndrome and Conotruncal anomalies face syndrome.

These syndromes are manifested as craniofacial malformation, cardiac defects and thymic and parathyroid gland hypoplasia. Chromosome 22 Di George syndrome critical region in human and the syntenic region of the mouse chromosome 16 are thought to contain genes involved in craniofacial development [11]. The fact that *YPEL1* is present on 22q11.2 in human and the mouse ortholog *Ypel1* is on chromosome 16 synteny region suggest that Ypel1 is indeed involved in craniofacial development. Consistent with this, in a study conducted in 2001 to define genes involved in craniofacial development, the mouse yippee like 1 (*Ypel1*) gene was identified as a novel gene involved in craniofacial development. It was found that *Ypel1* is expressed in the ventral half of early embryo including branchial arches from which face develops [11]. Moreover, Ypel1 as a GFP-fusion protein in transfected NIH3T3 mouse fibroblast cell line was observed to localize to the nucleus and to induce alterations in cytoskeletal and adhesion properties of cells resulting in an epithelial-like transition. These findings could underlie the mechanistic effects of Ypel1 to regulate the cellular morphology and differentiation that could be important in the development of craniofacial complex [11]. In keeping with this, *YPEL1* was found to be the highest ranking Di George Syndrome candidate gene in a study to prioritize candidate genes underlying biological processes or diseases. To validate the results, Ypel1 protein levels in zebrafish were knocked down using morpholinos [12]. It was shown that Ypel1 knockdown embryos develop craniofacial defects including underdeveloped jaw, defective pharyngeal arch cartilage formation and reduced mandibular arch [12].

A developmentally important gene is expected to be conserved through evolutionary time. Ypel1 having homologues in flies, plants, *C.elegans* and yeast fulfils this expectation. Aforementioned studies carried out in mouse embryos and cell lines as well as in zebrafish suggest that Ypel1 is involved in craniofacial development. However, the presence of homologues in species wherein craniofacial development is not an issue could underlie the fact that Ypel1 is not particularly involved in craniofacial development rather it has a basic cellular function common to the diverse range of organisms, that was co-opted during craniofacial development in vertebrates. Association of single nucleotide

polymorphism (SNP rs4821217) in *YPEL1* with breast cancer risk in the log-additive model [13] exemplifies this possibility.

1.2.3. Human YPEL gene family

Hosono *et. al*. [14] in 2004 identified the human *YPEL1* on chromosome 22q11.2 based on its homology to mouse *Ypel1*. A further analysis led to the discovery of four human paralogs on four different chromosomes; *YPEL2* (17q23.2), *YPEL3* (16p11.2), *YPEL4* (11q12.1) and *YPEL5* (2q23.1). Hosono *et. al*. also cloned all five genes of the mouse Ypel family [14].

To study the expression of YPEL genes in different tissues of human and mouse, Hosono *et. al.* performed PCR analysis with tissue cDNA libraries [14]. In human, *YPEL1* was found to be expressed only in the testis and fetal brain. *YPEL2* is expressed in the heart, kidney, lung, pancreas, placenta, skeletal muscle, leukocyte, prostate, spleen, testis, fetal brain, fetal heart, fetal kidney, fetal liver, fetal lung, fetal skeletal muscle and fetal spleen. The expression of *YPEL4* is in the brain, lung, placenta, colon, ovary, small intestine, spleen, testis, fetal brain, fetal heart, fetal liver, fetal lung, fetal spleen and bone marrow. On the other hand, *YPEL3* and *YPEL5* appear to be expressed in all tissues examined. In mouse, Ypel1-5 show the same expression patterns as observed in human [14].

The analysis of amino-acid sequences among mouse and human Ypel proteins and paralogs revealed a high sequence homology. Of the human Ypel proteins, Ypel 1-4 have 83.2% identical amino-acid sequence while Ypel5 has the lowest homology to other Ypel proteins (43.8-49.5% amino-acid identity to Ypel1-4). Mouse and human Ypel orthologs share at least 99.2 % amino-acid sequence identity. The identity of human and mouse Ypel1-4 to Drosophila Yippee is between 43.4 and 48.5 %, respectively; while human and mouse Ypel5 are more similar to Drosophila Yippee with 70.8 % amino-acid sequence identity [14].

Hosono *et. al*. [14] have also identified 100 YPEL family genes in 68 species including mammal, bird, amphibia, fish, protochordate, insect, nematode, coelenterate, echinoderm, protozoan, plant, and fungi. Amino-acid sequence alignments based on highly conserved cystine and histidine residues of all identified Ypel proteins revealed a remarkable homology with many identical residues. Thus, a consensus sequence is deduced as follows: C-X_2-C-X_{19}-G-X_3-L-X_5-N-X_{13}-G-X_8-C-X_2-C-X_4-GWXY-X_{10}-K-X_6-E. Interestingly, in the consensus sequence, the number of non-consensus residues, designated as X, is identical for all species examined and they are conserved among closely related species [14]. This suggests that through evolutionary time, as species diverge, YPEL genes acquired some sequence changes, but the core sequence has been conserved even in distantly related species. Therefore, there might be a conserved basic function of the ancestral Yippee related protein but different members of YPEL family may have gained specialized functions. This notable conservation of the amino-acid sequences among Ypel proteins in various species could pose the idea that 1) each Ypel protein has different spatio-temporal expression in response to distinct signaling pathways, which could lead to the execution of the same function, an indication of a functional redundancy; 2) as a result of minor structural differences, each Ypel has a distinct function leading to different cellular responses in a time and space dependent manner.

Although there are five YPEL genes in human, mouse and African green monkey, there are four YPEL genes in birds and fish [15]; three in amphibia; two in nematodes, insects and protochordates [14]. In yeast, there is one YPEL homolog. It is therefore tempting to speculate that increase in the number of YPEL family genes correlates with the complexity of organism.

To begin to understand the function of Ypel proteins, *Hosono et. al.*, initially studied the cellular localization of Ypel proteins [14]. To accomplish this, they generated two antibodies against synthetic peptides corresponding to C-terminal regions of Ypel1 and Ypel5. The antibody based on Ypel1 was suggested to recognize Ypel1-4 while Ypel5 antibody was stated to specifically recognize Ypel5 [14].

However, no supporting information for the specificity of these two antibodies was presented. Indirect immunofluorescent staining of COS7 cells derived from transformed African green monkey kidney fibroblast-like cells, with 'Ypel1 antibody' suggests that Ypel1-4 are nuclear proteins. In interphase cells, Ypel1-4 are localized in nucleoli and the centrosome. In the mitotic phase, Ypel1-4 become localized on or close to the mitotic apparatus rather than in the centrosome [14]. It was further observed that Ypel1-4 are found at different locations in different stages of cell cycle. These observations suggest the involvement of Ypel1-4 in mitotic cell division. Results of immunofluorescent staining with Ypel5 antibody are summarized in Section 1.2.6.

1.2.4. YPEL3

The mouse Ypel3 was initially identified as small unstable apoptotic protein (SUAP) in 2003. The *SUAP* mRNA was shown to be up-regulated during interleukin-3 (IL3) deprivation-induced apoptosis in myeloid precursor cells [16]. Suap was also suggested to be target of ubiqutination and subsequent proteasomal degradation [16].

In 2010, the human *YPEL3* mRNA was shown to increase during DNA damage in p53 positive colorectal carcinoma cell line Hct116, likely through two functional p53 binding sites present on the *YPEL3* gene promoter [17]. In addition, Ypel3 protein levels in Hct116 cells, assessed by western blot (WB) analysis using a commercially available YPEL3 antibody, were suggested to increase in response to DNA damage [17]. The protein detected by this commercial antibody was shown to be lost after transduction of cells with a lentivirus bearing shYPEL3. However, it should be stated that the observed changes in protein levels may include Ypel1-4 proteins rather than Ypel3. This is because, according to the product specification of the company, this Ypel3 antibody recognizes Ypel1-4 proteins. Furthermore, in this study, the shRNA sequence used for silencing of *YPEL3* is not disclosed. Given the high level of sequence homology among YPEL genes, it is also possible that the shRNA used in this study target other YPEL mRNAs. It is therefore likely that DNA damage induces an increase in Ypel1-4 protein levels rather than only that of Ypel3.

Later studies by Tuttle *et al.* suggest that DNA damage augments the YPEL3 gene expression resulting in cellular senescence [17]. In ER-positive MCF7 cell line derived from a breast adenocarcinoma and ER-negative U2OS cells derived from an osteosarcoma, when exogenously genome integrated *YPEL3* was expressed by a tetracycline inducible expression system at levels comparable to endogenous mRNA levels detected upon DNA damage, cells displayed growth inhibition and increased cellular senescence as determined by colony formation assay, β-galactosidase staining and the appearance of senescence associated heterochromatin foci within the nuclei (SAHF) [17].

In addition to DNA damage, Tuttle *et al.* also showed that E2 regulates *YPEL3* expression [18]. When MCF7 cells were grown in the absence of E2, *YPEL3* mRNA levels increase while the addition of E2 at circulating levels (1nM) for 24 hours decreases the expression of *YPEL3* mRNA levels. This repression was reversed by the addition of selective estrogen receptor modulator, tamoxifen (TMX) or the knockdown of ERα by siRNA [18]. Moreover, MCF7 cells undergo cellular senescence in the absence of E2, whereas the repression of Ypel3 synthesis with siRNA rescues cells from senescence [18]. These findings suggest that the E2-ER mediated repression of *YPEL3* expression is critical for the growth of MCF7 cells. Conversely, the treatment of MCF7 cells with TMX induces cellular senescence which is not observed when *YPEL3* is silenced. Although the authors reported changes in Ypel3 protein levels in MCF7 cells correlating with alterations in mRNA levels by treatments [18], the results need to be cautioned due to the limitations of the antibody specificity as discussed above.

Tuttle *et. al.* also reported that the expression of YPEL3 may be associated with carcinogenesis [17], [19]. It appears that the expression of *YPEL3* is repressed in colon tumor samples from commercial sources and patient colon adenocarcinoma [19], commercial and patient ovarian tumors [17], and lung tumor samples from commercial sources [19].

1.2.5. YPEL4

In a yeast two hybrid screen of a human brain cDNA library using Ypel4 as a bait, the major vault protein (MVP) was found to interact with Ypel4 [20]. Ypel4-MVP interaction was confirmed with mammalian two hybrid reporter assay, co-localization, co-immunoprecipitation and GST pull-down assays [20].

The MVP, the main component of vault ribonucleoparticles [21], is a highly conserved protein with 90% amino-acid sequence identity among mammals and 60% identity between mammals and lower eukaryotes [22]. Vault particles are composed of the MVP, the vault poly (ADP-ribose) polymerase (vPARP), the telomerase-associated protein 1 (TEP1) and a number of short untranslated RNAs (vRNA) [21]. Vaults are present in a diverse range of eukaryotes including protozoa, mollusks, the slime mold, echinoderms, fish, amphibians, avians and mammals [22]. Although the cellular function of vaults is yet unclear, vault particles are suggested to be critical factors in chemotherapy resistance and regulation of several cellular processes including transport mechanisms, signal transmissions and immune responses [21] by regulating the activities of proteins involved in phosphoinositide 3-kinase/Akt (PI -3-kinase/Akt), JAK/STAT and MAPK pathways [20]. Elk-1 is a transcription factor regulated by MAPK signaling pathway [20]. In a mammalian two hybrid reporter assay in COS-7 cells, *YPEL4* over-expression was shown to increase the activity of Elk-1, while decreasing *YPEL4* mRNA levels with siRNA led to the suppression of the Elk-1 activity [20]. The same study also showed that the over-expression of *MVP* alone causes a decrease in Elk-1 activity. The over-expression of *MVP* together with *YPEL4* prevents the observed increase in Elk-1 activity when *YPEL4* is over-expressed alone [20]. This result can be interpreted in two ways: 1) the MVP prevents the ability of Ypel4 to activate Elk-1; or 2) the MVP has a dominant repressive effect on the Elk-1 activity even Ypel4 is present.

There appears to be similarities between vault ribonucleoprotein, or MVP in particular, and Ypel proteins, or Ypel4 in particular. Although, the roles are yet unclear, both are highly conserved proteins displaying remarkable homology

among distantly related species. Ypel proteins are suggested to localize in the nucleus, nucleoli, and centrosome. In addition, the Yippee domain of various proteins is thought to be associated with kinetechore and/or microtubule, while MVP was shown to co-purify with microtubules.

Independent studies for the initial characterization of Ypel proteins or vault particles may therefore point to the fact that these yet understudied proteins could have an important basic cellular function.

1.2.6. YPEL5

After the initial characterization of the YPEL family proteins, Hosono *et al.* [14] conducted studies on Ypel5 as a distinct member of the YPEL family proteins. Studies showed that the knock-down of *YPEL5* in COS7 cells results in the suppression of cell growth by extending the durations of G_1 and G_2+M phases. Furthermore, the knockdown of *YPEL5* homolog Ypel-b in medaka fish caused malformations in embryos; embryogenesis was found to be interrupted as a result of the suppression of cell proliferation and induction of apoptosis [15].

To address one underlying mechanism by which Ypel5 could affect cellular proliferation and death, a yeast two hybrid approach was used to screen the protein partner(s) of Ypel5. Using a human brain cDNA library and Ypel5 as the bait, four interacting proteins were identified. These are: Ran binding protein in the microtubule organizing center (RanBPM or RanBP9), cannabinoid receptor interacting protein 1 (CNRIP1), neuro-oncological ventral antigen 1 (NOVA1) and COP9 constitutive photomorphogenic subunit 5 (COPS5) [15]. Further studies showed that the SPRY (SPla and the RYanodine Receptor) domain of RanBPM is required for RanBPM-Ypel5 interactions. It was also demonstrated that Ypel5 interacts with RanBP10, a paralog of RanBPM in COS-7 cells. RanBPM was reported to be localized to the centrosome in interphase while RanBP10 was associated with microtubules [15]. Moreover, Ypel5 was shown to be localized in the nucleus and on the centrosome during interphase using the antibody mentioned in Section 1.2.3. At prophase, Ypel5 was detected at spindle poles, while during

metaphase and early anaphase it became associated with mitotic spindles. At mid-anaphase and telophase, Ypel5 was localized in the spindle midzone. During cytokinesis, Ypel5 was observed only on the midbody. Based on these observations, Ypel5 was suggested to function in cell division and cell cycle progression through interactions with RanBPM and RanBP10 [15].

Other studies also suggest that *YPEL5* is expressed in human peripheral T cells in G_0 stage and it is down-regulated upon activation by immobilized anti-CD3. In addition, the transfection of YPEL5 cDNA into HeLa cells causes a decrease in cellular proliferation [23].

In contrast, Ypel5 as the fusion protein with a truncated PPP1CB was reported to enhance proliferation and colony formation in B-CLL (B-cell chronic lymphocytic leukemia)-related cells [24]. Reciprocal RNA chimeras of *YPEL5* and *PPP1CB* (protein phosphatase 1, catalytic subunit, beta isozyme) genes are recurrently and exclusively detected in chronic lymphocytic leukemia. *YPEL5/PPP1CB* and *PPP1CB/YPEL5* chimeras form as a result of an intergenic splicing event without marks of possible genomic breakpoints. *YPEL5/PPP1CB* RNA chimera results in the formation of a truncated PPP1CB protein which shows a decreased phosphatase activity. This RNA chimeric fusion product acts a dominant negative protein that inhibits the function of wild type PPP1CB protein [24].

In addition, based on results from breast cancer cell models, it was suggested that de-regulation of the *YPEL5* gene expression may be involved in breast cancer. In non-transformed, non-tumorigenic and ER negative MCF-10F cell line derived from human breast epithelium, it was found that the treatment with E2, the concentration of which is not stated, of cells grown in collagen matrix (3D-cultures) induces phenotypical changes indicative of neoplastic transformation manifested by impaired ductal morphogenesis. In these E2-transformed, by yet an unknown mechanism, MCF-10F cells the levels of YPEL5 mRNA was reported to be repressed compared to controls [25].

1.2.7. YPEL2

There is, to our knowledge, no study on *YPEL2* gene expression and regulation as well as Ypel2 protein structure and function.

However, gene analysis studies show that *YPEL2* is localized on 17q23 amplicon, which is an amplified segment associated with breast cancer. Although the importance of the observation is yet unclear, of the 29 known genes in the 17q23 amplicon, only *YPEL2* does not show copy number dependent over-expression [26]. In addition, SNP (rs16943468b) of the *YPEL2* gene is suggested to be significantly associated with breast cancer risk in the log-additive model [13]. These findings suggest that de-regulated expression of *YPEL2* gene could be important in cancer initiation and progression.

1.2.8.Aim of the study

Based on the information available for *YPEL* genes and proteins, we propose that E2-ER responsive gene product Ypel2 could be involved in E2 mediated cellular proliferation, differentiation, senescence and/or death. In this study, we started with cloning and initial characterization of Ypel2 in cell lines.

CHAPTER 2

MATERIALS AND METHODS

2.1. *In silico* analyses

2.1.1. Multiple Sequence Alignment

MUSCLE (**MU**ltiple **S**equence **C**omparison by **L**og- **E**xpectation) tool (http://www.ebi.ac.uk/Tools/msa /muscle/) [27] was used for the alignment of *YPEL1-5* nucleotide and Ypel1-5 amino-acid sequences.

2.1.2. Family/Domain Analyses

To analyze Ypel1-5 and Yippee for prediction of functional units or domains and post-translational modifications; Interpro (http://www.ebi.ac.uk/interpro/) [28], eukaryotic linear motif (ELM, http://elm.eu.org/) [29] and ScanProsite (http://prosite.expasy.org/scanprosite/) [30] tools were used.

2.1.3. Homology modeling

Phyre2 (Protein Homology/analogY Recognition Engine V 2.0) program was used for homology modeling [31].

2.1.4. Prediction of transcription factor binding sites

The Champion ChiP Transcription Factor Search Portal (http://www.sabiosciences.com/chipqpcrsearch.php?app=TFBS) is used to predict binding sites in 20 kb upstream and 10 kb downstream sequences of *YPEL1-5*.

2.1.5. Prediction of ERE sequences

To examine the existence of putative ERE sequences Dragon ERE finder (System for Identification and Interactive Analyses of Estrogen Response Elements in DNA Sequences) version 3 (http://datam.i2r.a-star.edu.sg/ereV3/) was used [32].

2.2. Cell lines and treatments

2.2.1. Cell lines and growth conditions

MCF7, T47D and MCF10A cells were a kind gift from Dr. A. Elif Erson Bensan. HeLa cells were obtained from Dr. Can Özen. MDA-MB-231 and COS-7 cells were a kind gift from Dr. Özgür Şahin.

MCF7, HeLa, MDA-MB-231 and COS-7 cells were maintained in high glucose Dulbecco's modified eagle medium (DMEM) without phenol red (Biochrom AG, Germany, F0475) supplemented with 8% fetal bovine serum (FBS, Biochrom AG, Germany S0115), 1% penicillin streptomycin and 0.5% L-glutamine. T47D cells were maintained in RPMI 1640 without phenol red (Biochrom AG, Germany F1275) supplemented with 8% fetal bovine serum (FBS, Biochrom AG, Germany S0115), 1% penicillin streptomycin and 0.5% L-glutamine. MCF10A cells were grown in DMEM/Ham's F12 medium (Biochrom AG, Germany FG 4815) supplemented with 5% donor horse serum (Biochrom AG, Germany, S 9133), 20 ng/ml EGF, 500 ng/ml hydrocortisone, 100 ng/ml cholera toxin, 10 μg/ml insulin and 1% penicillin/strepromycin.

All cells were grown as monolayer in humidified incubator with 5% $C0_2$ at 37°C and passaged every three days for a maximum of 6 weeks.

2.2.2. Treatments

To test the effects of ligands on *YPEL* expressions, we used E2 (Sigma Aldrich, Germany) and Imperial Chemical Industries (ICI) 182,780, fulvestrant® (Tocris Biosciences,UK) the latter which is considered as the complete antagonist for ERα. For 10^{-9} M E2 or 10^{-7} M ICI treatment, MCF7 cells were grown in respective complete growth media supplemented with charcoal-dextran stripped FBS (CD-FBS) instead of FBS to reduce endogenous steroid hormone content. Protocol for CD-FBS preparation is given in Appendix B.

For RNA isolation from cells treated with E2 and/or ICI, $7.5x10^4$ MCF7 cells were seeded into T25 flasks in DMEM-CD-FBS. After 48 hours, ligands were diluted in fresh DMEM-CD-FBS medium as follows: E2 was added from a 10^{-3} M stock

solution in molecular grade ethanol for the final concentration of 10^{-9} M. ICI was added from 10^{-3} M stock solution in molecular grade ethanol for the final concentration of 10^{-7} M. For vehicle treated cells, ethanol was added in the same amount present in E2 or ICI treated cells. Cells were incubated with ligand containing media for 3, 6 or 24 hours then subjected to RNA isolation (Section 2.3.2).

For protein isolation, MCF7 cells were seeded as $2.5x10^5$ cells/ T75 flask. E2 or ICI treatment was performed as described above. 24h after ligand treatment cells were lysed for protein extraction.

2.3. mRNA expression analysis

2.3.1. Primer design

To ensure that primers to be used in RT-qPCR are unique to each YPEL, RefSeq mRNA sequences obtained from NCBI database were aligned with MUSCLE (**MU**ltiple **S**equence **C**omparison by **L**og- **E**xpectation) tool [27]. Primer sequences were analyzed and assessed for specificity with NCBI Primer Blast [33]. All primers used in this study are listed in Table A 1.

2.3.2. Total RNA isolation

Cells grown for RNA isolation were trypsinized, collected in fresh growth medium, and centrifuged at 250xg for six minutes. Pellets were washed with PBS and centrifuged at 300xg for seven minutes. Pellets re-suspended in PBS were aliquoted into three in 1.5 ml centrifuge tubes and centrifuged at 3000 rpm for five minutes in a microcentrifuge. PBS was aspirated and dry cell pellets were stored at -80°C until RNA isolation.

Total RNA isolation was carried out with QuickRNA Miniprep Kit (ZymoResearch) according to manufacturer's instructions including on-column DnaseI digestion.

Concentration and purity of RNAs were assessed with NanoDrop 2000 (Thermo Scientific, USA).

2.3.3. The control of genomic DNA contamination in RNA samples

To confirm the absence of genomic DNA, 500 ng of isolated RNA was used as template in a PCR reaction with glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) primers: GAPDH_FP: 5'-GGGAGCCAAAAGGGTCATCA-3' and GAPDH_RP: 5'-TTTCTAGACGGCAGGTCAGGT-3. Reactions were carried out with Taq polymerase (Thermo Scientific,USA) using KCl buffer, 2mM $MgCl_2$, 500 nM of each primer and 200 μM of each dNTP. Reaction conditions were as follows: Five minutes initial denaturation at 95°C; 40 cycles of 30 seconds at 95°C; annealing for 30 seconds at 65°C; extension for 30 seconds at 72°C; 10 minutes final extension at 72°C; and a subsequent infinite hold at 4°C. 50 ng genomic DNA was used as a positive control. A representative PCR result is depicted in Figure C. 1.

2.3.4. cDNA synthesis

500 ng RNA was used as template for cDNA synthesis using RevertAid First Strand cDNA synthesis kit (Thermo Scientific, USA) with oligo $(dT)_{18}$ primers according to manufacturer's instructions. Complete reaction conditions are given in Table 2. Reaction were performed with one hour incubation at 42°C and terminated by a five minute incubation at 70°C. Synthesized cDNAs were stored at -80°C.

Table 2. Reaction conditions of cDNA synthesis

Components	**Amount taken**	**Final concentration**
Total RNA	500 ng	-
Oligo(dT)$_{18}$ primer (100 µM 0.5 µg/µl (15 A260 u/ml))	1 µL	5 µM
5X Reaction Buffer (250 mM Tris-HCl (pH 8.3), 250 mM KCl, 20 mM $MgCl_2$, 50 mM DTT)	4 µL	1X
10 mM dNTP mix	2 µL	1 mM each
RiboLock™ RNase Inhibitor (20U/µl) *One unit of RiboLock™ RNase Inhibitor inhibits the activity of 5 ng RNase A by 50%	1 µL	1 U/µl
RevertAid™ M-MuLV Reverse Transcriptase (200U/µL) *One unit of RevertAid™ M-MuLV RT incorporates 1 nmol of dTMP into a polynucleotide fraction (adsorbed on DE-81) in 10 min at 37°C	1 µL	10 U/µl
Nuclease-free water	to 20 µL	-

2.3.5. RT-qPCR analysis

RT-qPCR analyses were performed using FastStart Universal SYBR Green Master mix (Roche Applied Science, Switzerland) with 300 nM final concentration of each primer and four µl of 1:20 diluted cDNA in 20 µl reaction volume using RotorGene 6000 (Qiagen, Germany).

All primer sequences are listed in Table A 1. Primers used in the study

For *YPEL2* (amplicon length: 138 bp) and *YPEL3* (amplicon length: 115 bp), reaction conditions were as follows: incubation at 95° C for 10 minutes, 40 cycles of 30 seconds at 94°C, 30 seconds at 65°C and 30 seconds at 72°C.

pS2/TFF1 (trefoil factor 1) was used as positive control for E2 and ICI treatments. The fold change in *YPEL2*, *3* and *pS2/TFF1* was normalized against ribosomal protein, large, P0 (*RPLP0*) and/or homolog of Pumilio, Drosophila, 1 (*PUM1*)

genes which are proposed to be the most reliable reference genes for normalization of RT-qPCR results in breast carcinomas [34].

For *pS2/TFF1* (amplicon length: 209 bp), reaction conditions were as follows: incubation at 95° C for 10 minutes, 40 cycles of 30 seconds at 94°C, 30 seconds at 55°C and 30 seconds at 72°C.

For *RPLP0* (amplicon length: 191 bp) and *PUM1* (amplicon length: 111 bp), reaction conditions were as follows: incubation at 95° C for 10 minutes, 40 cycles of 30 seconds at 94°C, 30 seconds at 60°C and 30 seconds at 72°C.

For each gene, reactions were run with three technical replicates and experiments were repeated with three biological replicates. For the relative quantification, the reaction efficiency incorporated ΔΔCq formula was used [35]. Mean fold change of three biological replicate ± standard deviation (SD) was used for statistical analysis. In E2 treatments, levels in EtOH treated samples were set to 1 for each time point. E2 treated sample in each time point was compared to the EtOH treated sample in the corresponding time point. In ICI treatments, levels in EtOH treated samples were set to 1. E2, ICI or E2+ICI treated sample was compared to the EtOH treated sample. One tailed paired t-test, with 95% confidence interval was [36] performed using GraphPad Prism version 5.00 for Windows (GraphPad Software, San Diego California USA).

2.4. Cloning of YPEL1-5

2.4.1. Primer design for cloning of YPEL1-5

For cloning of the open reading frames of YPEL1-5, two set of primers were designed for each gene based on the sequences in the NCBI database. For YPEL3, we used transcript variant 1 which has the identical core sequence with transcript variant 2 with extra 5' sequences. The first set of primers consisted of a forward primer from 5'UTR and a reverse primer from 3'UTR designed for RT-qPCR. The second set of primers internal to the first set of primers, or so-called nested primers, contained restriction enzyme sites that were designed to provide flexibility in cloning as depicted in Figure 2.

In the engineering of all constructs, the first ATG was designed to be embedded into the Kozak sequence for efficient translation [37]. Reverse cloning primers were also designed to contain a polyA signal that embeds a stop codon (TAA) in mRNAs.

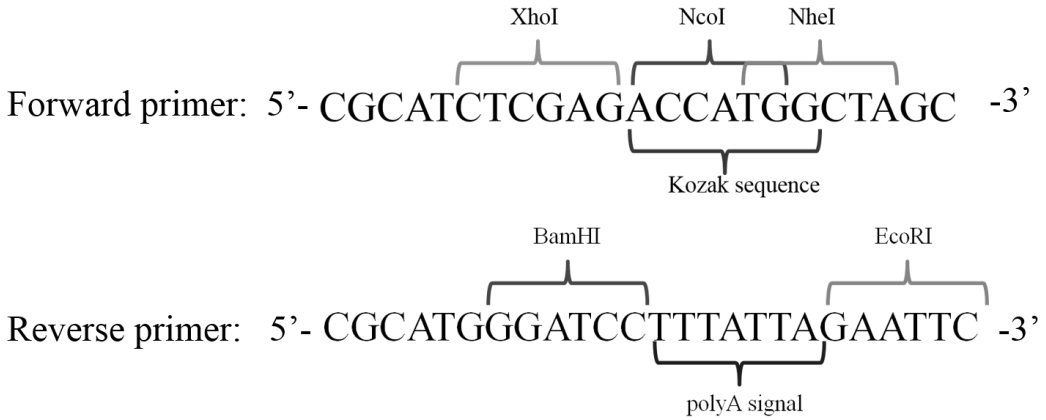


Figure 2: Structure of cloning primers. Forward cloning nested primers are designed to add XhoI, NcoI, NheI cut sites and Kozak sequence to 5' end of ORFs. Each forward primer contains the extensions shown here followed by around 20 nucleotides starting from the nucleotide after the first ATG. Each reverse primer adds EcoRI and BamHI cut sites and polyA signal to 3' ends of ORFs. 3' ends of reverse primers continue with around 20 nucleotides from the end of ORFs, omitting the stop codon sequences.

2.4.2. PCR amplification of YPEL1-5 ORFs from MCF7 cDNA library

The first PCR reactions using 5' UTR forward primers and 3' UTR reverse primers for YPEL1 through 5 were carried out using Phusion High-Fidelity DNA Polymerase (Thermo Scientific, USA) with GC buffer, 500 nM final primer concentration and 200 μM of each dNTP. For YPEL1, 2 and 3, two-step reactions were performed as follows: two minutes initial denaturation at 98°C; 40 cycles of 10 seconds at 98°C; 30 seconds at 72°C; 10 minutes final extension at 72°C; and an infinite hold at 4°C. For YPEL4 and 5, three-step reactions were performed as follows: two minutes initial denaturation at 98°C, 40 cycles of 10 seconds at 98°C, 15 seconds at 55°C and 15 seconds at 72°C, 10 minutes final extension at 72°C followed by an infinite hold at 4°C.

Products were run on 1% agarose gel and fragments with the correct sizes were excised from the gel. The excised DNA fragments were purified using Zymoclean™ Gel DNA Recovery Kit (Zymo Research, USA).

The second PCR were carried out with 50 ng of purified fragments as templates and with cloning nested primers. Reactions were performed using Phusion High-Fidelity DNA Polymerase (Thermo Scientific, USA) with GC buffer, 500 nM final primer concentration and 200 μM of each dNTP. For all five genes two-step cycling conditions were used: two minutes initial denaturation at 98°C; 40 cycles of 10 seconds at 98°C; 30 seconds at 72°C; 10 minutes final extension at 72°C; and an infinite hold at 4°C.

Products of the second PCR were purified using QIAquick PCR Purification Kit (Qiagen, Germany) after assessment of correct sizes on 1% agarose gel.

2.4.3. Cloning of wild type *YPEL1-5* into pBS-KS vector

For cloning of the wild type *YPEL1-5*, PCR purified ORFs with added restriction enzyme cut sites and pBS-KS (-) vector were double digested with XhoI (Thermo Scientific, USA) and BamHI (Thermo Scientific, USA) restriction ezymes using Buffer B (Roche, Switzerland) at 37°C for 1.5 hours. Digested vector was treated with FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific, USA) for 30 minutes at 37°C to prevent vector recircularization.

Double digested PCR products and the vector were run on 1% agarose gel and expected size of DNA fragments were recovered using Zymoclean™ Gel DNA Recovery Kit (Zymo Research, USA).

Ligation reaction was performed using Rapid DNA Ligation Kit (Thermo Scientific, USA) for 15 minutes at room temperature. For each reaction, 3:1 insert: vector molar ratio is used. Ligation mixtures were then transformed into chemically-transformation-competent XL1-Blue cells (Agilent-Stratagene, USA).

Colonies grown in LB agar ampicillin plates were assessed for the presence of inserts by colony PCRs using cloning primers specific for each gene with

abovementioned reaction conditions. Colonies containing inserts were grown in 100 µg/ml ampicillin containing LB medium for overnight at 37°C with shaking. Plasmid isolation was carried out from bacterial cultures using QIAprep Spin Miniprep Kit (Qiagen, Germany). Plasmids were then sent to the METU Central Laboratories for sequencing.

2.4.4. Cloning of *YPEL1-5* into pBS-KS vector with Flag tag

The Flag tag composed of DYKDDDDK amino-acid sequence is widely used in various protein purification and detection applications (Sigma-Aldrich, USA). To obtain Flag-tagged constructs of *YPEL1-5*, pBS-KS (-) bearing the WT cDNA of each gene and pBS-KS(-) vector containing an in-frame flag-tag sequence upstream of the NheI cut site were double digested with NheI (Thermo Scientific, USA) and BamHI (Thermo Scientific, USA) restriction enzymes in 1X Tango Buffer (Thermo Scientific, USA). Ligation, transformation and colony PCR were performed as described above.

2.4.5. Construction of c-myc, HA, StrepII and V5 tagged pBS-KS vectors

In order to have five different tags for five YPEL genes, oligonucleotides for c-myc (EQKLISEEDL), HA (YPYDVPDYA), StrepII (WSHPQFEK) and V5 (GKPIPNPLLGLDST) tags were designed to exchange the in-frame Flag-tag in pBSKS-Flag constructs. Upstream oligonucleotides contain sequences coding for a specific-tag sequence with 5'-CATGG-3' extension at their 5' ends. This sequence carries sequences that would give rise to an NcoI restriction enzyme site following the digestion of the vector with NcoI. NcoI cut site was followed by two additional nucleotides generating a codon for glycine to prevent the formation of a frame shift. The 3'end of upstream oligonucleotides contained a single G nucleotide that would combine with the 5'-CTAGC-3' hangover generated with the NheI digestion of vector. Downstream oligonucleotides were reverse complements of upstream oligonucleotides. The 5' end of downstream oligonucleotides contains a 5'-CTAGC-3' extension, the hangover that would be generated by NheI digestion. The 3'end of downstream oligonucleotides contains a single C nucleotide which would be ligated to 5'-GGTAC-3' hangover in the vector generated by the NcoI digestion.

Table 3. Oligo sequences for c-myc, HA, strepII and V5 tags

Oligo name	Oligo sequences (5'-3')
c-myc_UP	CATGGGTCAGAAGCTGTCTCAGAGGAGGACCTGG
c-myc_DOWN	CTAGCCAGGTCCTCCTCTGAGATCAGCTTCTGACC
HA_UP	CATGGGTTACCCATACGATGTTCCAGATTACGCTG
HA_DOWN	CTAGCAGCGTAATCTGGAACATCGTATGGGTAACC
StrepII_UP	CATGGGTTGGAGCCACCCGCAGTTCGAAAAGG
StrepII_DOWN	CTAGCCTTTTCGAACTGCGGGTGGCTCCAACC
V5_UP	CATGGGTGGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTACGG
V5_DOWN	CTAGCCGTAGAATCGAGACCGAGGAGAGGGTTAGGGATAGGCTTACCACC

Oligos were dissolved in TE buffer (Tris 10mM EDTA 1mM, pH 8.0) as one µg/µl. To anneal oligos, 10 µg of upstream and 10 µg of downstream oligonucleotides were combined in the same tube and the total volume was brought to 50 µl with TE buffer. Reactions were incubated at 95°C on a dry heat block for 5 minutes, the heat block was then unplugged to allow the temperature of the block to gradually decrease to room temperature.

Annealed oligonucleotides were measured for double stranded DNA and ligated to pBS-KS-Flag vector digested with NcoI (Thermo Scientific, USA) and NheI in 1X Tango buffer as described above. Transformed bacterial colonies were assessed with colony PCR for the insertion of tag sequences into the cloning vector using tag-specific upstream oligonucleotides and a vector-specific reverse primer (T7 reverse primer: 5'-GTAATACGACTCACTATAGGGC-3') with following reaction conditions: five minutes initial denaturation at 95°C; 30 cycles of 30 seconds at 95°C; annealing for 30 seconds at 55°C; extension for 30 seconds at 72°C; 10 minutes final extension at 72°C; and a subsequent infinite hold at 4°C.. Plasmids with inserts were then sent to sequencing to the METU Central Laboratories.

2.4.6. Cloning of *YPEL1-5* constructs into pcDNA3.1 (-) vector

Both the WT and Flag-tagged constructs of *YPEL1-5* were double digested with XhoI (Thermo Scientific, USA) and BamHI (Thermo Scientific, USA) restriction

enzymes using Buffer B (Roche, Switzerland) at 37° for 1.5 hours. pcDNA3.1 (-) vector similarly digested was treated with FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific, USA) for 30 minutes at 37°C to prevent vector recircularization.

Each YPEL construct fragment was ligated into 50 ng pcDNA3.1 (-) vector with 6:1 insert:vector ratio. Ligations were performed using Rapid DNA Ligation Kit (Thermo Scientific, USA) for 15 minutes at room temperature. Subsequent transformation and colony PCR reactions were carried out as described in Section 2.4.3.

2.5. Fluorescent microscopy

2.5.1. Transient transfection

For fluorescent imaging studies, MCF7 cells, $6x10^4$ cells/well, were seeded onto round coverslips placed into 12-well tissue culture plates. 48 hours later, transfection was carried out. One μg DNA was diluted in 100 μL phenol red free DMEM high glucose medium. Two μL Turbofect *in vitro* transfection reagent (Thermo Scientific, USA) was added into diluted DNA and the mixture was vortexed briefly. After a 30 minute incubation at room temperature for the formation of transfection complexes, the spend medium in wells was removed and cells were incubated with fresh complete growth medium. Transfection mixture was then added dropwise onto the cells. Four hours later, the spend medium was changed to the complete growth medium.

2.5.2 Immunocytochemistry

18, 24 or 36 hours (indicated for each experiment) after transfection, the medium was removed and cells were washed three times with PBS. Cells were then fixed with freshly prepared 2% paraformaldehyde for 30 minutes at room temperature. All traces of fixative were subsequently removed with three PBS washes. Cells were permeabilized with 0.4% Triton-X in PBS for 10 minutes at room temperature. After three washes with PBS, cells were blocked with 10% normal goat serum (NGS) for S-14 Ypel antibody (Santa Cruz Biotechnology, USA) and

LaminA antibody (ab8980, Abcam, USA), or with 10% bovine serum albumin (BSA) for Flag-M2 antibody (Sigma Aldrich, Germany) and LaminB1 antibody (ab16048, Abcam, USA) for one hour at room temperature with gentle agitation. The blocking solution was removed and primary antibody solutions were added. S-14 Ypel and LaminA antibodies were diluted in 2% NGS, while the Flag-M2 and LaminB1 antibodies were diluted in 3% BSA. The S-14 Ypel antibody was used at 1:50, the Lamin A was used at 1:200, Flag-M2 antibody was used at 1:250, LaminB1 antibody was used at 1:200 dilution. After two hour incubation at room temperature, the primary antibody solution was removed and cells were washed three times with PBS. In double labeling studies, primary antibodies were used sequentially. An Alexa Fluor® 488 conjugated goat anti-mouse (ab150113, Abcam,USA) secondary antibody is used at 1:1000 dilution for the Flag-M2 antibody; Alexa Fluor® 488 conjugated goat anti-rabbit (ab150077, Abcam, USA) secondary antibody was used at 1:1000 dilution for the S-14 Ypel antibody; An Alexa Fluor® 647 conjugated goat anti-rabbit (ab150083, Abcam, USA) secondary antibody was used at 1:250 for the LaminB1 antibody, an Alexa Fluor® 647 conjugated goat anti-mouse (ab150119, Abcam, USA) secondary antibody was used at 1:250 for the LaminA antibody. Secondary antibodies diluted in 2% NGS for Ypel and LaminA or in 3% BSA for Flag-M2 and LaminB1 were added onto the cells. After 30 minute incubation at room temperature in dark, the secondary antibody solution was removed and cells were washed three times with PBS. In double labeling studies secondary antibodies were used sequentially. For labeling of nuclei, DAPI (4', 6-Diamidino-2-Phenylindole, Dilactate, Life Technologies, USA) was used in 300 nM final concentration in PBS. After 10 minute incubation at room temperature in dark, cells were washed three times with PBS. Coverslips were then mounted onto glass slides with a drop of Quick hardening mounting medium (Sigma Aldrich, Germany), edges of coverslips were sealed with nail polish and slides were left to dry in dark. Imaging was carried out with a fluorescence microscope (Leica) in the laboratory of Dr. Çağdaş Son.

2.6. Western Blot

2.6.1. Transfection

For western blotting (WB) studies, MCF7 cells, $1.6x10^4$ cells/well, were seeded into six-well tissue culture plates. 48 hours later, transfection was carried out using two μg DNA diluted in 200 μL phenol red-free DMEM high glucose medium and four μL Turbofect in vitro transfection reagent (Thermo Scientific, USA). After 30 minute incubation at room temperature for the formation of transfection complexes, the spend medium was replaced with the complete growth medium. Transfection mixture was added drop-wise onto the cells. Four hours later, the spend medium was refreshed with the complete growth medium. Cells were maintained for 36h until protein isolation.

2.6.2. Total protein isolation

Total protein isolation was carried out using M-PER Mammalian protein extraction reagent (Thermo Scientific, USA). For cells grown in six-well plates, the spend medium was removed, cells were washed with PBS. 250 μL/well M-PER with protease inhibitor cocktail (Roche, Switzerland) was added directly onto the cells. For cells grown in T-25 or T-75 flasks, cells were trypsinized, collected with growth medium, pelleted at 250xg for six minutes. Cell pellets were washed once with PBS and pelleted again at 300xg for seven minutes. Based on the number of cells, an appropriate amount of M-PER with protease inhibitor cocktail (Roche, Switzerland) was added onto the pelleted cells. After 15 minutes incubation with agitation at room temperature, lysates were transferred into microcentrifuge tubes. Samples were centrifuged at 14,000xg for 10 minutes at 4°C. Supernatants were transferred into new tubes, aliquoted and stored at -80°C.

2.6.3. Nuclear and cytoplasmic protein isolation

For separation of nuclear and cytoplasmic fractions, NE-PER Nuclear and Cytoplasmic Protein Extraction Kit (Thermo Scientific, USA) was used as directed by the manufacturer. For 80-90% confluent cells in T75 tissue culture flasks, CERI-CERII-NER buffers were used in the 200-11-100 μL amounts.

2.6.4. Western Blot

Extracts were denatured in 6x Laemmli buffer at 95^{o}C for 5 min and were separated on a 15% SDS-PAGE. Proteins were transferred onto a PVDF membrane (Roche, Switzerland) using semi-dry transfer with 2 mA/cm^2 of membrane for 70 minutes. Membranes were blocked with 5% skim milk in 0.05 % TBS-T (Tris Buffered Saline- Tween) for the S-14 Ypel, beta actin and HDAC antibodies for one hour at room temperature. For the Flag-M2 antibody, blocking was done with 5% BSA in 0.1 % TBS-T overnight at 4°C. The S-14 Ypel antibody was used at 1:100, beta actin antibody (Abcam, USA) was used at 1:15000, the HDAC (Abcam, USA) antibody was used at 1:5000 and Flag-M2 antibody was used at 1:1000 dilution in respective blocking solutions. Membranes were incubated with the Flag-M2, beta actin or HDAC antibodies for one hour at room temperature. The S-14 Ypel antibody was incubated overnight at 4°C. A horseradish peroxidase (HRP) conjugated goat anti-rabbit secondary antibody (Santa Cruz Biotechnology, USA) was used at 1:2500 dilution for the S-14 Ypel, beta actin and HDAC antibodies; An HRP conjugated goat anti-mouse secondary antibody (Santa Cruz Biotechnology, USA) was used at 1:2500 dilution for the Flag-M2 antibody. Secondary antibody dilutions in blocking solutions were incubated for one hour at room temperature. Protein bands were visualized using an enhanced chemiluminescence (ECL) kit (Clarity Western ECL Substrate, BioRad, USA) used with 1:3 luminol-enhancer reagent to peroxide reagent ratio. Developed blots were imaged with ChemiDoc™ MP system (Biorad, USA) and generated images were analyzed with Image Lab 5.1 (Biorad, USA). Precision Plus Protein™ WesternC™ (Biorad, USA) standards were used as molecular weight marker.

It should be noted that, in the process of choosing an antibody for detection of Ypel protein(s) in ICC and WB studies, we have examined the product datasheets of antibodies provided by several manufacturers. We chose S-14 Ypel antibody provided by Santa Cruz Biotechnology as it was the only antibody which was shown to detect a protein band around the expected molecular masses of Ypel proteins.

2.7. Generation of stable cell line

For generation of stable cell lines, following experimental groups were utilized: 1) un-transfected cells grown in normal medium; 2) un-transfected cells grown in G418 containing medium; 3) transfected cells grown in normal medium; 4) and transfected cells grown in G418 containing medium. MCF7 cells, $1.6x10^4$ cells/well, were seeded into six-well tissue culture plates. 24 hours later, transfection was carried out using one μg DNA diluted in 100 μL phenol red free DMEM medium with high glucose and two μL Turbofect transfection reagent (Thermo Scientific, USA). After a 30 minute incubation at room temperature for the formation of transfection complexes, the spend medium was changed to the fresh complete growth medium. The transfection mixture was added dropwise onto the cells. 4 hours later, the spend medium was refreshed with the complete growth medium. 48 hours later, 1000 μg/ml G418 (>700 μg active compound) containing medium was added onto wells. Medium changes were carried out every three days. After establishing that transfected cells grown in non-selective medium expend in number and un-transfected cells grown in G418 containing media are dead (7-9 days), while colonies in transfected wells became visible, we employed colony selection. After reaching a size visible with naked eye, individual colonies were picked with a one mL pipette tip fitted onto one mL pipette and transferred into wells of a 48 well plate for cell expansion. The remaining cells in wells after monoclone selection were trypsinized and transferred to T25 flask for polyclonal expansion. Monoclonal and polyclonal cell populations were then grown for storage in liquid nitrogen, genomic DNA and RNA isolation as well as for the extraction of nuclear and cytoplasmic proteins.

Genomic DNA isolation was carried out using Quick-gDNA™ MiniPrep kit (ZymoResearch) following manufacturer's instructions. 50 ng of genomic DNA of YPEL2-polyclonal MCF7 cells was used as template for PCR using pcDNA3.1 (-) specific primers: T7_FP: 5' TAATACGACTCACTATAGGG-3' and BGH_REP: 5' TAGAAGGCACAGTCGAGGC-3'. Reactions were carried out with Taq polymerase (Thermo Scientific, USA) using KCl buffer, 2mM $MgCl_2$, 500 nM of each primer and 200 μM from each dNTP. Reaction conditions were as follows:

five minutes initial denaturation at 95°C; 40 cycles of 30 seconds at 95°C; annealing for 30 seconds at 55°C; extension for 30 seconds at 72°C; 10 minutes final extension at 72°C, which was followed by an infinite hold at 4°C. 50 ng MCF7 genomic DNA and 50 ng pcDNA-Flag-YPEL2 vector were used as negative and positive controls, respectively.

RNA isolation, genomic DNA contamination control and cDNA synthesis were accomplished as described in sections 2.3.2., 2.3.3. and 2.3.4. cDNA synthesized from YPEL2-polyclonal MCF7 cells was used as template for PCR reaction using a forward primer that anneal to Flag sequence (Flag_FP: 5'-GATTACAAGGATGACGACGATAAG-3') and a reverse primer within the ORF of YPEL2 (Transfected YPEL2_REP: 5'-GTCCTTGACTTCCTTGGAATGACT-3').

CHAPTER 3

RESULTS AND DISCUSSION

3.1. *In silico* analyses

The structural and functional studies on members of the YPEL family of genes are limited. Due to the high level of conservation of nucleotide sequences, it appears that Ypel proteins share common properties and functions. In order to better understand the features of Ypel proteins, we initially performed extensive *in silico* analyses of genome, transcript and protein structures using publically available bioinformatics tools.

3.1.1. Gene, transcript and protein properties of *YPEL*1-5

According to data available in NCBI gene, nucleotide and protein databases, information regarding gene, transcript and protein structures of the human *YPEL* family is depicted in Figure 3 and summarized in Table 4.

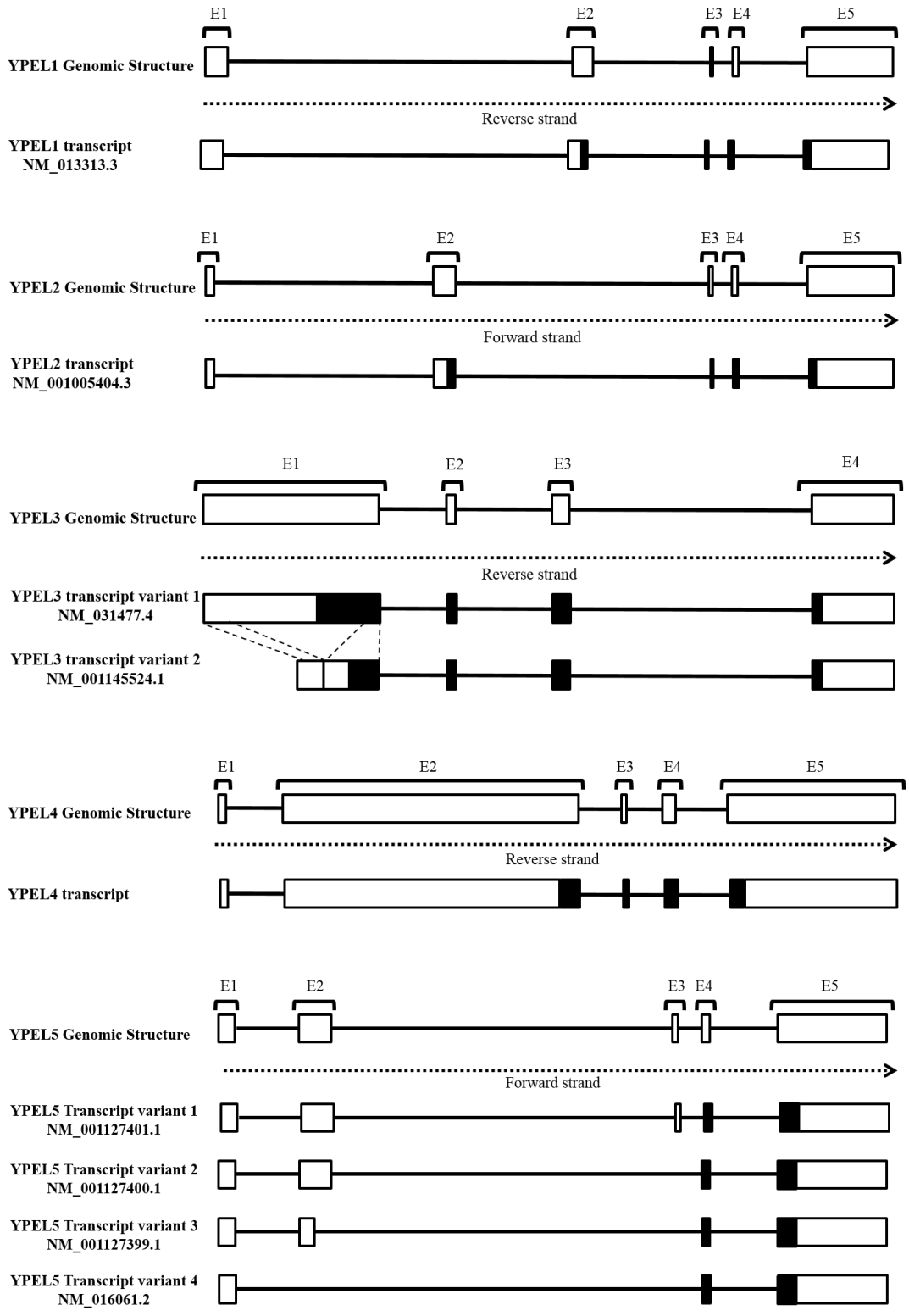


Figure 3. Gene and transcript organizations of YPEL1-5. Boxes represent exons (**E**); black boxes are coding exons, and white boxes are noncoding exons. Lines connecting the boxes represent introns.

Table 4. Characteristics of human *YPEL*1-5 gene, transcript and proteins.

	RefSeq Gene ID **Size of gene (kb)**	**Location**	**RefSeq Transcript Accession number** **Transcript size (bp)**	**Protein Accession number; Protein length, aa**
YPEL1	29799; 38.29	22q11.2	NM_013313.3; 4301	NP_037445.1; 119
YPEL2	388403; 70.03	17q23	NM_001005404.3; 5242	NP_001005404.1; 119
YPEL3	83719; 3.9	16p11.2	NM_031477.4;1588	NP_113665.3; 157
			NM_001145524.1; 940	NP_001138996.1; 119
YPEL4	219539; 4.86	11q12.1	NM_145008.2; 1731	NP_659445.1; 127
YPEL5	51646; 13.59	2p23.1	NM_001127401.1; 2639	NP_001120873.1; 121 NP_001120872.1; 121 NP_001120871.1; 121 NP_057145.1; 121 (identical proteins)
			NM_001127400.1; 2,578	
			NM_001127399.1; 2,342	
			NM_016061.2; 2,281	

Information retrieved from NCBI databases.

Figure 3 demonstrates diagrams of the gene and the transcript organization of human *YPEL1-5* depicted according to NCBI reference GRCh38 primary assembly (http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/). Although *YPEL1-5* encode small proteins with molecular masses ranging between 13,500 to 17,500 Da (Table 4), the sizes of genes are large, primarily for *YPEL1* and *2* due to the presence of long intron and non-coding exon sequences. *YPEL1* and *2* have five exons, one of which is noncoding and the translation initiation codon is located at the end of the second exon. Gene organizations of *YPEL3*, *4* and *5* are somewhat different possibly due to the divergence from *YPEL1* and *2* earlier in evolution.

3.1.2. Primer design for RT-qPCR

Given the high similarity between *YPEL1-5* nucleotide sequences, it was of profound importance to design primers that would bind only to a specific *YPEL* gene sequence. For this purpose, MUSCLE Multiple Sequence Alignment Tool (http://www.ebi.ac.uk/Tools/msa/muscle/) [27] was used to align the open reading frame (ORF) sequences of *YPEL1-5* (Figure 4). The MUSCLE result of the alignment of the ORF sequences demonstrates a high level of sequence similarity among the members of the *YPEL* gene family. This alignment prevented us from designing primers within the ORF sequence of a specific *YPEL* that would permit the differentiation of each *YPEL* gene expression from the other members of the family.

As the next step, we aligned the 3' UTR sequences of *YPEL1-5* using MUSCLE tool once again. The 3' UTR sequences were found to be more readily differentiable compared to the ORF sequences. Figure 5 shows the alignment of the first 300 nucleotides of the 3'UTR sequences of RefSeq mRNAs including transcript variants listed in NCBI Nucleotide database (http://www.ncbi.nlm.nih.gov/nucleotide). Primers for each *YPEL* were designed from sequences with a low similarity to other *YPEL* genes. For *YPEL3* and *YPEL5*, the 3' UTRs of transcript variants are identical to each other, thus primers designed for *YPEL3* and *YPEL5* could recognize whichever variant is expressed without differentiating between them.

Primers were then analyzed with NCBI Primer BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) against the human RefSeq mRNA collection to make sure that they do not have any unintended target.

Figure 4. MUSCLE results of ORF sequences of YPEL1-5. ORF sequences were aligned and output was generated in HTML format. Green and gray highlights indicate the average BLOSUM62 score of pairs of letters in the column is ≥3 or ≥0.2, respectively.

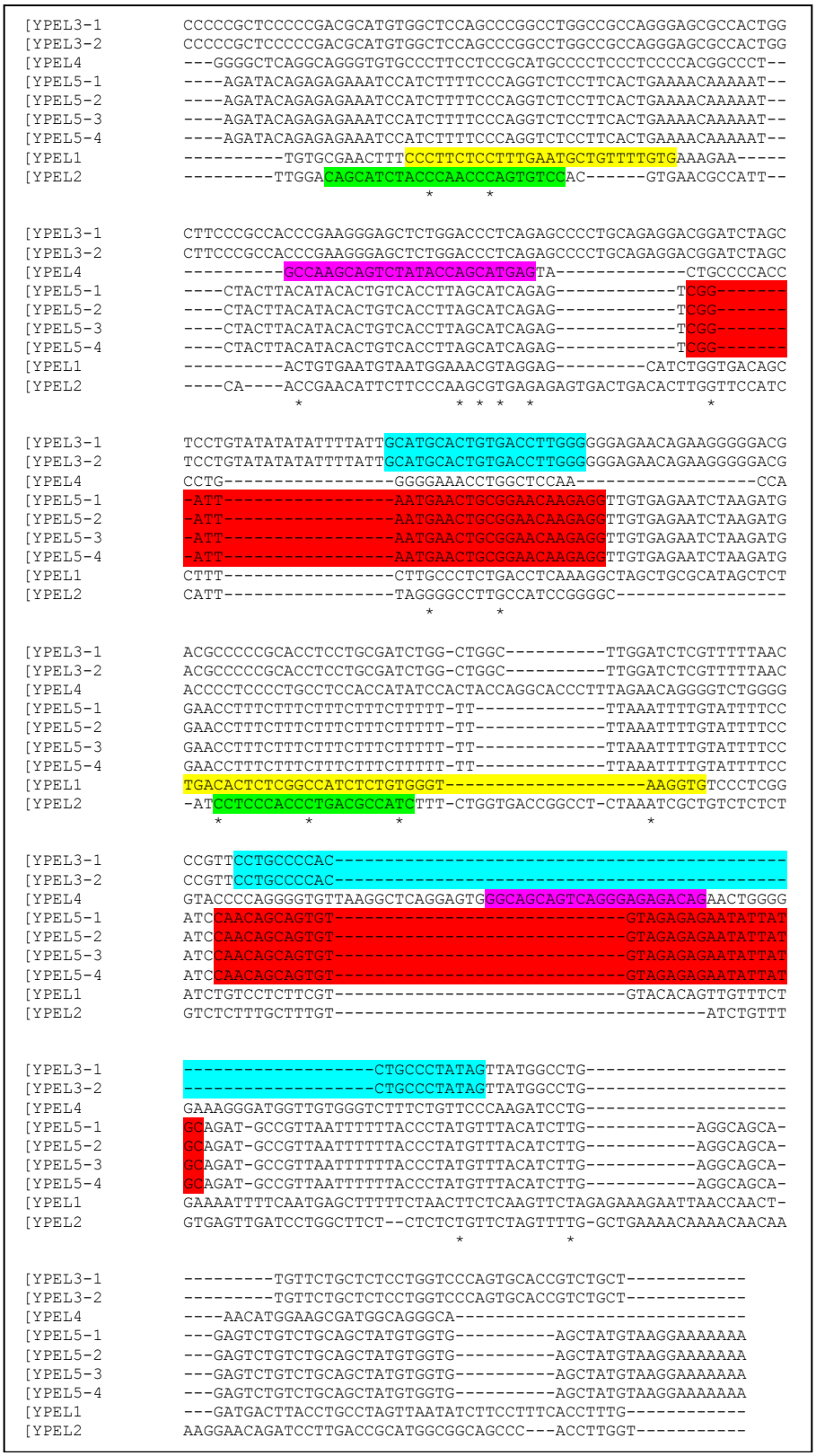


Figure 5. MUSCLE results of the 3'UTR sequences of YPEL1-5. Alignment of 300 nucleotides starting immediately from the 3' of the stop codon sequences of *YPEL1-5* RefSeq mRNAs is given in ClustalW format. *YPEL3-1* and *YPEL3-2* represent *YPEL3* transcript variants *1* and *2*, respectively. *YPEL5-1*, *YPEL5-2*, *YPEL5-3* and *YPEL5-4* represent *YPEL5* transcript variants *1*, *2*, *3* and *4* respectively. Sequences chosen for primer design are highlighted in yellow, green, light blue, pink and red for *YPEL1*, *2*, *3*, *4* and *5*, respectively.

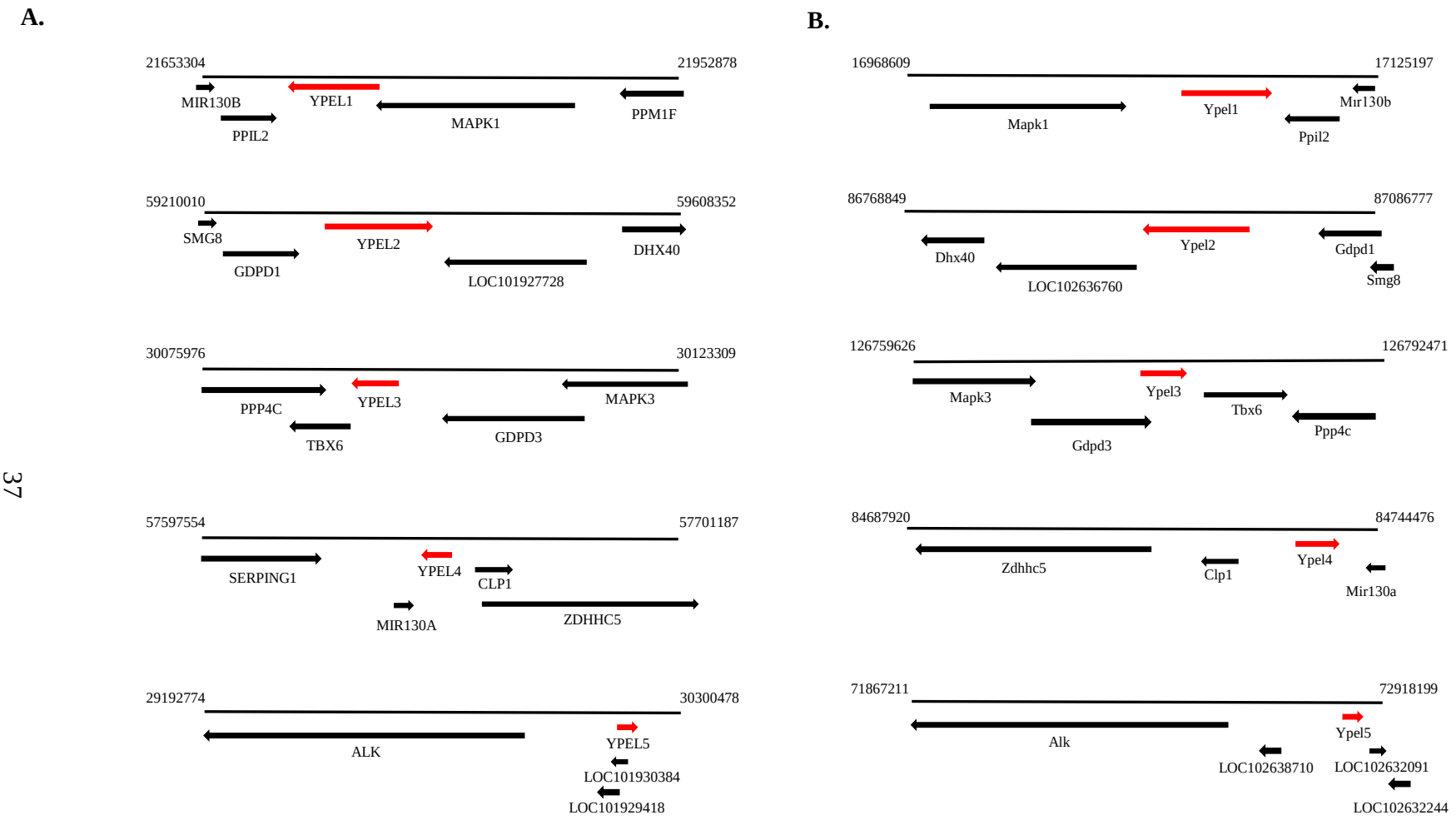


Figure 6. Local chromosome orders of human and mouse *YPEL* family genes. Based on the data available in NCBI Gene database, human *YPEL* (A) and mouse *Ypel* (B) genes (red arrows) and neighboring genes (black arrows) are depicted. Direction of arrows indicates genes that are on the leading or complementary strand. Chromosomal positions for each represented region are given.

3.1.3. Local chromosome orders

Figure 6 depicts the neighboring genes of the human and mouse *YPEL* family based on NCBI gene database (http://www.ncbi.nlm.nih.gov/gene). Strikingly, among the human and mouse orthologs, most of the neighboring genes are identical and in the same order. This implies that the chromosomal regions where *YPEL* genes are located are syntenic between human and mouse. It should be noted that with the exception of YPEL5, the human *YPEL1-4* and mouse *Ypel 1-4* and neighboring genes are reversely oriented.

3.1.4. Amino-acid sequence alignments of Ypel1-5

```
Ypel5           -------------------------------------------MGRIFLDHIGGT-RLFS
Ypel4           ------------------------------MPSCDPGPGPACLPTKTFRSYLPRCHRTYS
Ypel1           -------------------------------------MVKMTKSKTFQAYLPNCHRTYS
Ypel2           -------------------------------------MVKMTRSKTFQAYLPSCHRTYS
Ypel3           MCVAQVLTAHLLPPRQALGSLCSPWAAPRVGPLPPAPAMVRISKPKTFQAYLDDCHRRYS
                                                      . *  ::    * :*

Ypel5           CANCDTILTNRSELISTRFTGATGRAFLFNKVVNLQYSEVQDRVMLTGRHMVRDVSCKNC
Ypel4           CVHCRAHLAKHDELISKSFQGSHGRAYLFNSVVNVGCGPAEQRLLLTGLHSVADIFCESC
Ypel1           CIHCRAHLANHDELISKSFQGSQGRAYLFNSVVNVGCGPAEERVLLTGLHAVADIYCENC
Ypel2           CIHCRAHLANHDELISKSFQGSQGRAYLFNSVVNVGCGPAEERVLLTGLHAVADIYCENC
Ypel3           CAHCRAHLANHDDLISKSFQGSQGRAYLFNSVVNVGCGPAEERVLLTGLHAVADIHCENC
                * :* : *::..:***. * *: ***:***.***:  . .::*::*** * * *: *:.*

Ypel5           NSKLGWIYEFATEDSQRYKEGRVILERALVRESEGFEEHVPSDN
Ypel4           KTTLGWKYEQAFETSQKYKEGKYIIEMSHMVKDNGWD-------
Ypel1           KTTLGWKYEHAFESSQKYKEGKFIIELAHMIKDNGWE-------
Ypel2           KTTLGWKYEHAFESSQKYKEGKYIIELAHMIKDNGWD-------
Ypel3           KTTLGWKYEQAFESSQKYKEGKYIIELNHMIKDNGWD-------
                ::.*** ** * * **.****. *:*   : :.:*::
```

Figure 7. MUSCLE results for amino-acid sequences of Ypel1-5 Alignment of amino-acid sequences of Ypel1, 2, 3 isoform 1, 4 and 5 are given in ClustalW format.

3.1.5. Family/ Domain analyses

Interpro [28] (http://www.ebi.ac.uk/interpro/) is a resource used for classifying proteins into families and predicting the presence of domains using information available in several databases. Searches carried out with Interpro revealed the same results for all Ypel proteins as shown in

Table 5. In Pfam database, proteins that contain a domain called 'Yippee zinc-binding/DNA-binding/Mis18, centromere assembly domain' include Yippee like proteins and Mis18 kinetochore proteins. This family contains 862 sequences displaying 10 different architectures in 301 species. These proteins particularly Mis18 are involved in the priming of centromeres for recruiting CENP-A [38].

Table 5 Interpro results of Ypel1-5

	Position	**Domain/Other matches**	**Database**	**ID**
Ypel1	20-111	Yippee-Mis18 domain	Pfam	PF03226
	1-119	FAD NAD Binding Oxidoreductases	PANTHER	PTHR13847
Ypel2	20-111	Yippee-Mis18 domain	Pfam	PF03226
	1-119	FAD NAD Binding Oxidoreductases	PANTHER	PTHR13847
Ypel3	59-150	Yippee-Mis18 domain	Pfam	PF03226
	39-157	FAD NAD Binding Oxidoreductases	PANTHER	PTHR13847
Ypel4	28-120	Yippee-Mis18 domain	Pfam	PF03226
	7-127	FAD NAD Binding Oxidoreductases	PANTHER	PTHR13847
Ypel5	14-106	Yippee-Mis18 domain	Pfam	PF03226
	1-121	FAD NAD Binding Oxidoreductases	PANTHER	PTHR13847

The eukaryotic linear motif (ELM, http://elm.eu.org/) [29] resource is used to identify short linear motifs (SLiMs). The database contains 200 different motif classes with over 2400 experimentally validated instances manually curated from 2000 scientific publications. SLiMs or linear motifs are defined as compact, degenerate and convergently evolvable interaction modules found in disordered regions. SLiMs are important for protein-protein interactions and protein function. Linear motifs are demonstrated to direct processes including control of cell cycle progression, tagging proteins for proteasomal degradation, modulating the efficiency of translation, targeting proteins to specific sub-cellular localizations and stabilizing scaffolding complexes [29].

Results of ELM motif search for Ypel1-5 and Yippee are represented in Table 6.

Table 6 ELM results

Elm name	Elm Description	Ypel1	Ypel2	Ypel3	Ypel4	Ypel5	Yippee
CLV_PCSK_SKI1_1	Subtilisin/kexin isozyme-1 (SKI1) cleavage site	65-69	65-69	65-69	73-77	3-7 41-45 59-63 98-102	
DEG_APCC_DBOX_1	An RxxL-based motif that binds to the Cdh1 and Cdc20 components of APC/C thereby targeting the protein for destruction in a cell cycle dependent manner	64-72	64-72	64-72	72-80	2-10	2-10
DOC_CYCLIN_1	Substrate recognition site that interacts with cyclin and thereby increases phosphorylation by cyclin/cdk complexes.	65-68	65-68	65-68	73-76	103-106	
DOC_MAPK_1	MAPK interacting molecules carry docking motif that help to regulate specific interaction in the MAPK cascade.	101-109	101-109			93-101	46-53 93-101
LIG_BRCT_BRCA1_1	Phosphopeptide motif which directly interacts with the BRCT (carboxyl-terminal) domain of the Breast Cancer Gene BRCA1 with low affinity	6-10 37-41	6-10 37-41	37-41	45-49	31-35 108-112	31-35
LIG_LIR_Gen_1	Canonical LIR motif that binds to Atg8 protein family members to mediate processes involved in autophagy.	20-24	20-24		28-32		52-56 102-106

Table 6. (continued)

Elm name	Elm Description	Ypel1	Ypel2	Ypel3	Ypel4	Ypel5	Yippee
LIG_SH2_STAT5	STAT5 Src Homology 2 (SH2) domain binding motif.	13-16 49-52 78-81	13-16 49-52 78-81 105-108	13-16 49-52 105-108	21-24 57-60 113-116		43-46
LIG_SH2_SRC	Src-family Src Homology 2 (SH2) domains binding motif.	90-93	13-1-16 90-93				
LIG_SUMO_SBM_1	Motif that mediates binding to SUMO proteins non-covalently.	66-69	66-69	66-69	74-77	99-102	48-52 99-102
LIG_SUMO_SBM_2	Inverted version of LIG_SUMO_SBM_1 that mediates binding to SUMO proteins non-covalently.	53-57	53-57	53-57	61-65		
MOD_CK1_1	CK1 phosphorylation site	-	-		89-95		
MOD_GSK3_1	GSK3 phosphorylation recognition site	2-9 37-44	2-9 13-20 37-44	2-9 37-44	45-52 74-81 86-93	9-16 25-32 29-36	25-32 29-36 111-118
MOD_NEK2_1	NEK2 phosphorylation motif with preferred Phe, Leu or Met in the -3 position to compensate for less favorable residues in the +1 and +2 position.	4-9 41-46 50-55 94-99	4-9 41-46 50-55 94-99	41-46 50-55 94-99	49-54 58-63 74-79 86-91 102-107	30-35	30-35 51-56
MOD_NEK2_2	NEK2 phosphorylation motif with specific set of residues in the +1 and +2 position to compensate for less favorable residues in the -3 position.	-	-		14-19		
MOD_OFUCOSY	Site for attachment of a fucose residue to a serine.	17-23	17-23	17-23	25-31		

Table 6. (continued)

Elm name	Elm Description	Ypel1	Ypel2	Ypel3	Ypel4	Ypel5	Yippee
TRG_ENDOCYTIC_2	Tyrosine-based sorting signal responsible for the interaction with mu subunit of AP (Adaptor Protein) complex	21-24	21-24		29-32-	53-56	103-106
TRG_LysEnd_APsAcLL_1	Sorting and internalization signal found in the cytoplasmic juxta-membrane region of type I transmembrane proteins. Targets them from the Trans Golgi Network to the lysosomal-endosomal-melanosomal compartments. Interacts with adaptor protein (AP) complexes	63-68	63-68	63-68	71-76		
LIG_TRAF2_1	Major TRAF2-binding consensus motif.	61-64	61-64	61-64			87-90
DOC_USP7_1	The USP7 NTD domain binding motif variant based on the MDM2 and P53 interactions.	93-97	93-97	93-97			117-121
LIG_14-3-3_3	Consensus derived from reported natural interactors which do not match the Mode 1 and Mode 2 ligands.	6-11 19-24	6-11 19-24				
CLV_NRD_NRD_1	N-Arg dibasic convertase (NRD/Nardilysin) cleavage site.	-	-	19-21			
MOD_PKA_1	Main preference for PKA-type AGC kinase phosphorylation.	-	-	19-25			

Table 6. (continued)

Elm name	Elm Description	Ypel1	Ypel2	Ypel3	Ypel4	Ypel5	Yippee
DOC_PP1_RVXF_1	Protein phosphatase 1 catalytic subunit (PP1c) interacting motif binds targeting proteins that dock to the substrate for dephosphorylation.					77-83	77-83
LIG_PTB_Apo_2	Phosphorylation-independent motifs that bind to Dab-like PTB domains.					47-54	47-54
LIG_PTB_Phospho_1	Phosphorylation-dependent motif that binds to Shc-like and IRS-like PTB domains.					47-54	
LIG_TYR_ITIM	ITIM (immunoreceptor tyrosine-based inhibitory motif).					51-56	
MOD_N-GLC_1	Generic motif for N-glycosylation					25-30	22-27 25-30 49-54
TRG_PEX_1	Wxxx[FY] motifs present in N-terminal half of Pex5 bind to Pex13 and Pex14 at peroxisomal and glycosomal membranes to facilitate entrance of PTS1 cargo proteins into the organellar lumen.					82-86	82-86
MOD_PIKK_1	(ST)Q motif which is phosphorylated by PIKK family members.	41-47 94-100	41-47 94-100	41-47 94-100	102-108	88-94	25-31 88-94
MOD_PKA_2	Secondary preference for PKA-type AGC kinase phosphorylation.	-	-	2-8 19-25		33-39 106-112	33-39

ELM results allow some predictions common for Ypel1-5 and Yippee to be made. Firstly, it appears that the human Ypel1-5 share functions conserved in the Drosophila Yippee. This suggests that the functional features of the ancestral protein have been conserved in all five Ypel members that could be indispensable for cellular processes. Secondly, we also observed some distinct motifs unique to Ypel5 and Drosophila Yippee. As Ypel5 is the most similar among Ypel1-5 to Drosophila Yippee, this finding supports the idea that Ypel5 is closer to the ancestral Yippee related protein. This in turn suggests that the ancestral protein diverged into two main branches: one giving rise to Ypel5 and other giving rise to Ypel1-4. Among Ypel1-4, Ypel3 appears to contain some distinct motifs. Since SLiMs are implicated in protein-protein interactions, even minor changes in amino-acid sequences could result in specialized molecular interaction networks that result in distinct functions.

Prosite (http://prosite.expasy.org/) is a database of protein families and domains. ScanProsite [30] tool was also used to scan sequences against Prosite collection of motifs to identify protein domains, families and functional sites as well as associated patterns and profiles.

ScanProsite analysis for Ypel1-5 and Yippee are depicted in Table 7, wherein amino-acids are numbered.

Table 7 ScanProsite results of Ypel1-5 and Yippee

	Ypel1	**Ypel2**	**Ypel3**	**Ypel4**	**Ypel5**	**Yippee**
N-myristoylation	58-63	58-63	96-101	66-71		77-82
PKC phosphorylation	97-99	97-99	135-137	17-19 89-91 105-107	25-27 32-34 39-41 63-65 72-74 91-93	25-27 32-34 39-41 63-65 91-93
Tyrosine kinase phosphorylation		99-105	137-143	107-113		
cAMP/cGMP-dependent protein kinase phosphorylation			57-60			
Casein kinase II phosphorylation				81-84		107-110
N-glycosylation					26-29	23-26 26-29 50-53

Results imply that each Ypel protein can be post-translationally modified in response to different signaling cascade affecting protein function. In accord with our previous motif and functional unit predictions, post-translational modification predictions also point out that Ypel1-4 may have similar processing while Ypel5 and Yippee have common modifications that are not shared by Ypel1-4.

3.1.6. Homology modeling

In order to predict tertiary (3D) structures of the human Ypel1-5 and Drosophila Yippee, we used Phyre2 program (Protein Homology/analogY Recognition Engine V 2.0; http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) [31].

Phyre2 uses the alignment of hidden Markov models via HHsearch, which is an open-source software program for protein sequence and structure prediction, to significantly improve accuracy of alignment and detection rate. Phyre2 also incorporates a new *ab initio* folding simulation called Poing to model regions of proteins with no detectable homology to known structures.

Similarities among Ypel1-5 and with Yippee have so far been demonstrated at the primary sequence level. Since functionality of a protein is directly correlated with its conformation, a comparative analysis of 3D structures of Ypel1-5 and Yippee could help to delineate whether the high homology observed in amino-acid sequences is also reflected in 3D structures of the Ypel family proteins. Since homology modeling based on similarity to other known proteins is one approach in predicting 3D structures, we decided to use Phyre2 to model a 3D structure of, and consequently to propose a putative function for, Ypel proteins.

Phyre2 results for Ypel1-5 and Yippee gave the same PDB headers, namely oxidoreductase, Mss4-like and hydrolase with >90 confidence (Table 8) meaning that Ypel1-5 and Yippee display 3D structural similarities to proteins belonging to these three groups with at least 90% probability.

Oxidoreductase header includes methionine-r-sulfoxide reductases (MsrA and MsrB) and zinc-binding methionine sulfoxide reductase. Methionine sulfoxide reduction, carried out by methionine-R-sulfoxide reductases MsrA and Msr-B, is an important biochemical process that protects cellular proteins from oxidative damage [39] during neurodegeneration [40] and aging [41].

Mss4-like header contains the evolutionarily highly conserved Mss4 (mammalian suppressor of Sec4) superfamily of proteins. Mss4 shares structural similarity to nucleotide exchange factors. Mss4, acting primarily as guanine exchange factor, plays a role in protecting cells against programmed cell death in response to stress stimuli [42].

Table 8. Phyre2 results of Ypel1-5 and Yippee. PDB headers for high confidence (>90) templates. Numbers represent the lowest and highest confidence values for each PDB header member**.**

PDB header	Ypel1	Ypel2	Ypel3	Ypel4	Ypel5	Yippee
Oxidoreductase	90.5-95.4	93.6-95.6	90.1-95.3	91.3-94.7	90.8-91.8	90.4-92.9-
Mss4-like	93.6-95	94.4-95.3	93.5-94.3	91.8-92.3	90.9	90.4-90.8
Hydrolase	91.2-92.8	90-92.8	93.5	91.3-93.4	90.9-91.4	92.2

Hydrolase header contains the probable ATP-dependent RNA helicase DHX58, interferon-induced helicase c domain-containing protein (mda5) and retinoic acid inducible protein (RAI). Helicases are nucleic acid-dependent ATPases that unwind DNA or RNA. Helicases are involved in DNA replication and repair, transcription, translation, ribosome synthesis, RNA maturation and splicing, and nuclear export processes. RNA virus infection is recognized by retinoic acid-inducible gene (RIG)-I-like receptors (RLRs), RIG-I, and melanoma differentiation-associated gene 5 (MDA5) in the cytoplasm [43]. RLRs function as sensors for the detection of double-stranded viral RNA to initiate antiviral responses. LGP2, or DHX58, a homolog of RIG-I and MDA5, facilitates viral RNA recognition by binding to double-stranded viral RNA and/or interacting directly with RIG-I and MDA5 [43].

As seen in Figure 8, 3D structure models of Ypel1-5 and of Yippee show some similarities as the highly conserved 'Yippee' domain was modelled based on the same group of template proteins. However, it appears that Ypel1 and 2 are more similar to each other than other Ypel proteins. Ypel3 shares similar regions to Ypel1 and 2 while it possesses more disordered structures corresponding to regions where there are extra sequences that are absent in Ypel1 and 2. Ypel4 and 5 exhibit similar structure to each other, yet they are more different than Ypel1-3. The prediction of 3D structures of Ypel1-5 and Yippee confirms that the high levels of nucleotide and amino-acid sequence similarities are also reflected in similarities in the tertiary structures of Ypel1-5 and Yippee. The predicted 3D structure of Ypel proteins together with aforementioned biochemical and cellular studies therefore suggest that Ypel proteins could function as nucleotide binding proteins and be involved in a variety of cellular processes during development, stress and aging.

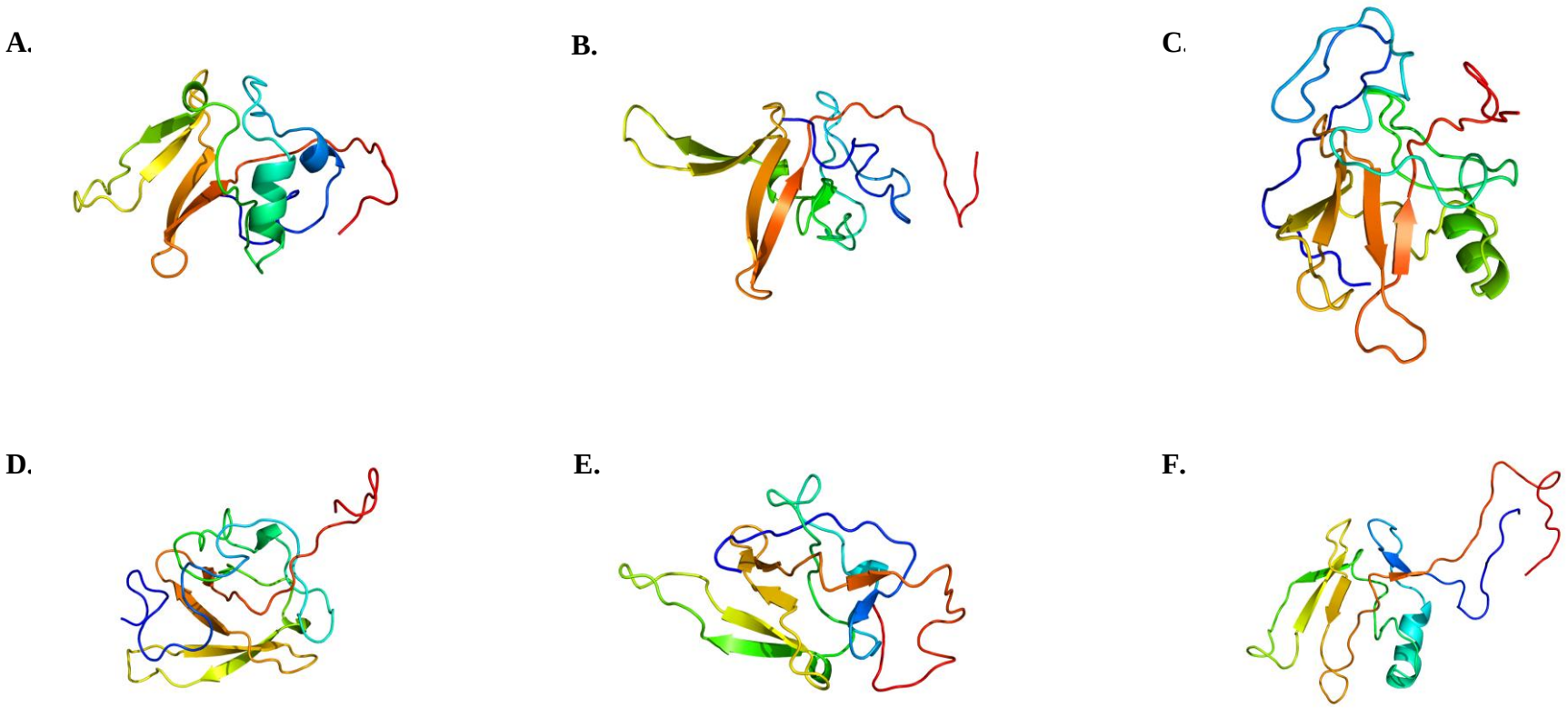
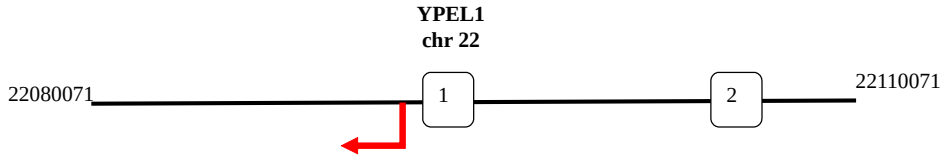


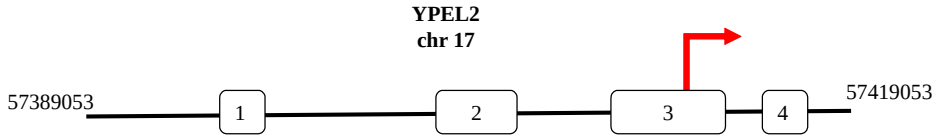
Figure 8. Predicted 3D structures of human Ypel1-5 and Drosophila Yippee. Phyre2 tool is used for homology modeling of A) Ypel1, B) Ypel2, C) Ypel3, D) Ypel4, E) Ypel5 and F) Yippee to construct 3D models. Arrows indicate beta strands, helices correspond to alpha helices and loops are disordered structures.

3.1.7. Transcription factor binding sites

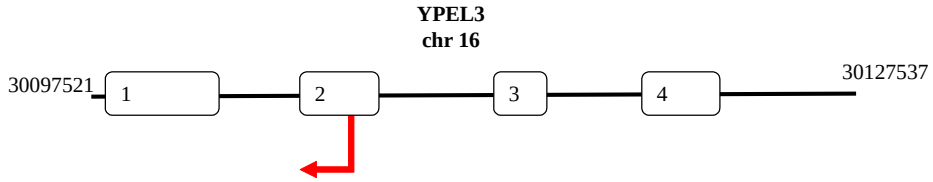
Despite the fact that all *YPEL* genes share a high degree of nucleotide sequence similarity, the expression of these genes could be regulated differentially by various signaling pathways in a spatiotemporal manner. This in turn implies that each *YPEL* locus contains distinct regulatory sequences recognized by different transcription factors. To examine this possibility, we performed a transcription factor binding site analysis (Figure 9).



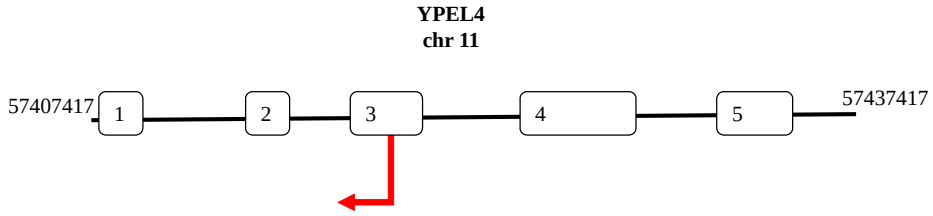
1: RFX1, c-Myb, CP2,SRY, Sox9, CBF-A, CBF-B, CBF-C
2: MIF-1, Nkx2-5, RFX



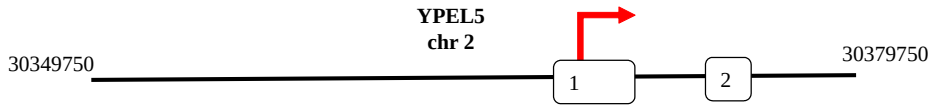
1: POU2F1a, POU2F1
2: E4BP4, Hlf
3: POU3F1, Ik-3, FOXO3, FOXO3a, FOXO3b, POU2F1a, POU2F1, HNF-1
4: POU2F1a, POU2F1



1: CREB, deltaCREB, c-Myc, Max1, POU2F1a
2: CREB, deltaCREB, Max1, POU2F1a, POU2F2, POU2F2 (Oct-2.1), POU2F2B, POU2F2C, Oct-B1
3: CREB, deltaCREB, Max1,
4: CREB, deltaCREB



1: p53
2: FAC1, p53
3: ATF6, LUN-1, Pax-4a, STAT3
4: Elk-1, ATF6, LUN-1, FAC1, NRF-2, FOXF2, p53
5: Elk-1, NRF-2, STAT3, FOXD3



1:Elk-1, Bach2, FAC1, Arnt, c-Myc, Max, USF-1,
USF-1: USF-2
2: Arnt, aMEF-2, MEF-2A

Figure 9. Transcription factor binding site predictions for human YPEL1-5. Approximate locations of predicted binding sites for a group of TFs on the upstream and downstream of each gene are represented with numbers in the rectangular frames. Size of rectangles corresponds to the length of sequence. Names of TFs are provided just below the diagram for each gene. Red arrows indicate transcription start sites. Chromosomal positions are given for the start and end of the represented regions.

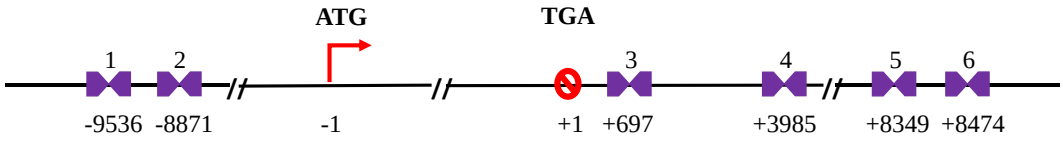
The Champion ChiP Transcription Factor Search Portal (http://www.sabiosciences.com/chipqpcrsearch.php?app=TFBS) is used to predict binding sites in 20 kb upstream and 10 kb downstream sequences of *YPEL1-5*. The Champion ChiP Transcription Factor Search Portal is based on SABiosciences' proprietary database known as DECODE (DECipherment Of DNA Elements). DECODE uses the data available in UCSC Genome Browser through SABiosciences' Text Mining Application.

The results of analysis with the most relevant TFs for each *YPEL* are depicted in Figure 9. Comparison of predicted TFs reveals that regulatory regions of *YPEL1-5* are indeed recognized by different TFs. This finding also supports our initial preposition that although *YPEL* genes have a high sequence similarity and consequently similar protein functions, each could have a different spatiotemporal expression as a result of responses to different signals.

3.1.8. Prediction of ERE sequences

YPEL3 is previously reported to be an E2 regulated gene in MCF7 cells [18]. *YPEL2*, on the other hand, appears to be an E2 responsive gene in ER negative MDA-MB-231 cells infected with a recombinant adenovirus bearing a Flag-ERα cDNA and that E2-ERα regulates the expression of *YPEL2* through the ERE-dependent signaling pathway [6]. However, to date, there is no study to show that ER indeed binds directly to its regulatory sequences of *YPEL2* or *YPEL3* (Figure 10).

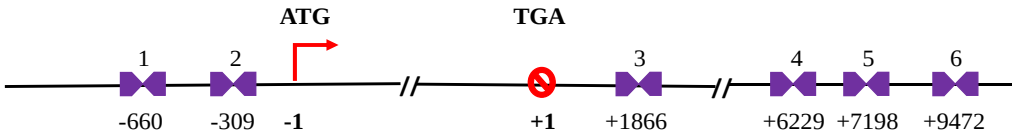
YPEL2:



Consensus ERE sequence: GGTCAnnnTGACC

YPEL2_ERE1: AGTCAgggAGACC
YPEL2_ERE2: AGGCAggtGGACC
YPEL2_ERE3: GGGTGcgaTGACT
YPEL2_ERE4: GGTCAgatTGCCT
YPEL2_ERE5: GGTCAccgCAACC
YPEL2_ERE6: GGTCAgacTGGTC

YPEL3:



YPEL3_ERE1: GGCCAgccCACCC
YPEL3_ERE2: GGTCGcaaGGACC
YPEL3_ERE3: GCTCAggcTGACC
YPEL3_ERE4: GGTCCctcTGCCC
YPEL3_ERE5: AGTCAgctTGCCC
YPEL3_ERE6: GGTCGtgtGGGCC

Figure 10. Predicted ERE like sequences of *YPEL2* and YPEL3. Dragon ERE finder is used to predict ER binding sites in the 10 kb upstream and 10 kb downstream sequences of YPEL2 (A) and YPEL3 (B). Positions of ERE like sequences relative to initiation codon (ATG, -1) and stop codon (TGA,+1) are marked. Sequences corresponding to each numbered position are given below the respective diagram.

To examine the existence of putative ERE sequences on *YPEL2* & *3* gene loci, we scanned 10 kb upstream and 10 kb downstream sequences using Dragon ERE finder version 3 (http://datam.i2r.a-star.edu.sg/ereV3/). Dragon ERE finder is a "System for Identification and Interactive Analyses of Estrogen Response Elements in DNA Sequences" [32]. Dragon ERE finder predicts the presence of ERE like sequences that could differ from the consensus ERE by a maximum of four mismatches in the provided DNA sequences. The consensus ERE sequence is defined as 5'-GGTCAnnnTGACC-3'; however sequences with deviations up to three nucleotides are shown to be functional ER binding sites [7].

The analysis of 10 kb upstream sequence of *YPEL2* resulted in the identification of two ERE like sequences that have two and three mismatches, respectively. In the 10 kb downstream sequence of *YPEL2*, there are four ERE like sequences of which three differ from the consensus ERE by two nucleotides while one has four mismatches. In the 10 kb upstream sequence of *YPEL3*, two ERE like sequences are predicted to have two or four mismatches. The 10 kb downstream sequence of *YPEL3* contains four predicted binding site: One sequence with one mismatch, two sequences with two mismatches and one sequence with three mismatches.

These predictions together with our finding described in Figure 17 and Figure 18 that the expression of *YPEL2* and *3* is regulated by E2 and ERα in hormone responsive breast cancer cell lines, provides the impetus to continue our studies directed towards the identification of ERα binding sites in *YPEL2* and/or *YPEL3* by various approaches to state that these two genes are *bona fide* E2 responsive genes regulated by ERα through the ERE-dependent signaling route. To accomplish this, we will perform chromatin immunoprecipitation (ChIP) assays. Chromatin regions immunoprecipitated with an ERα-specific antibody will be analyzed with primers designed to amplify DNA segments surrounding the predicted ERE sequences. The binding of ERα to sequences identified in ChIP studies will subsequently be confirmed *in vitro* with electrophoretic mobility shift essays (EMSA). Functionality of such ERE sequences will then be tested with luciferase reporter assays both within the context of endogenous gene promoter and a basic TATA

Box promoter. We have already designed primers to be used in ChIP-PCR for the predicted ERE like sequences of *YPEL2*. Similar approaches will also be used for *YPEL*3.

3.2. Cloning of *YPEL1-5* cDNAs from an MCF7 cDNA library

To study Ypel proteins in cells, the ORF of each *YPEL* gene was first cloned into an intermediary *E.coli* expression vector pBS-KS (-), then subcloned into the mammalian cell expression vector pcDNA 3.1 (-) as described in section 2.4. It should be noted that in cloning of the *YPEL3* cDNA, we used transcript variant 1 that encodes 157 aa long protein. As seen in Figure 3, *YPEL3* transcript variant 1 and 2 have an identical core sequence, while variant 1 has some extra 5' sequences. Constructs are designed to have flexibility that enables the in-frame insertion of a different tag(s) encompassing c-myc, Flag, HA, StrepII or V5 to the 5' of ORFs and/or 6His tag to the 3' of ORFs.

For each gene, forward primers from the 5'UTR and reverse primers from the 3'UTR, both of which were described in the Section 2.4.1 of Material & Methods used in first PCR reactions using the MCF7 cDNA library. Products of the first PCR reaction are shown in Figure 11.

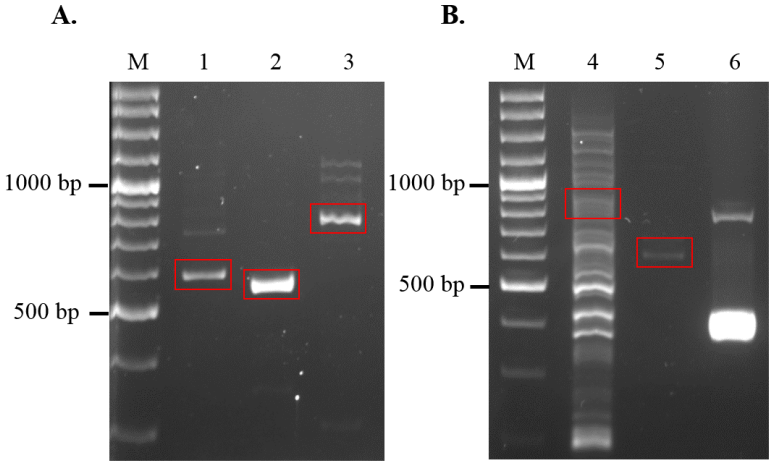


Figure 11. Products of PCR reactions using 5'UTR forward and 3'UTR reverse primers for open reading frames of *YPEL*1-5. An MCF7 cDNA library was used as template to amplify ORFs of *YPEL*1-5 with gene specific primers. PCR products were run on a 1% agarose gel, stained with ethidium bromide and visualized. A) M: DNA ladder, 1: *YPEL1*, 598 bp; 2: *YPEL2*, 567 bp; 3: *YPEL3*, 773 bp; B) M: DNA ladder, 4: *YPEL4*, 967 bp; 5: *YPEL5*, 683 bp; 6: *GAPDH*, 695 and 407 bp, as positive control.

Products with expected sizes were excised and extracted from agarose gels and used as templates for the second PCR with cloning primers. The second PCR products were run on a 1% agarose gel, visualized, excised and column purified (Figure 12).

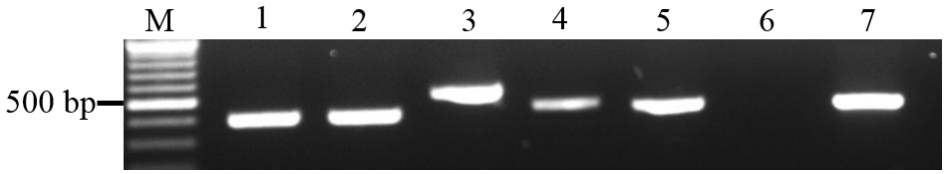


Figure 12. Products of second PCR reaction using cloning primers. Gel extracted products from the first PCR reaction carried out with the 5'UTR forward and the 3'UTR reverse primers were used as templates for the synthesis of corresponding *YPEL* cDNA. Gene specific cloning primers were designed to add restriction enzyme (RE) sites to the 5' and 3' ends of ORFs. PCR products were run onto a 2% agarose gel, stained with ethidium bromide and visualized. M: DNA ladder, 1: *YPEL*1, 360 bp; 2: *YPEL2*, 360 bp; 3: *YPEL3*, 474 bp; *YPEL4*, 384 bp; *YPEL5*, 366 bp; 6: negative control, 7: *GAPDH*, 407 bp, as positive control.

Column purified PCR products from the second PCR (Figure 12) and the pBS-KS (-) vector were double digested with XhoI and BamHI restriction enzymes, ran on a 1% agarose gel. PCR products with expected sizes were extracted and subjected to ligation. Ligation products were transformed into competent *E. coli* XL1-Blue cells. Transformed bacterial colonies were screened by colony PCR using gene specific cloning primers. Colonies with desired inserts were grown in selective LB media and plasmid isolation was carried out. The resultant cDNA containing pBS-KS (-) vectors were sequenced at the METU Central Laboratories. cDNA sequences were compared to sequences of the NCBI Nucleotide database and the correct ones were used for sub-cloning into a mammalian expression vector, pcDNA 3.1(-).

3.3. Investigation of endogenous and recombinant Ypel proteins in COS7 cells

3.3.1. Immunocytochemistry (ICC)

To examine the synthesis of Ypel proteins, we initially selected to use COS7 cell line derived from transformed African green monkey kidney fibroblast-like cells. COS7 cells were previously used to detect endogenous Ypel proteins using laboratory-made antibodies directed to Ypel1 and Ypel5 [14]. The antibody based on Ypel1 was suggested to recognize Ypel1-4, while the Ypel5 antibody was stated to specifically recognize Ypel5. Immunocytochemistry studies using the Ypel1 antibody suggest that Ypel1-4 are nuclear proteins that show distinct locations in cell cycle phases.

We therefore wanted to verify these results and to further examine the synthesis and intracellular locations of recombinant Ypel proteins in transiently transfected cells using a commercially available anti-YPEL antibody. This Ypel (S-14) antibody purchased from Santa Cruz Biotechnology (CA, USA) is raised against a synthetic peptide corresponding to an internal region of the human Ypel1 protein (http://www.scbt.com/datasheet-99727-ypel-s-14-antibody.html). It was predicted to recognize all five Ypel proteins although it was not used in any published work. Since we engineered recombinant proteins to bear a Flag epitope at the amino-terminus, using both the Flag-M2 and the S-14 Ypel antibodies would also allow us to differentiate which Ypel protein(s) is recognized by the S-14 Ypel antibody and to comparatively analyze the intracellular location of each protein.

We initially analyzed the intracellular localization of endogenous Ypel proteins in COS7 cells using the S-14 Ypel antibody (Figure 11). We also used an antibody against LaminA, a structural component of the nuclear lamina, to localize the nuclear envelope. DAPI (4', 6-Diamidino-2-Phenylindole, Dilactate) was used as a nuclear counterstain.

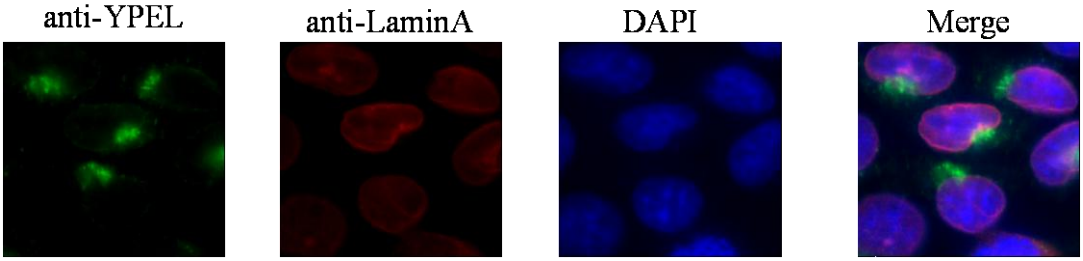


Figure 13. Immunocytochemistry of COS7 cells for endogenous Ypel expression. COS7 cells grown on coverslips in 12-well cell culture plates were fixed with 2% paraformaldehyde, permeabilized with 0.4% Triton-X100. After blocking with 10% normal goat serum (NGS) for one hour, cells were incubated with the S-14 Ypel antibody (1:50 in 2% NGS) for two hours and then with LaminA antibody (1:200 in 2% NGS) for two hours, sequentially. An Alexa Fluor®-488 conjugated goat anti-rabbit (green channel emission) secondary antibody at 1:1000 in 2% NGS was used for the visualization of Ypel protein. An Alexa Fluor®-647 conjugated goat anti-mouse (red channel emission) secondary antibody at 1:200 in 2% NGS was used for the visualization of LaminA. Nuclei were counterstained with DAPI. Merge denotes the overlapped images taken by different emission channels. Images are from a representative experiment. Similar results were obtained with two independent experiments.

Results revealed that the S-14 Ypel antibody detected a Ypel protein(s) localized at a distinct region just outside of the nucleus as the LaminA and DAPI staining indicate (Figure 13). Consistent with this finding, Hosono *et. al.* [14], using an antibody raised against Ypel1, observed that Ypel1-4 were localized in a cytoplasm region adjacent to nucleus that is likely centrosomes as this region was also stained with an anti-gamma tubulin antibody.

However, Hosono *et. al.* also detected a nucleolar staining, in clear contrast to our observations. This striking difference could be due to the antibodies used.

To comparatively examine the intracellular location of each Ypel protein comparatively and to investigate which Ypel protein is recognized by the S-14 Ypel antibody, we transiently transfected COS7 cells with pcDNA3.1 (-) bearing none (as control, empty vector: EV), a Flag-YPEL cDNA or Flag-ERα cDNA (as control for a nuclear protein). ICC using the Flag-M2 (Figure 14) or S-14 Ypel (Figure 15) antibody was performed at 36 hour post-transfection.

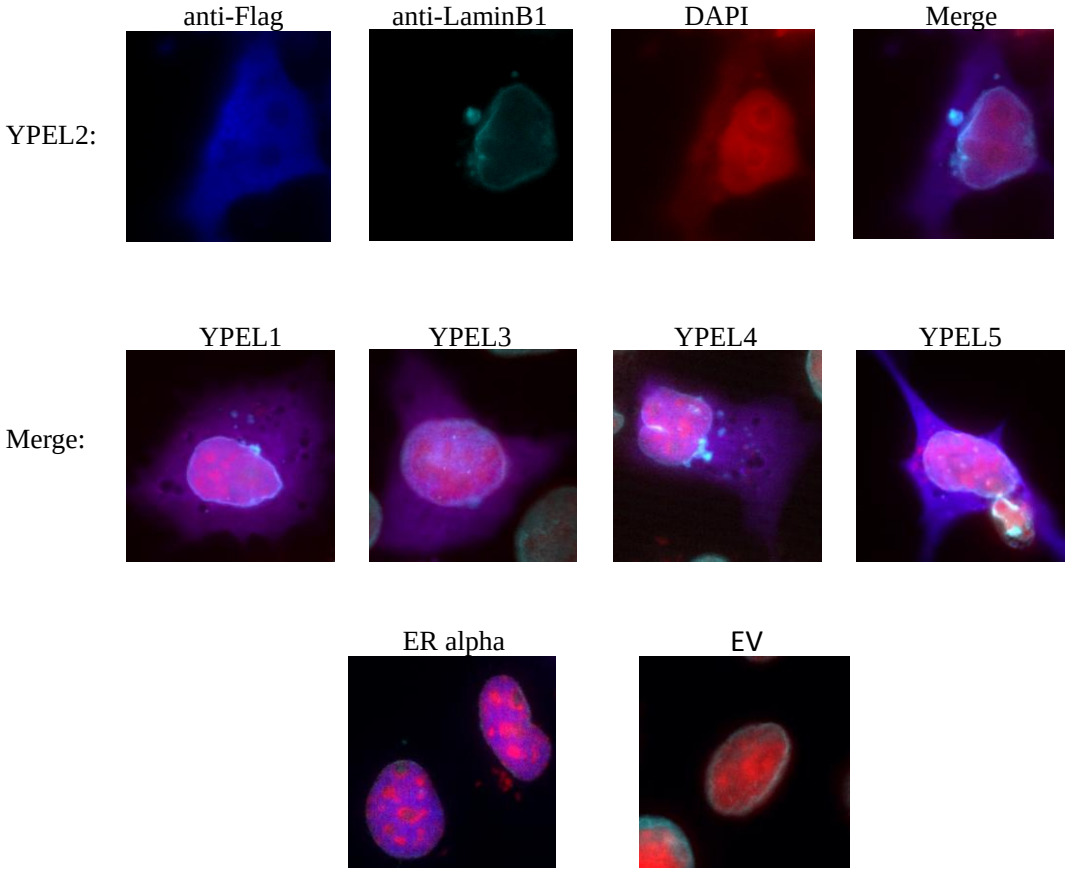


Figure 14. Immunocytochemistry of transiently transfected COS7 cells with Flag antibody. COS7 cells grown on coverslips in 12-well cell culture plates were transfected 48 hours after seeding. Cells were transfected with pcDNA3.1 (-) bearing none (as control, empty vector: EV), a Flag-YPEL cDNA or Flag-ERα cDNA (as control for nuclear staining). 36 hours after transfection, cells were fixed with 2% paraformaldehyde, permeabilized with 0.4% Triton-X100. After blocking with 10% bovine serum albumin (BSA) for one hour, cells were incubated with the Flag-M2 (1:250 in 3% BSA) for two hours and then with the LaminB1 (1:200 in 3% BSA) antibodies for two hours, sequentially. An Alexa Fluor®-488 conjugated (emission at green channel) goat anti-mouse and an Alexa Fluor®-647 conjugated (emission at red channel) goat anti-rabbit (1:1000 and 1:200 in 3% BSA, respectively) secondary antibodies were used for the visualization of Flag-Ypel and LaminB1, respectively. Nuclei were counterstained with DAPI (emission at blue channel). It should be noted that light emissions were collected using the aforementioned channels. However, to better visualize the overlapping regions we were in the need of using artificial coloring; accordingly the Flag staining is colored in blue, LaminB1 is in cyan and DAPI is in red. Merge denotes the overlapped images taken by different emission channels. Images are from representative experiments. Similar results were obtained with two independent experiments.

Results revealed that the Flag-M2 antibody detects Flag-ERα in the nucleus while cells transfected with the parent vector bearing no cDNA shows no staining. This indicates that the Flag-M2 antibody specifically recognizes Flag-ERα. As with Flag-ERα, the Flag-M2 antibody staining of cells transfected with a Flag-YPEL cDNA indicates that Flag tagged Ypel1-5 proteins are synthesized in cells.

We detected Flag-Ypel2 using the Flag-M2 antibody both inside and outside of the nucleus, whose boundary is indicated with the detection of the nuclear membrane LaminB1 protein in cells transfected with the Flag-YPEL2 cDNA. The DAPI staining indicates that some of DNA was diffused out from the nucleus into the cytoplasm in a pattern perfectly overlapping with the staining of Flag-Ypel2. Cells that were not stained with the Flag-M2 antibody as an indication of un-transfected cells within the same population showed a DAPI staining that was identical to that observed in cells transfected with the vectors bearing none (EV) or the Flag-ERα cDNA. This suggests that the over-expressed Ypel2 is directly or indirectly associated with DNA which is abnormally found in the cytoplasm.

We observed a similar pattern of staining in COS7 cells transfected with the cDNA of other YPEL family members, indication of an effect that is common to all Ypel proteins.

When we replicated ICC studies using the S-14 Ypel antibody, we detected a DNA staining pattern similar to experiments depicted in Figure 12. It should be noted that the DAPI staining overlapped with the staining of S-14 Ypel antibody when cells were transfected with YPEL1, 2 and 3 as well as 4, although to a much lesser degree as judged by the intensity of signal and the number of stained cells. We however observed no staining with the S-14 Ypel antibody in the cells transfected with the Flag-YPEL5 cDNA despite the fact that DNA showed an abnormal staining similar to other Ypel proteins when they were over-expressed.

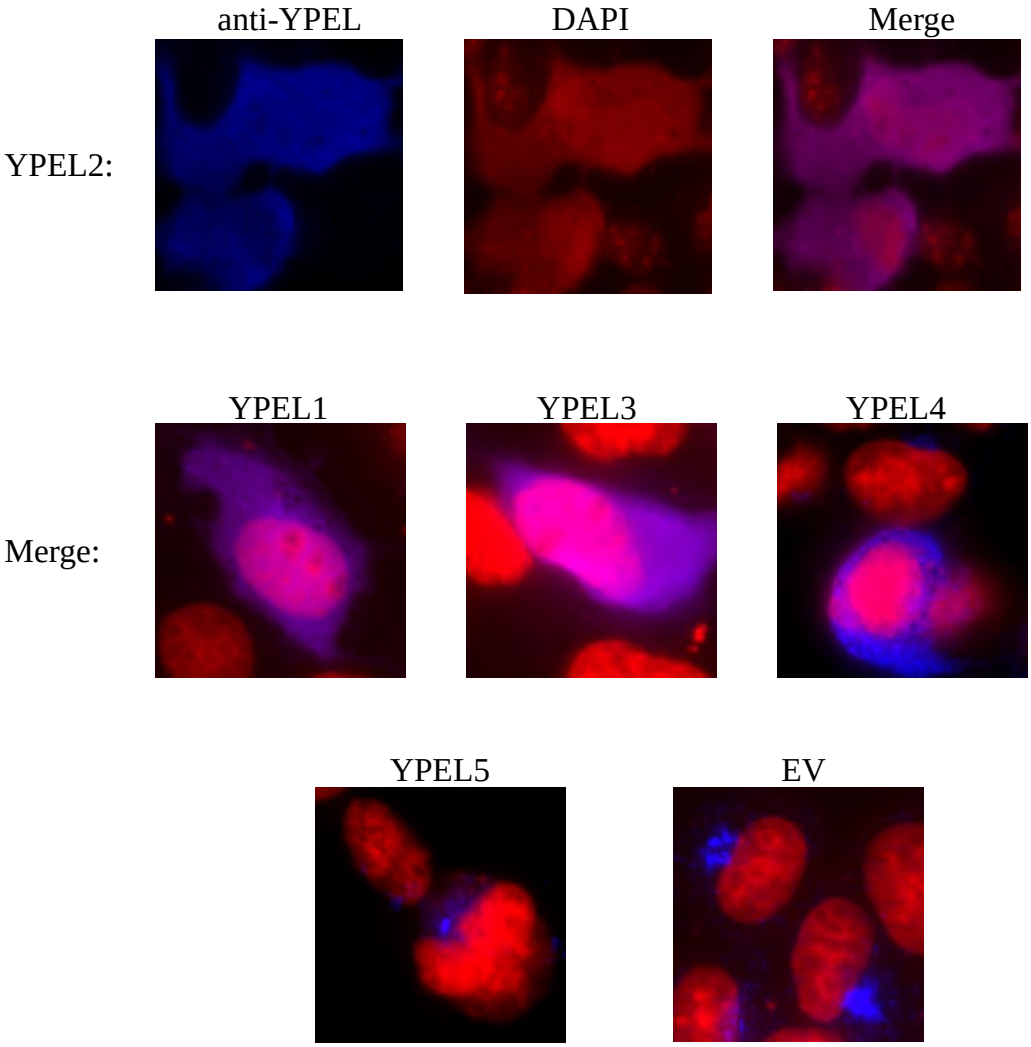


Figure 15. Immunocytochemistry of transiently transfected COS7 cells with YPEL antibody. COS7 cells grown on coverslips in 12-well cell culture plates were transfected 48 hours after seeding. Cells were transfected with pcDNA3.1 (-) bearing none (as control, empty vector: EV) or a Flag-YPEL cDNA. 36 hours after transfection, cells were fixed with 2% paraformaldehyde, permeabilized with 0.4% Triton-X100. After blocking with 10% normal goat serum (NGS) for one hour, cells were incubated with the S-14 Ypel antibody (1:50 in 2% NGS) for two hours. An Alexa Fluor®-488 conjugated (emission at green channel) goat anti-rabbit (1:1000 in 2% NGS) secondary antibody was used for the visualization of Ypel proteins. Nuclei were counterstained with DAPI (emission at blue channel). It should be noted that light emission were collected using the aforementioned channels. However, to better visualize the overlapping regions we were in the need of using artificial coloring; accordingly the Ypel staining is colored in blue and DAPI is in red. Merge denotes the overlapped images taken by different emission channels. Images are from representative experiments performed two independent times with similar results.

Based on ICC results obtained from transiently transfected COS7 cells, we conclude that:

1) The S-14 Ypel antibody recognizes Ypel1, 2 and Ypel3 variant 1 as well as Ypel4, albeit to a lesser degree, while the Flag-M2 antibody does recognize all Ypel proteins.

2) The endogenous Ypel protein in COS7 cells that show localization at a region adjacent to the nucleus could be Ypel1, 2, Ypel3 variant 1 and/or 4, but not Ypel5. We are not certain if Ypel5 is endogenously synthesized in COS7 cells.

3.3.2. Western Blot

To further confirm our ICC results and test the effectiveness of the antibodies to detect endogenous and recombinant proteins under denaturing conditions, we performed WB analyses with un-transfected and transfected COS7 cells.

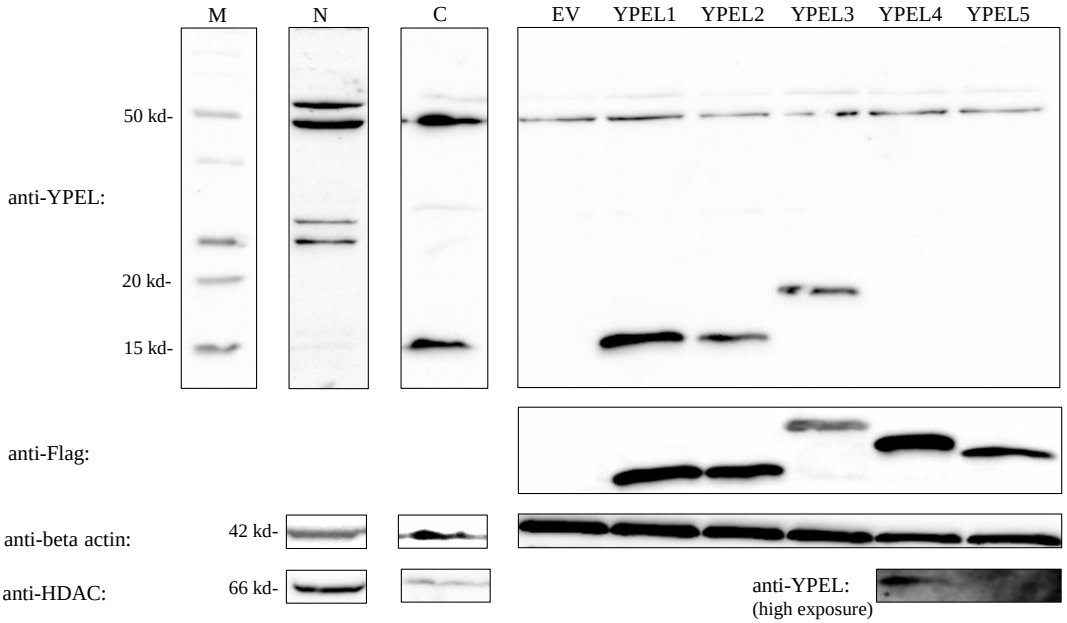


Figure 16. Composite Western Blot analysis of COS7 cells. COS7 cells were maintained for 48 hours after seeding. A group of cells were transfected with pcDNA3.1 (-) bearing none (as control, empty vector: EV) or a Flag-YPEL cDNA and maintained an additional 36h. Cells were then subjected to protein extraction. 200 µg of nuclear (N) and 200 µg of cytoplasmic (C) extracts from un-transfected COS7 cells; 100 µg of total proteins from transfected cells were loaded to 15 % SDS-PAGE, M: molecular weight marker. The S-14 Ypel and Flag-M2 antibodies were used to detect Ypel proteins in WB. Beta actin and HDAC antibodies were used as the cytoplasmic and nuclear loading controls, respectively. Experiments were performed two independent times with similar results. High exposure image of S-14 Ypel antibody is included for Ypel4.

In Figure 16, images from different experiments are presented together. Images are aligned according to the same molecular weight marker used in WBs.

We used both the nuclear and cytoplasmic extracts. In preliminary studies we used various concentrations, up to 200 µg, of cellular extracts. We found that we were able to detect a Ypel protein when we used 200 µg, but not less, cytoplasmic protein. This suggests that the amount of Ypel protein is very low in COS7 cells. Based on these findings, we used 200 µg cytoplasmic and nuclear proteins in subsequent experiments.

As depicted in Figure 16, we detected a protein band with a molecular mass of 15 kDa in the cytoplasmic extract of un-transfected COS7 cells. The size of this protein band corresponds to the size of Ypel1 or 2 but not Ypel 3 variant 1, 4 or 5,

in transfected cells suggesting that COS7 cells endogenously synthesize Ypel1, Ypel2 and/or Ypel3 variant 2. In the nuclear extracts, we also detected, albeit faintly, a protein band at the same molecular mass of the detected cytoplasmic protein, which could be due to cytoplasmic contamination of the nuclear extract as the levels of beta actin, a cytoplasmic protein, in nuclear extracts suggests.

In keeping with the results of ICC studies, Figure 16 also shows that Ypel1-5 in 100 µg of total protein extracts of transiently transfected cells were detected by the Flag-M2 antibody confirming the synthesis of Ypel proteins in WB as well. The recognition of Flag-Ypel proteins by the Flag-M2 antibody is antibody specific; this is because in extracts of cells transfected with the parent vector (EV), the Flag-M2 antibody does not detect a protein corresponding to the molecular masses of Ypel proteins.

Moreover, our results indicate that the S-14 Ypel antibody recognizes Ypel1, 2, 3 or 4, although at very low levels as detected at high exposures (Figure 16, bottom panel), but not Ypel5 in WB as observed in ICC. Ypel1 and 2 have 96.6% amino-acid sequence identity and the internal 88 amino-acids are identical (Figure 7). Since the antibody was raised against a fragment of this internal region, we expected that the S-14 Ypel antibody would recognize Ypel2 in WB as well. Despite the high level of amino-acid sequence conservation, subtle differences in amino-acid composition at the region against where the antibody was raised could be one explanation as to the why we barely detected Ypel4 when use 100 µg of total cell extracts at high exposures. The distinct amino-acid sequence of Ypel5 compared to that of Ypel1 likely renders Ypel5 unrecognizable by the S-14 Ypel antibody in WB and ICC.

WB analyses of endogenous and recombinant Ypel proteins in COS7 cells lead to the following conclusions:

1. The S-14 Ypel antibody recognizes Ypel1-4 but not 5 under denaturing conditions.

2. The molecular mass of the endogenous Ypel protein detected in the cytoplasmic fraction of un-transfected COS7 cells suggests that this endogenous protein is Ypel1, 2 and/or Ypel3 variant 2, but it is not Ypel3 variant 1, Ypel4 or 5.

3. Since we detected endogenous Ypel protein(s) in 200 µg of cytoplasmic but not in 100 µg of total protein extracts, our results suggest that the levels of endogenous Ypel proteins are low in COS7 cells. This could be due to transcriptional and/or post-transcriptional processes:

    a. The expression of *YPELs* is spatiotemporally regulated in that the gene expression is dependent upon distinct signaling pathways activated under certain conditions.

    b. *YPEL* expressions are cell cycle-dependent. In a non-synchronous growth condition, levels of *YPEL* expressions vary from cell-to-cell in a population, depending on which cell cycle stage a cell is at, leading to the protein synthesis restrictive to a certain phase of the cell cycle. Growth synchronization of cells could be necessary to detect stable protein levels.

    c. Levels of transcripts are regulated post-transcriptionally by processes that include non-coding RNAs and/or RNA binding proteins, leading to the degradation of transcripts and/or the prevention of mRNA interaction with ribosomes.

    d. Ypel proteins are subject to a rapid degradation through proteasome and/or lysosome dependent pathways.

3.4. Identification of *YPEL2* as an estrogen responsive gene

YPEL2 was identified as an estrogen responsive gene in a microarray study carried out with MDA-MB-231 cells infected with adenovirus bearing ERα or $ER\alpha_{DBD}$ (ERE binding defective but transcriptionally intact) cDNAs in our laboratory [6], [44]. Microarray results showed that *YPEL2* mRNA levels were augmented in ERα infected cells in response to E2 treatment, while E2-$ER\alpha_{DBD}$ had no effect on the expression of *YPEL2*. These findings indicate that the expression of *YPEL2* is regulated by the ERE-dependent signaling pathway. Although the signaling route is unknown, *YPEL3* was similarly shown to be regulated by E2 in ERα positive breast cancer cell lines [18].

In order to verify the microarray results for *YPEL2* while using *YPEL3* and *TIFF1/ps2*, a well-established E2-ER responsive gene, as positive controls, RT-qPCR was performed with RNA samples obtained from ERα-positive MCF7 cells treated with a physiological concentration, 10^{-9} M, E2 (Figure 17).

Following MIQE guidelines, optimizations were carried out to set the reaction efficiencies of each primer to 1 (0.9-1.2). Melt curve analyses were performed for each reaction (Appendix). For each gene, three biological replicates were used and each reaction was performed with three technical replicates. Results were normalized to the expression of the ribosomal protein, large, P0 (*RPLP0*) and/or homolog of Pumilio, Drosophila, 1 (*PUM1*) genes. These genes are proposed to be the most reliable reference genes for normalization of RT-qPCR results in breast carcinomas [34].

Results are represented as mean of three biological replicates with standard deviation. For statistical analysis of normalized fold changes we used one tailed paired t-test, with 95% confidence interval [36].

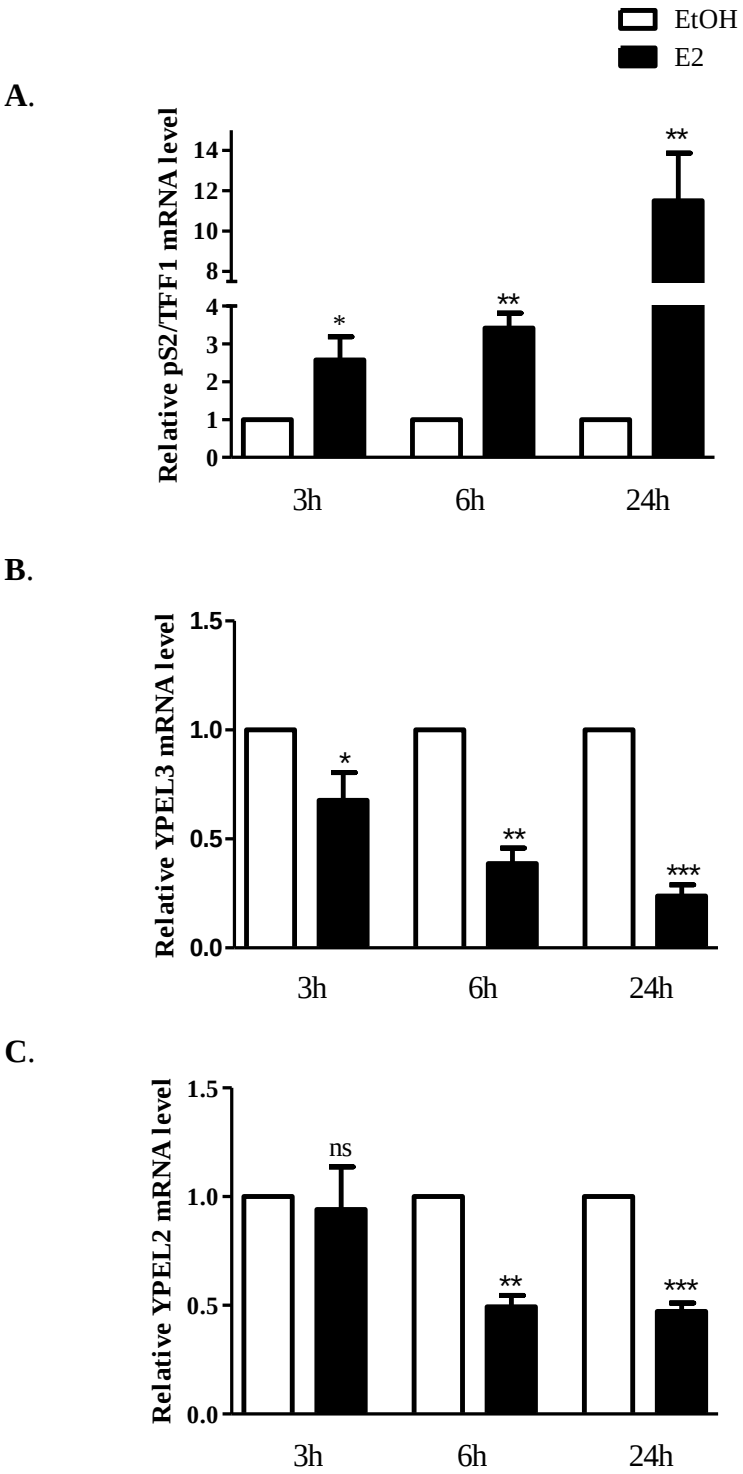


Figure 17. RT-qPCR results in E2 treated MCF7 cells. MCF7 cells grown in T25 tissue culture flasks in medium containing CD-FBS for 48h to reduce the endogenous steroid hormone concentrations were treated without (EtOH) or with E2 (10^{-9}M) for 3h, 6h and 24h. Cells were then subjected to RNA extractions and RT-PCR. Changes in mRNA levels normalized to *RPLP0* assessed by qPCR are depicted for A) *pS2/TFF1*, B) *YPEL3* and C) *YPEL2*. Results are the mean ± SD of three independent experiments performed in triplicate. *, ** and *** denotes significant effect of E2 compared with EtOH at a time point. ns defines non-significant. Statistical analysis was performed with one-tailed paired t-test, 95% confidence interval, $p < 0.05$.

We found that E2 represses the expression of *YPEL2* and *YPEL3* genes in a time dependent manner in MCF7 cells (Figure 17). This contrasts to the expression of the *TFF1/pS2* gene, which was augmented by the E2 treatment. Thus, as *TFF1* and *YPEL3* genes, *YPEL2* is an E2-ER responsive gene in MCF7 cells.

E2-ER signaling can increase or repress target gene expression. Genes augmented by E2 can be grouped as positive proliferation regulators including genes encoding growth factors, growth factor receptors and cell cycle progressors. Repressed genes by E2 include transcriptional repressors, anti-proliferative and pro-apoptotic genes [45]. ERα has mostly been considered as a transcriptional activator and molecular mechanism of E2 induced gene expression is well studied using model genes such as *TFF1/pS2* gene [46]. The transcription of target genes is a dynamic event and is modulated cyclically. Studies indicate that the interaction of the E2-ERα complex with ERE or other transcription factors bound to their responsive DNA elements leads to the recruitment of multi-subunit co-activator complexes, enzymes of the ubiquitin/proteasome pathway, and the basal transcription machinery together with RNA Polymerase II (Pol II) to initiate transcription. As the transcription continues, various chromatin modifying complexes are then recruited to the promoter of the transcribed gene. These complexes modify local chromatin structures that eventually result in the dissociation of associated factors including E2-ERα from the promoter, transcription termination and chromatin re-modeling, which renders the promoter non-permissive to transcription until the initiation of the next transcription cycle [47]

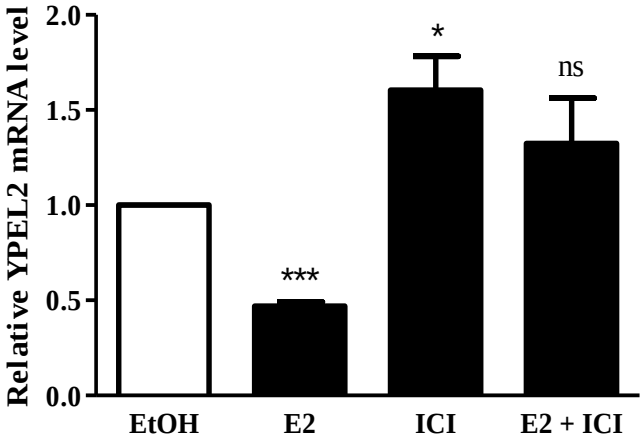
In microarray studies wherein E2 regulated genes were identified, the majority of genes were found to be down-regulated [48], [49], [50]. Despite the detailed knowledge of the mechanism of transactivation by E2-ER complexes, how E2-ERα mediates transcription repression is yet unclear. Although it has not been studied in details, one mechanism is suggested to involve the removal of the transcription machinery from the promoter by remodeling chromatin around the promoter upon binding of E2-ERα directly or indirectly to DNA at sites close to or distant from the promoter. This leads to the dissociation of the basal transcription machinery from

the promoter resulting in transcription termination [48] [49]. Another mechanism with an increasing body of evidence entails the ability of the E2-ERα complex to directly recruit co-repressor/Histone deacetylase (HDAC) complexes to gene promoter, thereby actively repressing the transcription of responsive genes [48], [49], [51]. Studies have demonstrated the E2-ER represses gene expression by recruiting NCoR, (nuclear receptor corepressor), HDAC1, and C-terminal-binding protein 1 (CtBP1) to the cyclin G2 (CCNG2) promoter; NCoR and SMRT (silencing mediator of retinoic acid receptor, RAR, and Thyroid hormone receptor, TR) to the vascular endothelial growth factor receptor 2 (*VEGFR2*) gene promoter; NCoR and Transforming growth factor β-activated protein kinase 1 (*TAK1*)-binding protein 2 (*TAB2*) to the Bone Morphogenic protein 7 (*BMP7*), ATP-binding cassette sub-family G member 2 (*ABCG2*), and B-cell CLL/lymphoma 3 (*BCL3*) gene promoters [45].

Our results with qPCR analysis showed that *YPEL2* is suppressed by E2 at 6 and 24h, while the expression of *YPEL3* gene is repressed at 3h after E2 treatment and remains repressed thereafter. This transcriptional repression of *YPEL2* and *3* genes by E2-ERα bound to DNA could involve an indirect effect that entails E2-ERα-mediated chromatin modeling of the gene promoters or through a direct effect that involves the recruitment of co-repressor/ HDAC complexes to the promoters. Studies utilizing ChIP-based sequence analysis coupled with bioinformatics tools are needed to examine these possibilities.

In order to show that the repression of *YPEL2* and *YPEL3* gene expressions is ER dependent, we used Imperial Chemical Industries 182,780 (ICI), which is considered to be a complete ER antagonist [52]. We found, as depicted in Figure 18, that ICI alone augments the level of *YPEL2* mRNA compared to vehicle (EtOH) treated cells and effectively blocked the E2-mediated repression. These findings suggest that ER regulates both the basal and E2-mediated expression of *YPEL2*. ICI on the other hand had no effect on the basal levels of *YPEL3*, while it effectively blocked the E2-mediated repression. Thus, it appears that the repression of *YPEL3* expression by ERα is dependent upon the presence of E2.

A.



B.

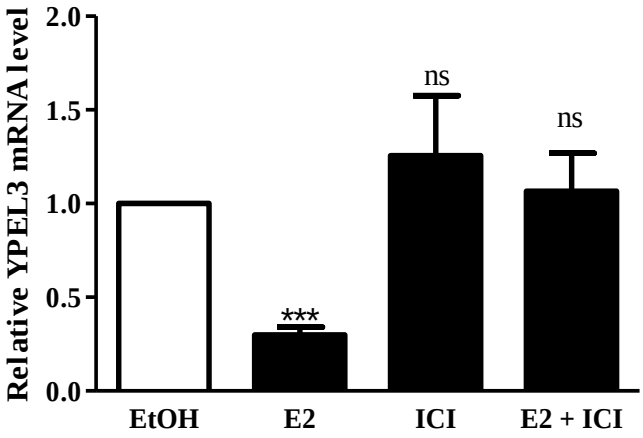


Figure 18. Effect of ICI treatment in MCF7 cells to E2 responsive genes. MCF7 cells grown in T25 tissue culture flasks in medium containing CD-FBS for 48h to reduce the endogenous steroid hormone concentrations were treated without (EtOH) or with E2 (10^{-9}M) and/or ICI (10^{-7} M) for 24h. Cells were then subjected to RNA extractions and RT-PCR. Changes in mRNA levels normalized to *PUM1* assessed by qPCR are depicted for A) YPEL2 and B) YPEL3. Results are the mean ± SD of three independent experiments performed in triplicate. *, ** and *** denotes significant effect of treatment compared with EtOH. ns indicates non-significant. Statistical analysis was performed with one-tailed paired t-test, 95% confidence interval, $p < 0.05$.

3.5. Investigation of endogenous and recombinant Ypel proteins in MCF7 cells

3.5.1. Immunocytochemistry

Based on our observations that the expressions *YPEL2* and *YPEL3* are regulated by E2-ERα in MCF7 cells, we envisioned that the use of this hormone responsive breast carcinoma cell model would allow us to delve further into the roles of Ypel proteins in E2-mediated cellular events that include proliferation and death.

Since we characterized the S-14 Ypel antibody to identify Ypel proteins in ICC and WB of COS7 cells, we employed the same approaches to define the endogenous as well as recombinant Ypel protein synthesis and localization in MCF7 cells.

We initially performed ICC studies using the S-14 Ypel antibody (Figure 19) in un-transfected MCF7 cells.

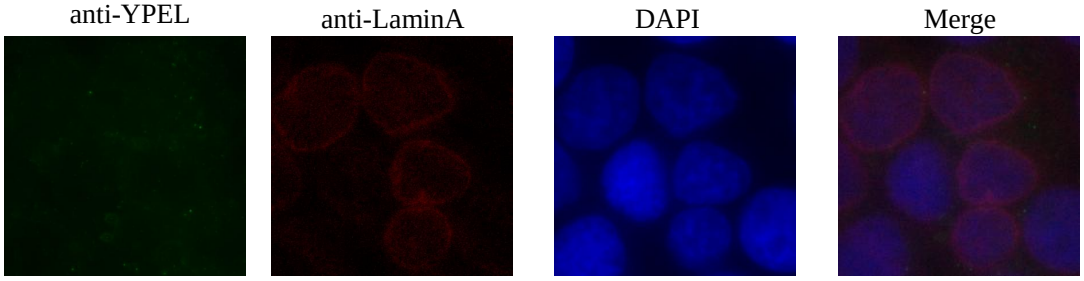


Figure 19. Immunocytochemistry of MCF7 cells for endogenous Ypel expression. MCF7 cells grown on coverslips in 12-well cell culture plates were fixed with 2% paraformaldehyde, permeabilized with 0.4% Triton-X100. After blocking with 10% normal goat serum (NGS) for one hour, cells were incubated with the S-14 Ypel antibody (1:50 in 2% NGS) for two hours and then with the LaminA antibody (1:200 in 2% NGS) for two hours, sequentially. Alexa Fluor®-488 conjugated (emission at green channel) goat anti-rabbit and Alexa Fluor®-647 (emission at red channel) goat anti-mouse (1:1000 and 1:200 in 2% NGS, respectively) secondary antibodies were used for the for the visualization of Ypel and LaminA, respectively. Nuclei were counterstained with DAPI. Merge denotes the overlapped images taken by different emission channels. Images are from a representative experiment performed two independent times with similar results.

Surprisingly, we could not see any endogenous Ypel staining in MCF7 cells at antibody concentrations sufficient to detect endogenous Ypel protein in COS7 cells. Based on these results, we performed additional studies with various concentrations of the primary and/or secondary antibodies. We unfortunately observed no staining for Ypel proteins.

Similarly, we could not detect any Ypel protein in total, nuclear or cytoplasmic extracts, at various conditions, of MCF7 cells in WB analyses (data not shown and Figure 21)

We nevertheless reasoned that the absence of any endogenous Ypel detection in MCF7 cells could provide us with a cell model to study the effect of ectopic expression of individual Ypel proteins without the contribution of endogenous Ypel protein background in contrast to COS7 cells.

For this purpose, we transiently transfected MCF7 cells with pcDNA3.1 (-) bearing none (empty vector, EV), a *YPEL* cDNA, or the Flag-ERα cDNA as a positive control for nuclear localization. Transfected cells were then subjected to ICC at 36h post-transfection using the Flag-M2 together with the LaminB1 antibody or the S-14 Ypel antibody. DAPI was used to define the nucleus.

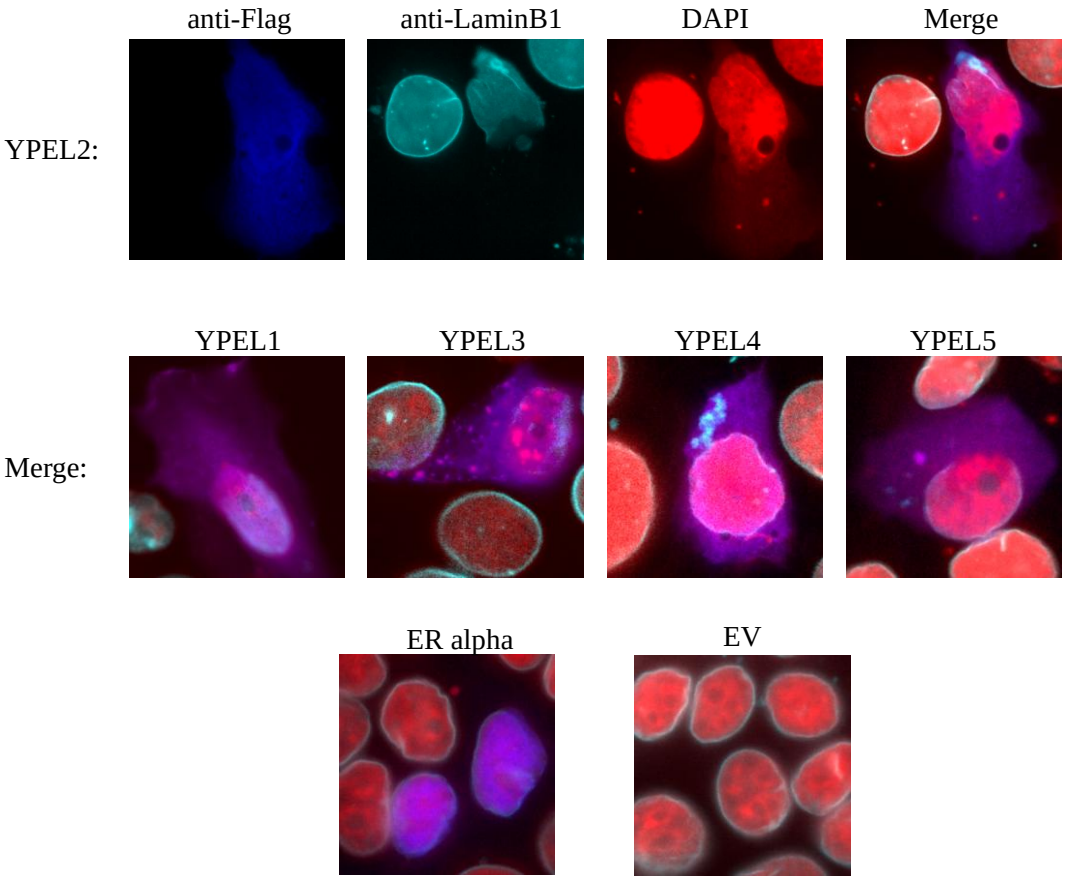


Figure 20. Immunocytochemistry of transiently transfected MCF7 cells with Flag antibody. MCF7 cells grown on coverslips in 12-well cell culture plates were transfected 48 hours after seeding. 36 hours after transfection, cells were fixed with 2% paraformaldehyde, permeabilized with 0.4% Triton-X100. After blocking with 10% bovine serum albumin (BSA) for one hour, cells were incubated with the Flag-M2 antibody (1:250 in 3% BSA) for two hours and then with the LaminB1 antibody (1:200 in 3% BSA) for two hours, sequentially. An Alexa Fluor®-488 conjugated goat anti-mouse (emission at green channel) and an Alexa Fluor®-647 (emission at red channel) conjugated goat anti-mouse (1:1000 and 1:200 in 3% BSA, respectively) secondary antibodies were used for the visualization of Flag-M2 and the LaminB1, respectively. Nuclei were counterstained with DAPI. It should be noted that light emissions were collected using the aforementioned channels. However, to better visualize the overlapping regions we were in the need of using artificial coloring; accordingly the Ypel staining was colored in blue and DAPI is in red. Merge denotes the overlapped images taken by different emission channels. Images are from a representative experiment performed two independent times with similar results.

Our results revealed that EV transfected cells treated with the Flag-M2 antibody showed no staining, while Flag-ERα displayed a distinct nuclear staining (Figure 20). The transfection of Flag-*YPEL1-5* cDNA to MCF7 cells led to a diffused DAPI staining that overlaps with each Ypel protein, as in COS7 cells, confirming that our observation is not dependent on cell type. Importantly however, we also observed in MCF7 cells differently from COS7 cells a disorganized/disrupted nuclear envelope structure as the LaminB1 antibody staining suggests. The nuclear lamina appeared to have compromised integrity together with disintegration and local dense stained regions.

We also performed ICC using the S-14 Ypel antibody (Figure 21). The ectopic expression of Flag-*YPEL1-5* in MCF7 cells led to the diffusion of DNA into cytoplasm as the DAPI staining indicates. The staining of Ypel1-4 showed a remarkable overlap with DNA staining. As in COS7 cells, the S-14 Ypel antibody recognized Ypel1-3. While we could not observe any staining in cells transfected with *Flag-YPEL4* or *5*, the apparent abnormal DNA staining suggests that although the proteins are synthesized, the S-14 Ypel antibody does not recognize Ypel4 or 5 in contrast to the Flag-M2 antibody which showed an overlapping staining with DNA. In keeping with our findings in COS7 cells, our results therefore indicate that the localization of Ypel proteins under over-expression conditions is not cell-specific and is not dependent upon antibody specificity.

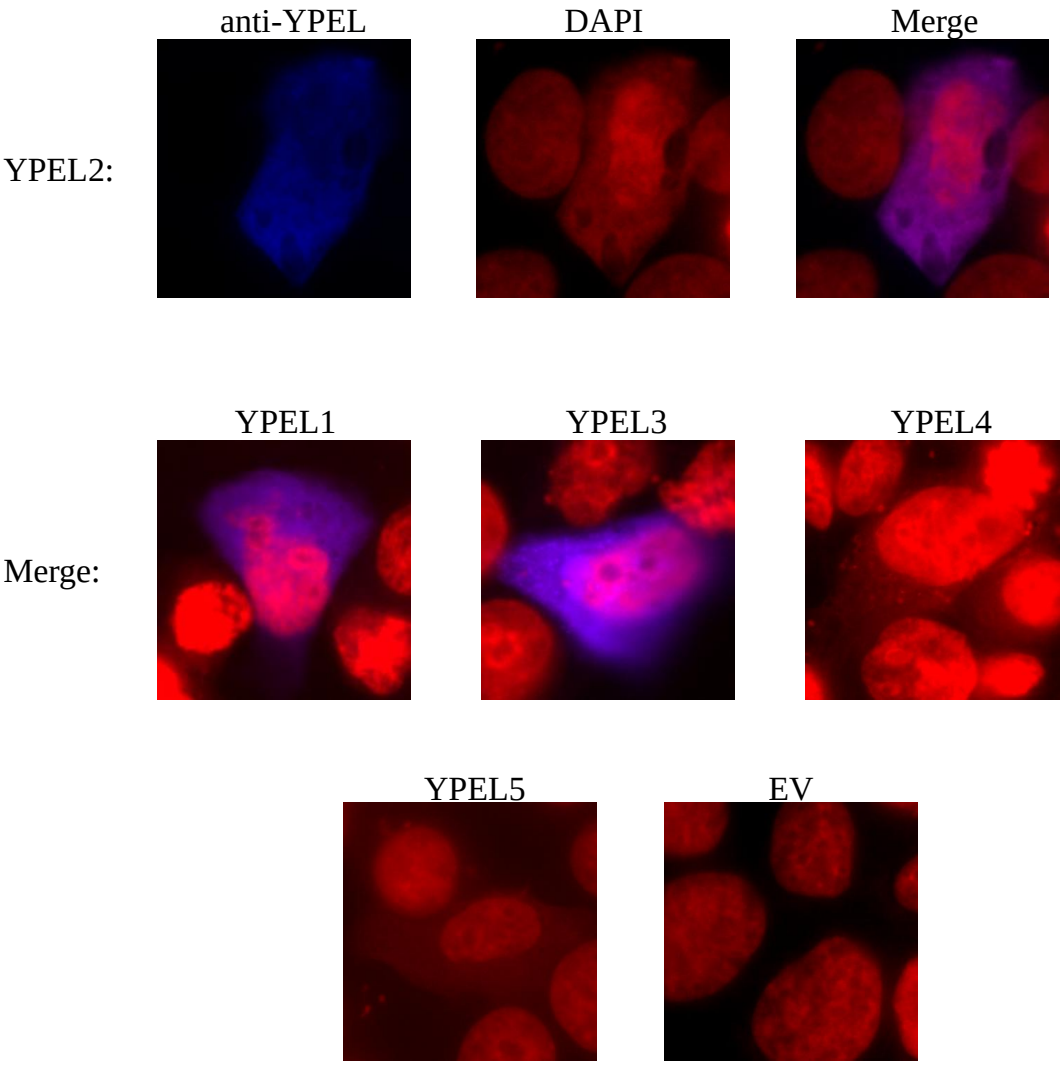


Figure 21. Immunocytochemistry of transiently transfected MCF7 cells with YPEL antibody. MCF7 cells grown on coverslips in 12-well cell culture plates for 48 hours after seeding were transfected. 36 hours after transfection, cells were fixed with 2% paraformaldehyde, permeabilized with 0.4% Triton-X100. After blocking with 10% normal goat serum (NGS) for one hour, cells were incubated with the S-14 Ypel antibody (1:50 in 2% NGS) for two hours. An Alexa Fluor®-488 (emission at green channel) conjugated goat anti-rabbit (1:1000 in 2% NGS) secondary antibody was used for the visualization of Ypel. Nuclei were counterstained with DAPI (emission at blue channel). It should be noted that light emissions were collected using the aforementioned channels. However, to better visualize the overlapping regions we were in the need of using artificial coloring; accordingly the Ypel staining was colored in blue and DAPI is in red. Merge denotes the overlapped images taken by different emission channels. Images are from a representative experiment performed two independent times with similar results.

All ICC studies were performed 36 hours after transfections. To examine the possibility that these *YPEL* over-expression-induced alterations in the nuclear membrane reflected as diffusion of DNA into the cytoplasm are time-dependent, we selected to use *Flag-YPEL2* to address this issue. We performed ICC 18, 24 and 36 hours after transfections using the Flag-M2 or S-14 Ypel antibody (Figure 22). At 18h post-transfection, we observed two patterns of staining: 1) Ypel2 and DNA show an overlapping staining in the nucleus and there is no DNA diffusion into the cytoplasm (indicated by the arrow in Figure 21A), and 2) DNA diffuses into the cytoplasm and shows a perfectly overlapped staining with Ypel2, reminiscing of the previous staining pattern observed at 36h of transfections. Indeed, ICC performed at 24h and 36h after transfection also gave virtually identical results in that all Ypel2 expressing cells displayed an overlap staining between Ypel2 and DAPI, the latter which showed staining outside of the nucleus as well.

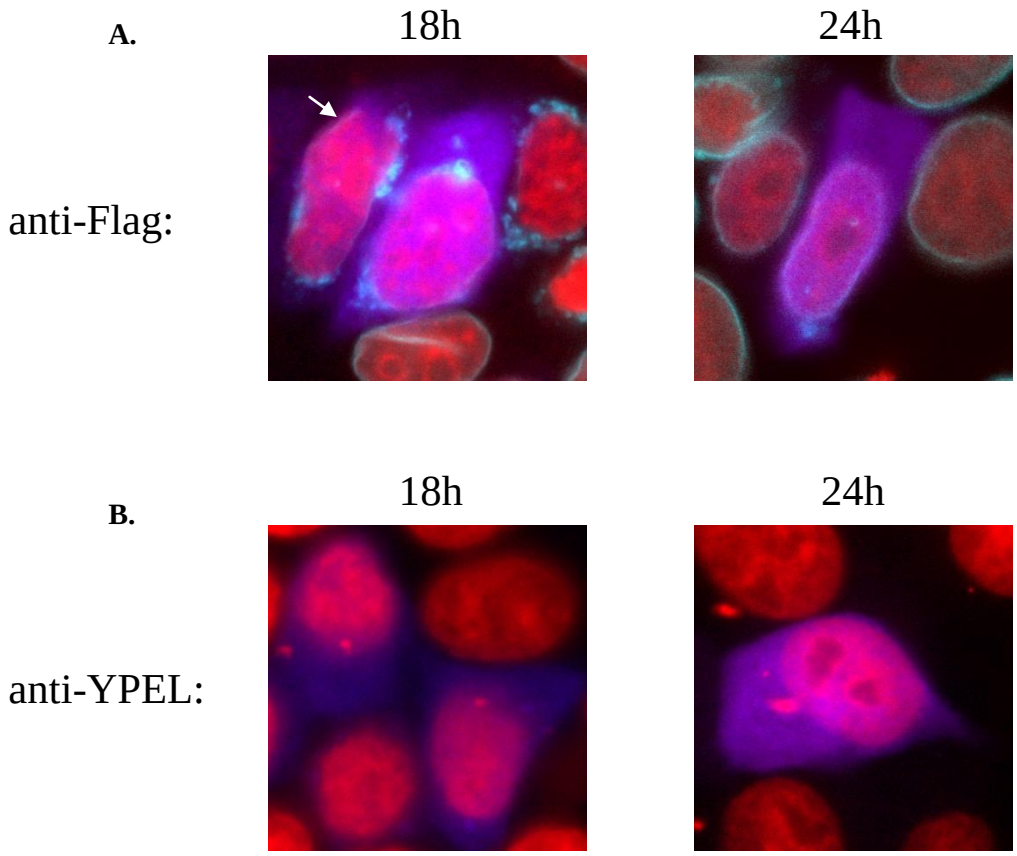


Figure 22. ICC of *YPEL2* transiently transfected MCF7 cells with FLAG and YPEL antibodies. MCF7 cells grown on coverslips in 12-well cell culture plates for 48 hours after seeding were transfected. 18 or 24 hours after transfection, cells were fixed with 2% paraformaldehyde, permeabilized with 0.4% Triton-X100. **A)** After blocking with 10% bovine serum albumin (BSA) for one hour, cells were incubated with the Flag-M2 antibody (1:250 in 3% BSA) for two hours and then with the LaminB1 antibody (1:200 in 3% BSA) for two hours, sequentially. An Alexa Fluor®-488 (emission at green channel) conjugated goat anti-mouse and an Alexa Fluor®-647 (emission at red channel) conjugated goat anti-rabbit (1:1000 and 1:200 in 3% BSA, respectively) secondary antibodies were used for the Flag-M2 and LaminB1 primary antibodies, respectively for visualization. Nuclei were counterstained with DAPI (emission at blue channel) **B)** After blocking with 10% normal goat serum (NGS) for one hour, cells were incubated with the S-14 Ypel antibody (1:50 in 2% NGS) for two hours. An Alexa Fluor®-488 conjugated (emission at green channel) goat anti-rabbit secondary antibody (1:1000 in 2% NGS) was used for the S-14 Ypel antibody. Nuclei were counterstained with DAPI. It should be noted that light emission were collected using the aforementioned channels. However, to better visualize the overlapping regions we were in the need of using artificial coloring; accordingly the Ypel or Flag staining was colored in blue, LaminB1 is in cyan, where applicable, and DAPI is in red. Merge denotes the overlapped images taken by different emission channels. Images are from a representative experiment performed two independent times with similar results.

3.5.2. Western Blot

As with ICC, we could not detect any endogenous Ypel protein in WB using 100 µg of total cell extracts from un-transfected MCF7 cells. Cells were transfected with pcDNA3.1 (-) without (empty vector, EV) or with a *YPEL* cDNA for 36h. After SDS-PAGE, WB analysis was carried out with the S-14 Ypel or the Flag-M2 antibody (Figure 23).

Similar to WB analyses of COS7 cells (Figure 16), Flag-M2 recognized Ypel1-5 confirming the synthesis at expected molecular masses while no protein was detected in extracts of cells transfected with the parent vector (EV). The use of the S-14 Ypel antibody allowed us to confirm that we cannot detect an endogenous Ypel protein and that the S-14 Ypel antibody recognized readily Ypel1-3 and but not Ypel4 and 5.

It should be noted that we performed additional WB analyses using nuclear and cytoplasmic extracts of MCF7 and various concentrations of the primary and/or secondary antibodies (data not shown). However, we could not detect any Ypel protein in a reproducible and reliable manner.

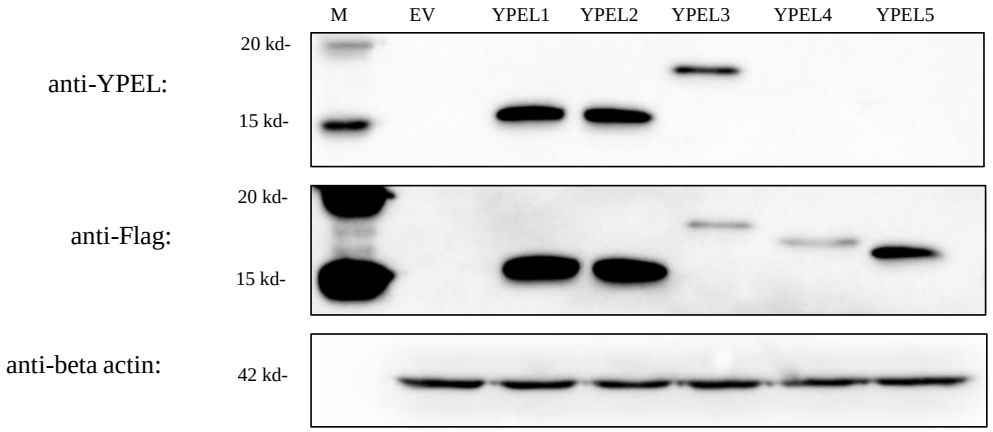


Figure 23. Western blot analysis in MCF7 cells. MCF7 cells grown on six-well cell culture plates for 48 hours after seeding were transfected with pcDNA3.1 expression vector bearing none (empty vector EV) or a *YPEL* cDNA. Cells 36h after transfection were subjected to protein extraction. 100 µg of total protein extracts from each transfection was subjected to a 15 % SDS-PAGE followed by WB analysis using the S-14 Ypel or the Flag-M2 antibody. The beta actin antibody was used as loading control. Experiments were repeated two independent times with similar results.

Based on ICC and WB analyses of MCF7 cells, we conclude that:

1. As in COS7 cells, we confirmed the synthesis of Flag-Ypel proteins detected by the Flag-M2 antibody under native and denaturing conditions.

2. Although at varying effectiveness, the S-14 Ypel antibody can recognize Ypel1-3 but not Ypel4 and 5 in ICC and WB.

3. We detected no protein suggestive of endogenous Ypel1, 2 or 3 in ICC or in WB. Since, the S-14 Ypel antibody does not recognize Ypel4 or Ypel5, we are not certain if Ypel4 or 5 is expressed in MCF-7 cells. Our inability to detect an endogenous Ypel protein could be due to:

   a. Protein levels of Ypel1-3 are very low in MCF7 cells. To examine this possibility, we performed WBs using 200 µg nuclear and cytoplasmic proteins but failed to obtain reproducible results (data not shown).

   b. As *YPEL2* and *YPEL3* are repressed by E2 at mRNA level, we anticipated that E2 could consequently affect Ypel2 and Ypel3 protein

levels. To test this prediction, we used cellular extracted from MCF7 cells treated with E2 and/or ICI. However, we could not detect in a reliable manner an endogenous protein suggestive of Ypel (data not shown).

c. The protein levels of Ypels might be low in cancer cells in general rather than specific to MCF7 cells. To examine this possibility, we screened by WB analyses of MCF10A (non-tumorigenic epithelial cell line derived from human breast tissue), MDA-MB-231 (triple negative, that is estrogen receptor, progesterone receptor and human epidermal growth factor receptor 2 negative, breast adenocarcinoma cell line), T47D (ER-positive breast ductal carcinoma) and HeLa (cell line derived from cervical adenocarcinoma) cell lines. However, we failed to suggest that these cells do synthesize Ypel proteins (data not shown).

It is possible that *YPEL*s are expressed only under certain conditions or at specific stages of cell cycle. We plan to examine YPEL expressions and Ypel synthesis under various conditions that include oxidative stress, DNA damage or mitotic arrest and cell cycle stages of growth synchronized cells.

We should also note that our several attempts to obtain MCF7 cell lines that stably express an endogenously introduced Flag-YPEL cDNA have failed. Although we are yet uncertain, the absence of a selectable sub-colony that synthesizes Flag-Ypel appears to be due to the lack of the Flag-Ypel2 synthesis (Figure F. 3) despite the fact that subclones have incorporated cDNAs (Figure F. 1) and express them (Figure F. 2).

These observations suggest that Ypel synthesis in stably transfected cells is regulated post-transcriptionally. We plan to address this issue as well.

CHAPTER 4

CONCLUSION AND FUTURE DIRECTIONS

Through extensive *in silico* analyses we predict that:

1. The high level of similarity among Ypel1-5 observed in nucleotide and amino-acid sequences is also manifested as similar predicted tertiary structures. We are in the process of recombinant Ypel2 protein production and purification. We are aiming to use the purified Ypel2 protein in atomic force microscopy and/or NMR studies to determine the protein structure, which could be a model for the other Ypel proteins.

2. Chromosomal regions where *YPEL1-5* are located are syntenic between the human and mouse genome suggest the ancestor specie of human and mouse have already acquired all five *YPEL* genes. This observation supports the suggestion that gene numbers in *YPEL* family are correlated with organismal complexity. The presence of five YPEL members in the common ancestor of human and mouse, or mammals in general, suggests all five genes are essential and each Ypel protein has specialized functions. Alternatively, they have a common function implying a functional redundancy in critical cellular processes.

3. Ypel1-5 contain the Yippee/Mis18 domain and share similarity to FAD/NAD binding oxidoreductases. Ypel proteins, like Mis18 and related proteins might be involved in microtubule-chromosome interactions. They may, on the other hand, be involved in responses to various stress conditions, including oxidative stress which poses a problem for many organisms whether simple or complex.

4. The common functionality for Ypel1-5 suggests their involvement in essential cellular mechanisms. Ypel5 exhibits more similarities to Yippee, indicating the ancestral protein diverged into two main branches: one giving rise to Ypel5 and other giving rise to Ypel1-4.

5. Ypel1-5 and Yippee are predicted to be regulated by post translational modifications involving phosphorylation, glycosylation and myristoylation that could further diversify their functions. These predictions again point out that Ypel5 is more similar to Yippee than Ypel1-4.

6. Each *YPEL* gene appears to be regulated by a different set of TFs. Despite being highly similar, each could be expressed to function in response to different signals converging on the same cellular process or the execution of diverse functions at different times or tissues. Analysis of promoter regions of each *YPEL* with reporter assays under different conditions could shed light on this possibility.

7. *YPEL2* and *3* are E2-ER responsive genes. Both gene loci contain several ERE-like sequences, functionalities of which will be assessed by ChIP, EMSA and luciferase reporter assays.

Based on our experimental studies in COS7 and MCF7 cells, we conclude that:

1. The S-14 Ypel antibody recognizes Ypel1, Ypel2 and Ypel3 isoform 1 well, but it recognizes Ypel4 very poorly under both the native and denaturing conditions. The S-14 Ypel antibody however does not recognizeYpel5. This is also consistent with our *in silico* prediction that Ypel5 is distinct from Ypel1-4.

2. COS7 cells endogenously synthesize Ypel protein(s). This Ypel protein localizing to a region just outside of the nucleus could be Ypel1, 2, and/or Ypel3 isoform 2, but not Ypel4 or 5. To determine which Ypel protein is synthesized in COS7 cells, we plan to use siRNA targeting each *YPEL* specifically.

3. We found that levels of Ypel protein(s) are low in COS7 cells and are not detectable in MCF7 cells. There appears to be mechanism that prevents the synthesis or induce degradation of Ypel protein(s). We plan to elucidate the underlying mechanisms by:

    a. Identifying signaling mechanisms that are activated under certain conditions. For this purpose, we have started our work on Yippee homolog in *S. cerevisae, MOH1*. We aim to identify conditions that regulate Moh1 levels in yeast, where there is only one Yippee-related protein and cellular events are controlled and studied more easily compared to mammalian systems. We have already some findings that indicate the involvement of Moh1 in nutrient stress, where *MOH1* knockout yeast strain has a low growth rate in restrictive media.

    b. Studying *YPEL* expressions in synchronous cell populations. Given the previous findings suggesting that the intracellular locations of Ypel1-5 change depending on the cell cycle stages, we envision that Ypel1-5 synthesis could also be cell cycle-dependent. For this purpose, we have started initial studies utilizing a thymidine-double block approach to synchronize cells.

    c. Identifying possible non-coding RNAs and/or RNA binding proteins that could be regulating Ypel protein levels post-transcriptionally.

    d. Preventing proteasome and/or lysosome dependent degradation of Ypel proteins. We have performed initial studies using proteasome inhibitor MG132 to address this issue. We will also study the effects of lysosomal degradation pathways on protein levels of Ypel.

4. We showed that E2 represses the expression of *YPEL2* and *YPEL3* genes in a time dependent manner in MCF7 cells. Also, we found that ER regulates both the basal and E2-mediated expression of *YPEL2* while it has no effect on the basal expression *YPEL3*. We will confirm our results in T47D, an

ER positive breast cancer cell line, as well as ER-negative MDA-MB-231 cell line stably transfected with ERα and the ERE binding defective ($ER\alpha_{DBD}$) cDNA. Furthermore, we will test the function of the predicted ERE sequences on gene loci with *in situ* and *in vitro* approaches to verify that *YPEL2* and/or *YPEL3* are *bona fide* estrogen responsive genes.

5. We have observed that over-expressions of *YPEL1-5* cause a diffusion of DNA from the nucleus into the cytoplasm in a pattern that overlaps with each Ypel protein. This observation suggests that Ypel proteins are directly or indirectly associated with DNA. Ypel proteins may be involved in processes regulating DNA and/or nucleus architecture during cell division and/or death. This was based on our observations that the DNA leakage is associated with disruptions in the nuclear lamina of MCF7 cells but not of COS7 cells. This cell-specific effect on the nuclear lamina is likely due to the origin (human breast *vs.* monkey kidney) and/or type (cancerous *vs.* transformed) of the cells we used. It is also possible that the absence of an endogenous Ypel synthesis in MCF7 cells renders the cells more vulnerable to the negative effects brought about by Ypel over-expressions as compared to COS7 cells.

REFERENCES

[1] Y. Huang, X. Li, and M. Muyan, "Estrogen receptors similarly mediate the effects of 17β-estradiol on cellular responses but differ in their potencies.," *Endocrine*, vol. 39, no. 1, pp. 48–61, Feb. 2011.

[2] J. Huang, X. Li, R. Hilf, R. a Bambara, and M. Muyan, "Molecular basis of therapeutic strategies for breast cancer.," *Curr. Drug Targets. Immune. Endocr. Metabol. Disord.*, vol. 5, no. 4, pp. 379–96, Dec. 2005.

[3] K. Polyak, "On the birth of breast cancer.," *Biochim. Biophys. Acta*, vol. 1552, no. 1, pp. 1–13, Nov. 2001.

[4] P. Ascenzi, A. Bocedi, and M. Marino, "Structure-function relationship of estrogen receptor (alpha) and (beta): Impact on human health," *Mol. Aspects Med.*, vol. 27, no. 4, pp. 299–402, 2006.

[5] Y. Bai and V. Giguére, "Isoform-selective interactions between estrogen receptors and steroid receptor coactivators promoted by estradiol and ErbB-2 signaling in living cells.," *Mol. Endocrinol.*, vol. 17, pp. 589–599, 2003.

[6] S. L. Nott, Y. Huang, X. Li, B. R. Fluharty, X. Qiu, W. V Welshons, S. Yeh, and M. Muyan, "Genomic responses from the estrogen-responsive element-dependent signaling pathway mediated by estrogen receptor alpha are required to elicit cellular alterations.," *J. Biol. Chem.*, vol. 284, no. 22, pp. 15277–88, May 2009.

[7] C. E. Mason, F.-J. Shu, C. Wang, R. M. Session, R. G. Kallen, N. Sidell, T. Yu, M. H. Liu, E. Cheung, and C. B. Kallen, "Location analysis for the estrogen receptor-alpha reveals binding to diverse ERE sequences and widespread binding within repetitive DNA elements.," *Nucleic Acids Res.*, vol. 38, no. 7, pp. 2355–68, Apr. 2010.

[8] M. Muyan, L. M. Callahan, Y. Huang, and A. J. Lee, "The ligand-mediated nuclear mobility and interaction with estrogen-responsive elements of estrogen receptors are subtype specific," *Journal of Molecular Endocrinology*, vol. 49. pp. 249–266, 2012.

[9] K. Roxström-Lindquist and I. Faye, "The Drosophila gene Yippee reveals a novel family of putative zinc binding proteins highly conserved among eukaryotes.," *Insect Mol. Biol.*, vol. 10, no. 1, pp. 77–86, Feb. 2001.

[10] K. Roxström-lindquist, “Innate Immunity in Insects , Function and Regulation of Hemolin from Hyalophora cecropia,” University of Stockholm, 2002.

[11] P. Farlie, C. Reid, S. Wilcox, J. Peeters, G. Reed, and D. Newgreen, “Ypel1: a novel nuclear protein that induces an epithelial-like morphology in fibroblasts.,” *Genes Cells*, vol. 6, no. 7, pp. 619–29, Jul. 2001.

[12] S. Aerts, D. Lambrechts, S. Maity, P. Van Loo, B. Coessens, F. De Smet, L.-C. Tranchevent, B. De Moor, P. Marynen, B. Hassan, P. Carmeliet, and Y. Moreau, “Gene prioritization through genomic data fusion.,” *Nat. Biotechnol.*, vol. 24, no. 5, pp. 537–44, May 2006.

[13] J. E. Olson, X. Wang, V. S. Pankratz, Z. S. Fredericksen, C. M. Vachon, R. a Vierkant, J. R. Cerhan, and F. J. Couch, “Centrosome-related genes, genetic variation, and risk of breast cancer.,” *Breast Cancer Res. Treat.*, vol. 125, no. 1, pp. 221–8, Jan. 2011.

[14] K. Hosono, T. Sasaki, S. Minoshima, and N. Shimizu, “Identification and characterization of a novel gene family YPEL in a wide spectrum of eukaryotic species.,” *Gene*, vol. 340, no. 1, pp. 31–43, Sep. 2004.

[15] K. Hosono, S. Noda, A. Shimizu, N. Nakanishi, M. Ohtsubo, N. Shimizu, and S. Minoshima, “YPEL5 protein of the YPEL gene family is involved in the cell cycle progression by interacting with two distinct proteins RanBPM and RanBP10.,” *Genomics*, vol. 96, no. 2, pp. 102–11, Aug. 2010.

[16] S. J. Baker, “Small unstable apoptotic protein, an apoptosis-associated protein, suppresses proliferation of myeloid cells.,” *Cancer Res.*, vol. 63, pp. 705–712, 2003.

[17] K. D. Kelley, K. R. Miller, A. Todd, A. R. Kelley, R. Tuttle, and S. J. Berberich, “YPEL3, a p53-regulated gene that induces cellular senescence.,” *Cancer Res.*, vol. 70, no. 9, pp. 3566–75, May 2010.

[18] R. Tuttle, K. R. Miller, J. N. Maiorano, P. M. Termuhlen, Y. Gao, and S. J. Berberich, “Novel senescence associated gene, YPEL3, is repressed by estrogen in ER+ mammary tumor cells and required for tamoxifen-induced cellular senescence.,” *Int. J. Cancer*, vol. 130, no. 10, pp. 2291–9, May 2012.

[19] R. Tuttle, M. Simon, D. C. Hitch, J. N. Maiorano, M. Hellan, J. Ouellette, P. Termuhlen, and S. J. Berberich, “Senescence-associated gene YPEL3 is downregulated in human colon tumors.,” *Ann. Surg. Oncol.*, vol. 18, no. 6, pp. 1791–6, Jun. 2011.

[20] P. Liang, Y. Wan, Y. Yan, Y. Wang, N. Luo, Y. Deng, X. Fan, J. Zhou, Y. Li, Z. Wang, W. Yuan, M. Tang, X. Mo, and X. Wu, "MVP interacts with YPEL4 and inhibits YPEL4- mediated activities of the ERK signal pathway," vol. 450, pp. 445–450, 2010.

[21] W. Berger, E. Steiner, M. Grusch, L. Elbling, and M. Micksche, "Vaults and the major vault protein: novel roles in signal pathway regulation and immunity.," *Cell. Mol. Life Sci.*, vol. 66, no. 1, pp. 43–61, Jan. 2009.

[22] M. H. Mossink, A. van Zon, R. J. Scheper, P. Sonneveld, and E. a C. Wiemer, "Vaults: a ribonucleoprotein particle involved in drug resistance?," *Oncogene*, vol. 22, no. 47, pp. 7458–67, Oct. 2003.

[23] D. Y. Jun, H. W. Park, and Y. H. Kim, "YPEL5, T cells, anti-CD3.pdf," *J. Life Sci.*, vol. 17, no. 12, pp. 1641–1648, 2007.

[24] T. Velusamy, N. Palanisamy, S. Kalyana-Sundaram, A. A. Sahasrabuddhe, C. a Maher, D. R. Robinson, D. W. Bahler, T. T. Cornell, T. E. Wilson, M. S. Lim, A. M. Chinnaiyan, and K. S. J. Elenitoba-Johnson, "Recurrent reciprocal RNA chimera involving YPEL5 and PPP1CB in chronic lymphocytic leukemia.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 8, pp. 3035–40, Mar. 2013.

[25] J. Russo, K. Snider, J. S. Pereira, and I. H. Russo, "Estrogen induced breast cancer is the result in the disruption of the asymmetric cell division of the stem cell," *Horm Mol Biol Clin Investig.*, vol. 1, no. 2, pp. 53–65, 2010.

[26] J. Pärssinen, T. Kuukasjärvi, R. Karhu, and a Kallioniemi, "High-level amplification at 17q23 leads to coordinated overexpression of multiple adjacent genes in breast cancer.," *Br. J. Cancer*, vol. 96, no. 8, pp. 1258–64, Apr. 2007.

[27] R. C. Edgar, "MUSCLE: multiple sequence alignment with high accuracy and high throughput.," *Nucleic Acids Res.*, vol. 32, no. 5, pp. 1792–7, Jan. 2004.

[28] S. Hunter, P. Jones, A. Mitchell, R. Apweiler, T. K. Attwood, A. Bateman, T. Bernard, D. Binns, P. Bork, S. Burge, E. de Castro, P. Coggill, M. Corbett, U. Das, L. Daugherty, L. Duquenne, R. D. Finn, M. Fraser, J. Gough, D. Haft, N. Hulo, D. Kahn, E. Kelly, I. Letunic, D. Lonsdale, R. Lopez, M. Madera, J. Maslen, C. McAnulla, J. McDowall, C. McMenamin, H. Mi, P. Mutowo-Muellenet, N. Mulder, D. Natale, C. Orengo, S. Pesseat, M. Punta, A. F. Quinn, C. Rivoire, A. Sangrador-Vegas, J. D. Selengut, C. J. a Sigrist, M. Scheremetjew, J. Tate, M. Thimmajanarthanan, P. D. Thomas, C. H. Wu, C. Yeats, and S.-Y. Yong, "InterPro in 2011: new developments in the family and domain prediction database.," *Nucleic Acids Res.*, vol. 40, no. Database issue, pp. D306–12, Jan. 2012.

[29] H. Dinkel, K. Van Roey, S. Michael, N. E. Davey, R. J. Weatheritt, D. Born, T. Speck, D. Krüger, G. Grebnev, M. Kuban, M. Strumillo, B. Uyar, A. Budd, B. Altenberg, M. Seiler, L. B. Chemes, J. Glavina, I. E. Sánchez, F. Diella, and T. J. Gibson, "The eukaryotic linear motif resource ELM: 10 years and counting.," *Nucleic Acids Res.*, vol. 42, no. Database issue, pp. D259–66, Jan. 2014.

[30] E. de Castro, C. J. a Sigrist, A. Gattiker, V. Bulliard, P. S. Langendijk-Genevaux, E. Gasteiger, A. Bairoch, and N. Hulo, "ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins.," *Nucleic Acids Res.*, vol. 34, no. Web Server issue, pp. W362–5, Jul. 2006.

[31] L. a Kelley and M. J. E. Sternberg, "Protein structure prediction on the Web: a case study using the Phyre server.," *Nat. Protoc.*, vol. 4, no. 3, pp. 363–71, Jan. 2009.

[32] V. B. Bajic, "Dragon ERE Finder version 2: a tool for accurate detection and analysis of estrogen response elements in vertebrate genomes," *Nucleic Acids Res.*, vol. 31, no. 13, pp. 3605–3607, Jul. 2003.

[33] J. Ye, G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen, and T. L. Madden, "Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction.," *BMC Bioinformatics*, vol. 13, p. 134, Jan. 2012.

[34] M. B. Lyng, A.-V. Laenkholm, N. Pallisgaard, and H. J. Ditzel, "Identification of genes for normalization of real-time RT-PCR data in breast carcinomas.," *BMC Cancer*, vol. 8, p. 20, Jan. 2008.

[35] S. Fleige, V. Walf, S. Huch, C. Prgomet, J. Sehm, and M. W. Pfaffl, "Comparison of relative mRNA quantification models and the impact of RNA integrity in quantitative real-time RT-PCR.," *Biotechnol. Lett.*, vol. 28, no. 19, pp. 1601–13, Oct. 2006.

[36] R. Goni, P. García, and S. Foissac, "The qPCR data statistical analysis," no. September, pp. 1–9, 2009.

[37] M. Kozak, "Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes," *Cell*, vol. 44, no. 2, pp. 283–292, Jan. 1986.

[38] T. Hayashi, Y. Fujita, O. Iwasaki, Y. Adachi, K. Takahashi, and M. Yanagida, "Mis16 and Mis18 Are Required for CENP-A Loading and Histone Deacetylation at Centromeres," *Cell*, vol. 118, no. 6, pp. 715–729, Aug. 2014.

[39] M. a Reeves and P. R. Hoffmann, “The human selenoproteome: recent insights into functions and regulation.,” *Cell. Mol. Life Sci.*, vol. 66, no. 15, pp. 2457–78, Aug. 2009.

[40] A. Hansel, S. H. Heinemann, and T. Hoshi, “Heterogeneity and function of mammalian MSRs: enzymes for repair, protection and regulation.,” *Biochim. Biophys. Acta*, vol. 1703, no. 2, pp. 239–47, Jan. 2005.

[41] E. R. Stadtman, “Protein oxidation and aging.,” *Free Radic. Res.*, vol. 40, no. 12, pp. 1250–8, Dec. 2006.

[42] B. M. Walter, C. Nordhoff, G. Varga, G. Goncharenko, S. W. Schneider, S. Ludwig, and V. Wixler, “Mss4 protein is a regulator of stress response and apoptosis.,” *Cell Death Dis.*, vol. 3, no. 4, p. e297, Jan. 2012.

[43] T. Satoh, H. Kato, Y. Kumagai, M. Yoneyama, S. Sato, K. Matsushita, T. Tsujimura, T. Fujita, S. Akira, and O. Takeuchi, “LGP2 is a positive regulator of RIG-I- and MDA5-mediated antiviral responses.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 4, pp. 1512–7, Jan. 2010.

[44] X. Li, S. L. Nott, Y. Huang, R. Hilf, R. a Bambara, X. Qiu, A. Yakovlev, S. Welle, and M. Muyan, “Gene expression profiling reveals that the regulation of estrogen-responsive element-independent genes by 17 beta-estradiol-estrogen receptor beta is uncoupled from the induction of phenotypic changes in cell models.,” *J. Mol. Endocrinol.*, vol. 40, no. 5, pp. 211–29, May 2008.

[45] S. Malik, S. Jiang, J. P. Garee, E. Verdin, A. V Lee, B. W. O’Malley, M. Zhang, N. S. Belaguli, and S. Oesterreich, “Histone deacetylase 7 and FoxA1 in estrogen-mediated repression of RPRM.,” *Mol. Cell. Biol.*, vol. 30, no. 2, pp. 399–412, Jan. 2010.

[46] K. a Green and J. S. Carroll, “Oestrogen-receptor-mediated transcription and the influence of co-factors and chromatin state.,” *Nat. Rev. Cancer*, vol. 7, no. 9, pp. 713–22, Sep. 2007.

[47] J. Huang, X. Li, C. A. Maguire, R. Hilf, R. A. Bambara, and M. Muyan, “Binding of estrogen receptor beta to estrogen response element in situ is independent of estradiol and impaired by its amino terminus.,” *Mol. Endocrinol.*, vol. 19, pp. 2696–2712, 2005.

[48] J. Frasor, J. M. Danes, B. Komm, K. C. N. Chang, C. R. Lyttle, and B. S. Katzenellenbogen, “Profiling of estrogen up- and down-regulated gene expression in human breast cancer cells: insights into gene networks and pathways underlying estrogenic control of proliferation and cell phenotype.,” *Endocrinology*, vol. 144, no. 10, pp. 4562–74, Oct. 2003.

[49] J. S. Carroll, C. A. Meyer, J. Song, W. Li, T. R. Geistlinger, J. Eeckhoute, A. S. Brodsky, E. K. Keeton, K. C. Fertuck, G. F. Hall, Q. Wang, S. Bekiranov, V. Sementchenko, E. A. Fox, P. A. Silver, T. R. Gingeras, X. S. Liu, and M. Brown, "Genome-wide analysis of estrogen receptor binding sites.," *Nat. Genet.*, vol. 38, pp. 1289–1297, 2006.

[50] C. Creighton, K. Cordero, J. Larios, R. Miller, M. Johnson, A. Chinnaiyan, M. Lippman, and J. Rae, "Genes regulated by estrogen in breast tumor cells in vitro are similarly regulated in vivo in tumor xenografts and human breast tumors.," *Genome Biol*, vol. 7, p. R28, 2006.

[51] F. Stossi, Z. Madak-Erdogan, and B. S. Katzenellenbogen, "Estrogen receptor alpha represses transcription of early target genes via p300 and CtBP1.," *Mol. Cell. Biol.*, vol. 29, no. 7, pp. 1749–59, Apr. 2009.

[52] C.-Y. Lin, A. Ström, V. B. Vega, S. L. Kong, A. L. Yeo, J. S. Thomsen, W. C. Chan, B. Doray, D. K. Bangarusamy, A. Ramasamy, L. a Vergara, S. Tang, A. Chong, V. B. Bajic, L. D. Miller, J.-A. Gustafsson, and E. T. Liu, "Discovery of estrogen receptor alpha target genes and response elements in breast tumor cells.," *Genome Biol.*, vol. 5, no. 9, p. R66, Jan. 2004.

APPENDIX A

PRIMERS

Table A 1. Primers used in the study

Primer Name	Sequence (5'-3')
YPEL1_3UTR_FP	CCCTTCTCCTTTGAATGCTGTTTTGTG
YPEL1_3UTR_REP	CACCTTACCCACAGAGATGGCC
YPEL2_3UTR_FP	CAGCATCTACCCAACCCAGTGTCC
YPEL2_3UTR_REP	GATGGCGTCAGGGTGGGAGG
YPEL3_3UTR_FP	GCATGCACTGTGACCTTGGG
YPEL3_3UTR_REP	CTATAGGGCAGGTGGGGCAGG
YPEL4_3UTR_FP	GCCAAGCAGTCTATACCAGCATGAG
YPEL4_3UTR_REP	CTGTCTCTCCCTGACTGCTGCC
YPEL5_3UTR_FP	CGGATTAATGAACTGCGGAACAAGAGG
YPEL5_3UTR_REP	GCATAATATTCTCTCTACACACTGCTGTTG
YPEL1_5UTR_FP	GCTTTTGTCAGAGAGGCCAGC
YPEL2_5UTR_FP	CCATCCTGTGGGAGTGCCTCG
YPEL3_5UTR_FP	CCAGTGTGACAGAGCGAGTCC
YPEL4_5UTR_FP	GGCCAGCTTTGCAGCCCCAGT
YPEL5_5UTR_FP	GCATACTTGTACTGGGTTAGC C
XNmetN_YPEL1_FP	CGCATCTCGAGACCATGGCTAGCGCACTGA GATGCATCCACTCAGCA
YPEL1_EpolyaB_REP	CGCATGGGATCCTTTATTAGAATTCCTCAAAGGGAAGGGTCA GTTTAAG

Table A.1. (continued)

XNmetN_YPEL2_FP	CGCATCTCGAGACCATGGCTAGCGTGAAGATGACAAGATCG AAG
YPEL2_EpolyaB_REP	CGCATGGATCCTTTATTAGAATTCGTCCCAGCCATTGTC
XNmetN_YPEL3_FP	CGCATCTCGAGACCATGGCTAGCTGTGTGGCCCAGGTCCTGA CAGCC
YPEL3_EpolyaB_REP	CGCATGGGATCCTTTATTAGAATTCGTCCCAGCCGTTGTCTTT GATCAT
XNmetN_YPEL4_FP	CGCATCTCGAGACCATGGCTAGCCCCAGCTGTGACCCCGGTC CGGGC
YPEL4_EpolyaB_REP	CGCATGGGATCCTTTATTAGAATTCGTCCCAGCCGTTGTCCTT CACCA
XNmetN_YPEL5_FP	CGCATCTCGAGACCATGGCTAGCGGCAGAA TTTTCCTTGATCATATC
YPEL5_EpolyaB_REP	CGCATGGGATCCTTTATTAGAATTCAGAGTTATCAGATGGTA CATGCTC
GAPDH_FP	GGGAGCCAAAAGGGTCATCA
GAPDH_REP	TTTCTAGACGGCAGGTCAGGT
pS2/TFF1_FP	TTGTGGTTTTCCTGGTGTCA
pS2/TFF1_REP	CCGAGCTCTGGGACTAATCA
RPLP0_FP	GGAGAAACTGCTGCCTCATA
RPLP0_REP	GGAAAAAGGAGGTCTTCTCG
PUM1_FP	AGTGGGGGACTAGGCGTTAG
PUM1_REP	GTTTTCATCACTGTCTGCATCC

APPENDIX B

CHARCOAL DEXTRAN TREATMENT OF FETAL BOVINE SERUM

To 500 ml fetal bovine serum, 10 g of Charcoal, dextran coated (C6241, Sigma Aldrich, Germany), is added. The mixture is incubated overnight at 4°C with agitation. Charcoal is pelleted by centrifuging at 10800xg for 30 minutes. Supernatant is pelleted again by centrifuging at 10800xg for an additional 30 minutes. Supernatant is filtered through a 0.45 μM sterile filter unit (Corning, polystyrene, cellulose acetate membrane, low protein binding) in a biological safety cabinet. CD-FBS is filtered once again through 0.22μM filter while adding to culture media.

APPENDIX C

CONTROL FOR GENOMIC DNA CONTAMINATION

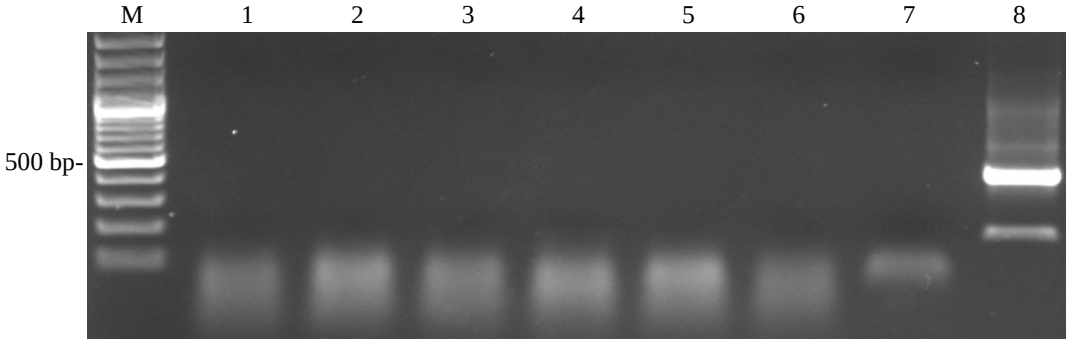


Figure C. 1. Lack of DNA contamination in RNA samples. 500 ng of RNA samples were used as template in PCR using *GAPDH* specific primers: GAPDH_F: 5'-GGGAGCCAAAAGGGTCATCA-3' and GAPDH_R: 5'-TTTCTAGACGGCAGGTCA GGT-3'. PCR conditions were as follows: incubation at 95^0C for 10 minutes, 40 cycles of 95^0C for 30 seconds, 65^0C for 30 seconds, and 72^0C for 30 seconds, and final extension at 72^0C for 5 minutes. M: DNA ladder, 1: EtOH-3h, 2: E2-3h, 3: EtOH-6h, 4: E2-6h, 5: EtOH-24h, 6: E2-24h, 7: no template reaction as negative control, 8: 50 ng genomic DNA as positive control. Image represents RNAs isolated from one set of E2 treated MCF7 cells. All other RNAs were tested likewise.

APPENDIX D

MIQE CHECKLIST

Table D 1. MIQE Checklist

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

Table D.1. (continued)

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

Table D.1. (continued)

qPCR PROTOCOL		
Complete reaction conditions	**E**	**YES**
Description of normalization method	**E**	**YES**
Reaction volume and amount of cDNA/DNA	**E**	**YES**
Primer, (probe), Mg++ and dNTP concentrations	**E**	**N/A**
Polymerase identity and concentration	**E**	**N/A**
Buffer/kit identity and manufacturer	**E**	**YES**
Exact chemical constitution of the buffer	D	**N/A**
Additives (SYBR Green I, DMSO, etc.)	**E**	**YES**
Manufacturer of plates/tubes and catalog number	D	**NO**
Complete thermocycling parameters	**E**	**YES**
Reaction setup (manual/robotic)	D	**YES**
Manufacturer of qPCR instrument	**E**	**YES**
qPCR VALIDATION		
Evidence of optimization (from gradients)	D	**NO**
Specificity (gel, sequence, melt, or digest)	**E**	**YES**
For SYBR Green I, Cq of the NTC	**E**	**YES**
Standard curves with slope and y-intercept	**E**	**YES**
PCR efficiency calculated from slope	**E**	**YES**
Confidence interval for PCR efficiency or standard error	D	**NO**
r2 of standard curve	**E**	**YES**
Linear dynamic range	**E**	**YES**
Cq variation at lower limit	**E**	**YES**
Confidence intervals throughout range	D	**N/A**
Evidence for limit of detection	**E**	**NO**
If multiplex, efficiency and LOD of each assay.	**E**	**N/A**
DATA ANALYSIS		
qPCR analysis program (source, version)	**E**	**YES**
Cq method determination	**E**	**YES**
Outlier identification and disposition	**E**	**N/A**
Results of NTCs	**E**	**YES**
Justification of number and choice of reference genes	**E**	**YES**

Table D.1. (continued)

Number and concordance of biological replicates	D	**YES**
Number and stage (RT or qPCR) of technical replicates	**E**	**YES**
Repeatability (intra-assay variation)	E	**YES**
Reproducibility (inter-assay variation, %CV)	D	**YES**
Power analysis	D	**NO**
Statistical methods for result significance	**E**	**YES**
Software (source, version)	E	**YES**
Cq or raw data submission using RDML	**D**	**N/A**

E: essential information, D: desirable information, N/A: not applicable

APPENDIX E

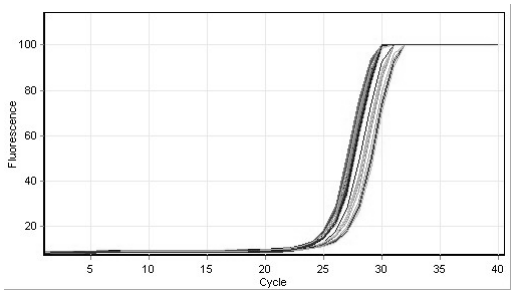
RT-QPCR ASSAY PERFORMANCE RESULTS

Results of *YPEL2* as the representative example, similar condiditons were used for *YPEL3*, *pS2/TFF1*, *RPLP0* and *PUM1*.

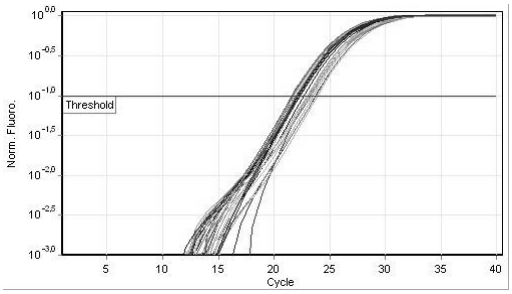
Quantitation information

Threshold	0,09822
Left Threshold	1,000
Standard Curve Imported	Yes
Standard Curve (1)	conc= 10^(-0,315*CT + 11,718)
Standard Curve (2)	CT = -3,179*log(conc) + 37,248
Reaction efficiency (*)	1,06338 (* = 10^(-1/m) - 1)
Start normalising from cycle	1
Noise Slope Correction	No
No Template Control Threshold	0%
Reaction Efficiency Threshold	Disabled
Normalisation Method	Dynamic Tube Normalisation
Digital Filter	Light
Sample Page	Page 1
Imported Analysis Settings	

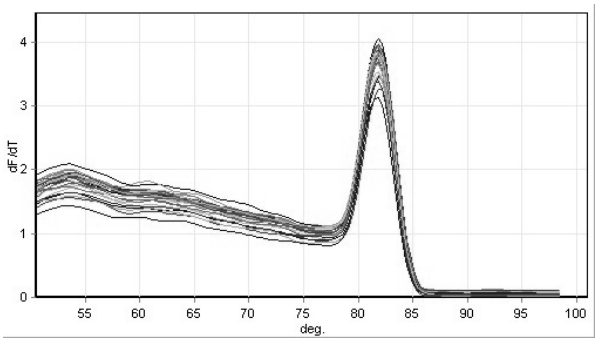
Raw data for Cycling A Green



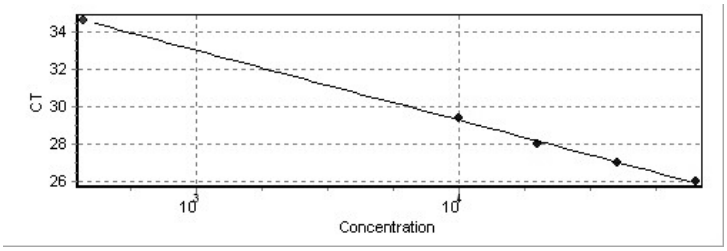
Quantitation data for Cycling A Green



Melt curve analysis



Standard curve



No.	Colour	Name	Type	Ct	Given Conc (copies/ml)	Calc Conc (copies/ml)	% Var
1		1:10 (e2 24)	Standard	26,00	80.000	75.975	5,0%
2		1:20 (e2 24	Standard	27,01	40.000	40.955	2,4%
3		1:40 (e2 24	Standard	28,00	20.000	22.275	11,4%
4		1:80 (e2 24	Standard	29,43	10.000	9.234	7,7%
5		nt	NTC	34,64		374	

APPENDIX F

GENERATION OF STABLE CELL LINE EXPRESSING FLAG-*YPEL2*

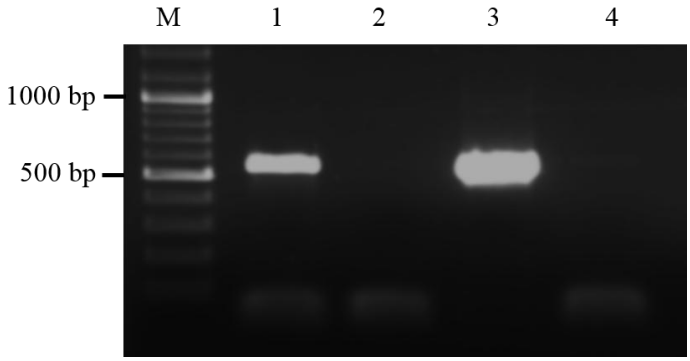


Figure F. 1. Genomic insertion of Flag-*YPEL2* construct into MCF7 cells. 50 ng of genomic DNA isolated from MCF7-Flag-*YPEL2* polyclonal cells was used as template in PCR performed with pcDNA3.1 (-) specific primers; T7_FP: 5' TAATACGACTCACTATAGGG-3' and BGH_REP: 5' TAGAAGGCACAGTCGAGGC-3'. Reaction was carried out with Taq polymerase (Thermo Scientific, USA) using KCl buffer, 2mM MgCl2, 500 nM of each primer and 200 µM from each dNTP. PCR conditions were as follows: incubation at 95^0C for 10 minutes, 40 cycles of 95^0C for 30 seconds, 55^0C for 30 seconds, and 72^0C for 30 seconds, and final extension at 72^0C for 5 minutes. M: DNA ladder, 1: Genomic DNA from MCF7-Flag-*YPEL2* polyclonal cells, 2: Genomic DNA from untransfected MCF7 cells, 3: pCDNA 3.1 (-) -Flag-*YPEL2* vector, as positive control 4: No template reaction, as negative control.

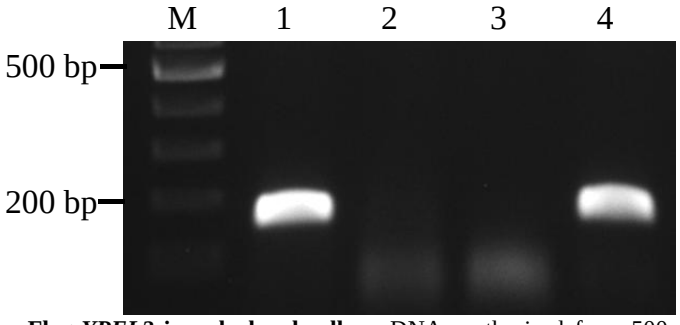


Figure F. 2. Expression Flag-*YPEL2* in polyclonal cells. cDNA synthesized from 500 ng of total RNA isolated from MCF7-Flag-*YPEL2* polyclonal cells was used as template in PCR performed with Flag_FP: 5'-GATTACAAGGATGACGACGATAAG-3' and a reverse primer within the ORF of YPEL2, Transfected YPEL2_REP: 5'-GTCCTTGACTTCCTTGGAATGACT-3'. Reaction was carried out with Taq polymerase (Thermo Scientific, USA) using KCl buffer, 2mM MgCl2, 500 nM of each primer and 200 µM from each dNTP. PCR conditions were as follows: incubation at 95^0C for 10 minutes, 40 cycles of 95^0C for 30 seconds, 55^0C for 30 seconds, and 72^0C for 30 seconds, and final extension at 72^0C for 5 minutes. M: DNA ladder, 1: cDNA synthesized from RNA isolated from MCF7-Flag-*YPEL2* polyclonal cells, 2: No reverse transcriptase reaction to ensure amplifications is not due to genomic DNA contamination of RNA sample, 3: No template reaction, as negative control 4: Genomic DNA isolated from MCF7-Flag-*YPEL2* polyclonal cells, as positive control.

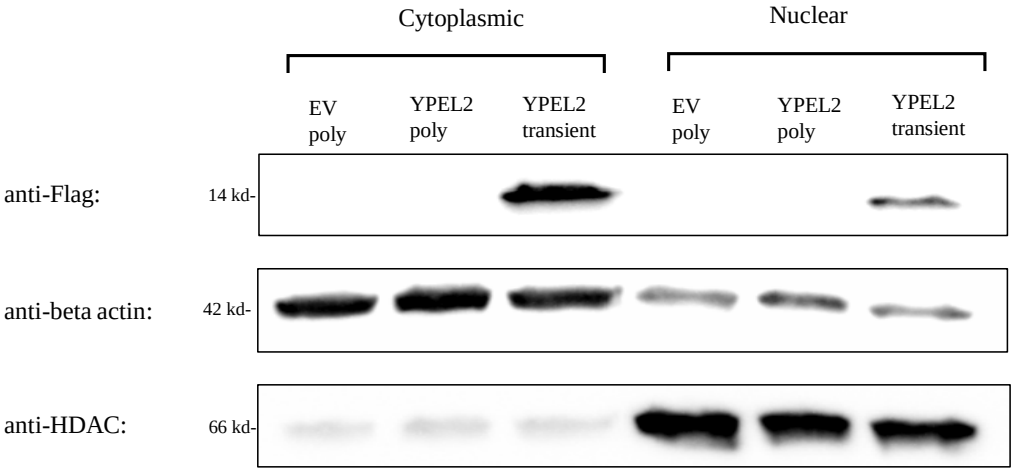


Figure F. 3. WB analysis of MCF7-Flag-*YPEL2* polyclonal cells. 100 µg of cytoplasmic and nuclear extracts isolated from MCF7-pCDNA3.1 empty vector polyclonal cells (EV poly), MCF7-Flag-*YPEL2* polyclonal cells (YPEL2 poly) and MCF7 cells transiently transfected with pcDNA3.1 (-) Flag-*YPEL2* (YPEL2 transient) were loaded to 15 % SDS-PAGE. Flag-M2 antibody was used to detect Flag-Ypel2 protein in WB with aforementioned conditions. There is no Flag-Ypel2 detection in EV-polyclonal cells, confirming that the detection of Flag-Ypel2 is antibody specific. Although Flag-*YPEL2* is inserted and expressed in MCF7-Flag-*YPEL2*-polyclonal cells, there is no Flag-Ypel2 synthesis. Flag-Ypel2 is detected in both cytoplasmic and nuclear extracts of MCF7 cells transiently transfected with pcDNA3.1 (-) Flag-*YPEL2*, used at positive control. Beta actin and HDAC antibodies were used as the cytoplasmic and nuclear loading controls, respectively. Experiments were performed two independent times with similar results.