INITIAL CHARACTERIZATION OF CXXC5 AS A PUTATIVE DNA BINDING PROTEIN

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

BY

PELİN YAŞAR

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

JUNE 2015

Approval of the Thesis:

INITIAL CHARACTERIZATION OF CXXC5 AS A PUTATIVE DNA BINDING PROTEIN

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

Prof. Dr. M. Gülbin Dural Ünver
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 _____

Assoc. Prof. Dr. Çağdaş D. Son
Biology Dept., METU _____

Assist. Prof. Dr. Bala Gür Dedeoğlu
Biotechnology Institute, Ankara University _____

Date: 09.06.2015

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: Pelin Yaşar

Signature :

ABSTRACT

INITIAL CHARACTERIZATION OF CXXC5 AS A PUTATIVE DNA BINDING PROTEIN

Yaşar, Pelin
M.S., Department of Biology
Supervisor: Assoc. Prof. Dr. Mesut Muyan
June 2015, 101 pages

17β-estradiol (E2), the main circulating estrogen hormone, is involved in the physiological and pathophysiological regulation of various tissue notably mammary tissue functions. E2 is responsible for the cellular proliferation, differentiation and/or death in target tissue. The E2 effect is mediated by the nuclear receptors, estrogen receptor α and β, as ligand-dependent transcription factors. Upon binding of E2, ER is converted to an active form and regulates the expression of target genes primarily through genomic signaling pathways leading to cellular responses. The binding of E2-ER to specific DNA sequences estrogen response elements (EREs) initiates the transcription through the ERE-dependent signaling pathway. The interaction of the E2-ER complex with the other transcription factors that are already bound to their cognate response elements mediates the transcriptional events via the ERE-independent signaling pathway. Previous genome-wide analysis studies of our laboratory suggested that the *CXXC5* gene expression is regulated by E2-ERα through the ERE-dependent signaling pathway. CXXC5 is a member of the ZF-CXXC domain protein family that contains a highly conserved CXXC domain and a nuclear localization signal. The ZF-CXXC family proteins, through their CXXC domain, bind to non-methylated CpG dinucleotides in CpG islands of transcriptionally active DNA regions. This binding prevents cytosine methylation and leads to the formation of a nucleation site for the direct or indirect recruitment of histone modifying proteins to DNA for transcription regulation. Although studies on functional features of CXXC5 are scarce, the protein appears to

participate as a transcription factor, an epigenetic regulator and/or a co-modulator in the transcriptional regulation of cellular events in response to various signaling pathways. We therefore predict that CXXC5, as a non-methylated CpG dinucleotide binding protein, also plays a fundamental role in E2-mediated cellular events.

CXXC5 is located on 5q31.2, oriented on the forward strand and encompasses 35-kb DNA. There are three exons and the parts of the second and third exons generate a 1447 bp long mRNA with 969 bp long open reading frame that encodes a 322 amino-acid long protein with a molecular mass of approximately 33 kDa.

To begin examining the structure-function of CXXC5, we initially verified our genome-wide findings that *CXXC5* is an estrogen responsive gene using MCF7 cells derived from a breast adenocarcinoma. We found by RT-qPCR that *CXXC5* is indeed an E2- and ERα-responsive gene. We subsequently cloned the ORF of the *CXXC5* transcript into a mammalian expression vector. We examined the synthesis and intracellular location of endogenous and exogenously introduced CXXC5 using western blot, immunocytochemistry and short interfering RNA approaches in cell models. We observed that CXXC5 is synthesized at varying amounts in cell lines of breast carcinomas. We also found that CXXC5 localized in the nucleus and shows a diffuse distribution in interphase appearing to overlap with DNA. However during cell division, CXXC5 displayed a distinct nuclear staining that does not overlap with DNA. This suggests that the intra-nuclear distribution of CXXC5 is cell-cycle dependent.

Based on our homology modeling and the presence of CXXC domain we predicted that CXXC5 as other protein members of the ZF-CXXC family is a non-methylated CpG dinucleotide binding protein. To verify this prediction, we analyzed the interactions of CXXC5 and CpG dinucleotide containing DNA fragments by electrophoretic mobility shift assay using whole cell extracts or recombinant protein obtained from a bacterial protein expression system. However, we did not observe any binding of CXXC5 to DNA. These findings suggest either that CXXC5, in contrast to our prediction, is not a DNA binding protein; or that bacterial expression

is not an appropriate system for CXXC5 to obtain a functional protein. We are currently addressing these issues using various eukaryotic expression systems.

Inclusion of studies on the identification of protein interaction partners of CXXC5 will be an important aspect of future studies aimed at the dissection of structural and functional features of the protein.

Keywords: estrogen, estrogen receptor, *CXXC5*, CXXC5

ÖZ

POTENSİYEL OLARAK DNA'YA BAĞLANAN CXXC5 PROTEİNİNİN ÖN KARAKTERİZASYONU

Yaşar, Pelin
Yüksek Lisans, Biyoloji Bölümü
Tez Yöneticisi: Doç. Dr. Mesut Muyan
Haziran 2015, 101 sayfa

17β-östradiyol (E2), dolaşımda bulunan ana östrojen hormonudur ve başta meme dokusu olmak üzere çeşitli dokuların fizyolojik ve patofizyolojik düzenlenmelerinden sorumludur. E2 hedef dokularında hücre çoğalması, farklılaşması ve ölümü olaylarında görev almaktadır. E2'nin etkileri, ligand-bağımlı nükleer transkripsiyon faktörleri olan östrojen reseptörü α ve β (ERα and β) tarafından gerçekleştirilmektedir. E2'nun bağlanması ile aktif hale gelen ER'ü, başta genomik sinyal yolakları üzerinden olmak üzere, hücresel cevapları oluşturacak şekilde, hedef genlerinin ekspresyonlarını düzenler. Oluşan E2-ER kompleksi, östrojen yanıt elemanlarına (estrogen response elements, ERE) doğrudan bağlanarak ERE-bağımlı sinyal yolağı üzerinden transkripsiyonu başlatır. ERE-bağımsız sinyal yolağında ise, E2-ER kompleksi, hali hazırda kendi ilgili yanıt elemanlarına bağlanmış halde bulunan başka transkripsiyon faktörlerine bağlanmak suretiyle genomik olaylara dahil olur. Laboratuvarımızda önceden yapılmış olan genom-düzeyindeki çalışmalarda, *CXXC5* geninin ekspresyonunun E2-ERα ile ERE-bağımlı sinyal yolağı üzerinden düzenlendiği öngörülmüştür. CXXC5 proteini ZF-CXXC domain protein ailesinin bir üyesidir, yüksek derece korunmuş olan CXXC domainini ve nükleer lokalizasyon sinyali içermektedir. ZF-CXXC protein ailesi üyeleri, içerdikleri CXXC domain sayesinde transkripsiyonel olarak aktif DNA bölgelerinde bulunan CpG adalarının metile-olmamış CpG dinükleotitlerine bağlanabilmektedirler. Bu bağlanma, sitozinlerin metillenmesini engelleyerek, histonları etkileyen proteinlerin doğrudan ya da dolaylı olarak DNA

bölgesine çağırılmasını ve böylece transkripsiyonun düzenlenmesini sağlamaktadır. CXXC5'in fonksiyonel özellikleri üzerine az sayıda çalışma olmasına rağmen, CXXC5'in çeşitli sinyal yolaklarına cevaben transkripsiyonu düzenlediği ve bunu bir transkripsiyon faktörü, epi-genetik düzenleyici ya da eş-düzenleyici olarak rol almak suretiyle gerçekleştirdiği düşünülmektedir. Bu nedenlerle, CXXC5'in metile-olmamış CpG dinükleotitlerine bağlanan bir protein olarak, E2 üzerinden oluşturulan hücresel olaylarda temel bir rol aldığını öneriyoruz.

CXXC5 kromozomda 5q31.2 konumunda, anlamlı zincir üzerinde 35-kb'lık bir DNA bölgesini kapsamaktadır. Toplamda üç eksonu bulunan *CXXC5*'in, 1447 bç uzunluğundaki mRNA'sı ikinci ve üçüncü eksonun bazı kısımlarından oluşturulmaktadır. 969 bç uzunluğundaki açık okuma çerçevesinden (open reading frame, ORF) 322 amino-asit uzunluğunda, yaklaşık olarak 33 kDa moleküler ağırlığında bir protein sentezlenmektedir.

CXXC5'in yapısal ve fonksiyonel özelliklerini incelemek üzere, öncelikle, meme adenokarsinomasından türetilmiş olan ER-pozitif MCF7 hücreleri kullanılarak, *CXXC5* ile ilgili genom-düzeyinde yapılmış olan çalışmada bulunan sonuçları doğruladık. Gerçek-zamanlı kantitatif PZR (Real-time quantitative PCR, RT-qPCR) kullanılarak, *CXXC5* geninin E2- ve ERα-yanıtlı olduğu gösterildi. Ardından, *CXXC5* transkriptinin açık okuma çerçevesi, memeli ekspresyon vektörüne klonlandı. İç-kaynaklı ve dış-kaynaklı CXXC5'in sentezi ve hücre-içi konumu western blot, immünsitokimya ve siRNA (short interfering RNA) yöntemleri ile incelendi. Yapılan hücre hattı taramasında, CXXC5'in taranan kanser meme hücrelerinde farklı seviyelerde sentezlendiği görüldü. Buna ek olarak CXXC5'in çekirdekte konumlandığı gözlemlenmiştir. CXXC5 proteini interfaz hücrelerinde DNA ile karışık/birlikte bir dağılım gösterirken; bölünmenin diğer evrenlerinde olan (metafaz, geç-anafaz/telofaz ve telofaz) hücrelerde ise DNA ile örtüşmeyen bir konumlanma göstermektedir. Yapılan bu gözlem, CXXC5'in çekirdek-içi konumlanmasının hücre-döngüsüne bağımlı olduğunu önermektedir.

Korunmuş olan CXXC domainini içeren CXXC5 için yapmış olduğumuz homolojik modelleme çalışmalarına dayanarak, CXXC5'in de diğer ZF-CXXC

protein ailesi üyeleri gibi metile-olmamış CpG dinükleotitlerine bağlandığını öneriyoruz. Bu öneriyi doğrulamak amacı ile meme kanser hücre ekstratı ve bakteriden elde edilmiş rekombinant CXXC5 proteini, CpG dinükleotiti içeren DNA parçaları ile EMSA deneylerinde kullanılmıştır. Ancak CXXC5'in DNA'ya bağlandığı gösterilememiştir. CXXC5'in DNA'ya bağlandığının gösterilememesi iki sonucu önermektedir: 1) CXXC5'in DNA'ya bağlanabildiği yönündeki öngörülerimizin aksine, CXXC5 DNA'ya bağlanan bir protein değildir. 2) Ya da ullanmakta olduğumuz bakteriyel ekspresyon sistemi, fonksiyonel CXXC5 proteinin elde edilmesi için uygun değildir. Bu nedenle, işlevsel bir proteinin elde edilebilmesi için ökaryotik ekspresyon sistemlerinin uygulamaya geçirilmesi için çalışmalar devam etmektedir.

Yapılacak olan CXXC5'in protein etkileşim partnerlerinin tanımlanması ile ilgili çalışmalar ise CXXC5'in yapısal ve fonksiyonel özelliklerinin aydınlatılabilmesi için önemli olacaktır.

Anahtar kelimeler: östrojen, östrojen reseptörü, *CXXC5,* CXXC5

To the beauty of the simplicity

ACKNOWLEDGEMENTS

First of all I would like to thank my supervisor Assoc. Prof. Dr. Mesut Muyan for his endless patience and support through my initial steps into scientific community. The things that I've learned under his scientific vision took me forward and provided a self-consistent and spirited resolution in me for present and also for the future. I am heartily thankful for his guidance and efforts for all the times.

I wanted to express my sincere gratitude to Assoc. Prof. Dr. A. Elif Erson-Bensan. She considered me as one of her own students and helped this study possible to be completed in time with her generous supports. I also wanted to thank Assoc. Prof. Dr. Çağdaş D. Son for his guidance in fluorescence microscopy and Assoc. Prof. Dr. Can Özen for cell culture facilities.

I owe a great many thanks to my lab mates for their kindness and friendship. I am very grateful to Gizem Güpür, best lab partner ever; her supports in all manners were golden. I am lucky to know Gamze Ayaz, her cheerful and energetic mood were the best. I want to thank Sırma Damla User for her kind helps and friendship. I also express my sincere feelings to Hamit İzgi, Batuhan Çağrı Yapan, Ece Çağla Olgun, Burcu Karakaya.

I want to thank Dr. Begüm Akman in particular. Her help, guidance and friendship were invaluable. I wanted to thank Cansaran Saygılı, Nihan Özdemirler, Merve Öyken, Hasan Hüseyin Kazan for the support and also, Tuna Çınkıllı, Esra Yavuz, Gülten Tuncel and Dr. Hızlan Hıncal Ağuş. I thank to Banerjee Lab members for their kindness and help. I also wanted to thank to Menekşe Ermiş for her help with confocal microscopy.

I cannot thank to Fatma Gözde Çilingir enough, your presence makes me stronger. Safiye Özkan, the friendship that we have means a lot to me. It is a fortunate for me to know Gizem Dinç, Hatice Sultan Yıldız, Seval Mutlu and Süleyman Özcan.

Finally, I would like to thank my precious family. I could not be the person who I am today without you. “You are worthy forever and ever.”

I would like to thank TÜBİTAK for supporting this study through 212T031 and 114Z243 projects.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES

LIST OF TABLES

TABLES

CHAPTER 1

INTRODUCTION

1.1 Estrogen signaling

17β-estradiol (E2) is the main circulating estrogen hormone in the body and involved in the physiological and pathophysiological regulation of various tissue notably breast tissue functions [1]. Breast cancer is a complex disease with distinct subtypes [2]. Genetic background, hormonal and environmental factors contributes to the occurrence of the breast cancer. However, E2 signaling is one primary factor for the uncontrolled growth and division of breast epithelial cells [3], [4].

E2 effect is mediated primarily by the transcription factors, estrogen receptors α and β (ERα and β). ERα and β are encoded by two distinct genes, *ESR1* and *ESR2,* but share a high degree of structural and functional characteristics. ERs are expressed in the same as well as different tissues at varying levels [1], [3]. ERα is the primary estrogen receptor synthesized in the breast tissue.

ERs are modular proteins and contain six functionally semi-independent domains [3]. A to F domains, named from N-terminal to C-terminal, are as follows in Figure 1.

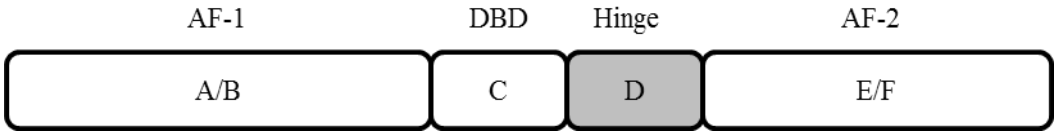


Figure 1: Six functional domains of ER are represented from N-terminal to C-terminal

The amino terminal A/B domain contains a ligand-independent transactivation (AF-1) function. The C domain is the DNA binding domain (DBD domain). The D domain is the hinge domain that provides flexibility between the amino- and

carboxyl-termini and also contains a nuclear localization signal. The E/F domain, also known as the ligand binding domain (LBD), is multi-functional that encompasses ligand binding, ligand-dependent transactivation function (AF-2) and dimerization functions [3]. Although small fractions of the newly synthesized ER dimerize and translocate to the perimembrane, cytoplasm, and mitochondria, the majority of the dimer ERα translocates to the nucleus independent from the presence of E2 [5]. Binding of E2 results in major conformational changes in the receptor. These structural changes transform the protein from an inactive form to an active one by generating protein interaction surfaces critical for the recruitment of the co-regulatory proteins and the basal transcription complex [6], [7].

E2-ER mediated genomic events occur through two distinct pathways: Estrogen response element (ERE)-dependent and ERE-independent. EREs are derivatives of a palindromic DNA sequence, 5' - GGTCAnnnTGACC - 3', separated by three non-specific nucleotides [6]. Studies showed that ERα even in the absence of E2 can interact with ERE dynamically [7]. Binding of E2 increases the association of E2-ER to ERE reflected as a decrease in the nuclear movement of ERα [7]. In the ERE-independent signaling pathway, E2-ER complex does not interact with the ERE sequences. Although the underlying mechanism is not fully clear, the regulation of target gene expressions by E2-ER through interactions with other transcription factors, such as activator protein 1 (AP-1) and stimulatory protein 1 (SP-1), bound to their cognate response elements denotes the ERE-independent pathway [5]. In addition to the nuclear events, the cytoplasmic ERα also responds to E2 by generating cytoplasmic signaling cascade or interacting with cascade proteins that are activated by other signaling pathways [8].

Our previous studies using genome-wide gene expression analyses suggest that *CXXC5* is one of the E2 responsive genes. E2-ERα mediates regulation of the expression of *CXXC5* through the ERE-dependent signaling pathway [5].

1.2. CXXC-Type Zinc Finger Protein 5 (CXXC5)

Studies in a hematopoietic stem and progenitor cell line CD34+ led to the cloning of *CXXC5* in 2000 [9]. *CXXC5* is located on the 5q31.2, oriented on the forward strand and encompasses 35-kb DNA. There are three exons and the parts of the second and third exons generate a 1447 bp long mRNA with 969 bp long open reading frame encoding a 322 amino-acid protein with a molecular mass of approximately 33 kDa (Figure 2). Chromosomal location of *CXXC5* is given in Figure 3.

Figure 2: Shaded parts of the exons represent the coding regions

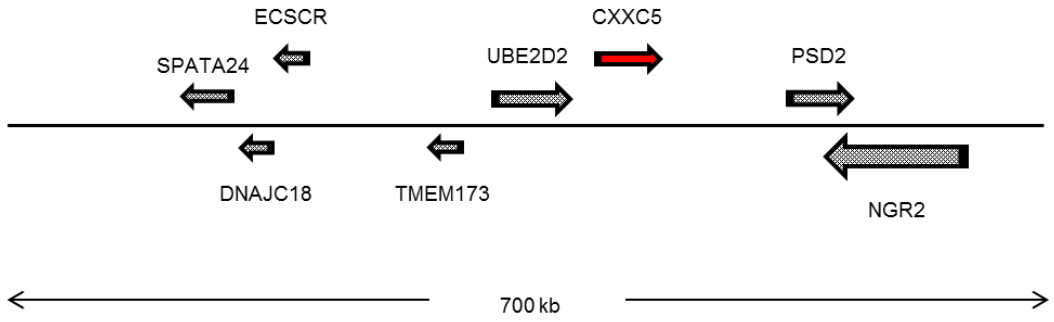


Figure 3: Depiction of the local order of *CXXC5* together with transcriptional directions of the leading and lagging genes on chromosome 5

The CXXC5 protein, also referred to as the CXXC Finger Protein 5 (CF5); Retinoid-Inducible Nuclear Factor (RINF); WT1-Induced Inhibitor of Dishevelled (WID), is considered to be a member of the ZF-CXXC domain protein family. As the other members, CXXC5 contains a conserved zinc-finger CXXC domain. The

vertebrate orthologs of human CXXC5 display high amino-acid identity varying between 67-81%. The important paralog of *CXXC5* is the *CXXC4*, or *IDAX*, gene.

1.2.1. The CXXC Domain Containing Proteins

The mixed lineage leukemia (MLL) protein is a common target of chromosomal translocations found in leukemia. Translocations of *MLL* could lead to the formation of a chimeric protein with an amino-terminal fragment of MLL fused to another protein [10]. Studies have identified 64 fusion protein partners for MLL [11]. MLL contains a putative DNA binding domain structured by two zinc ions each of which coordinate four conserved cysteines within a CGXCXXC motif with which MLL binds to non-methylated CpG dinucleotides [12]. Subsequent studies indicated that the CXXC domain of MLL binds in a wedge-like manner into the major groove of CpG dinucleotide containing DNA mediated by the electrostatic interactions, hydrogen bonds and several water mediated interactions [13]. The CXXC domain of MLL, as also of Cfp1 protein [13], interacts with a short six base pair long sequence of the DNA centered on the CpG dinucleotide motif [14]. Changing residues in the MLL-CXXC domain critical for DNA interactions disrupt the ability of the protein to bind to DNA. Moreover, the exogenous expression of this mutant MLL-fusion protein into cancer cell models leads to an increased DNA methylation of CpG nucleotides in various loci, decreased gene expressions, immortalization potential and inability to induce leukemia in mice [14]. Based on these observations, it was suggested that the prevention of DNA methylation by binding of MLL-fusion proteins through the CXXC domain to DNA and consequently protecting against DNA methylation is essential for the development of MLL fusion leukemia [14].

Structural and functional homology studies using the MLL protein as the template led to the discovery of the Zinc finger (ZF)-CXXC domain family proteins that specifically recognize and bind to non-methylated CpG containing DNA. This binding prevents DNA methylation. ZF-CXXC domain family proteins are involved in the gene expressions by interacting with proteins involved in the post-

translational modification of histone proteins [15]. ZF-CXXC protein family members include, and listed in Table 1 [15], CFP1 (CXXC protein finger 1, CXXC1), MBD1 (Methyl-CpG Binding Domain Protein 1, CXXC3), KDM2A & B (Lysine (K)-Specific Demethylase 2A & 2B; CXXC8 & CXXC2), DNMT1 (DNA (Cytosine-5-)-Methyltransferase 1, CXXC9) and TET1, 2, and 3 (Tet Methylcytosine Dioxygenase 1, 2, 3) [15].

Table 1: The ZF-CXXC family proteins (taken from Long 2013)

Gene name	CXXC nomenclature	Other names	NCBI Gene ID
KDM2A	CXXC8	Fbxl11, Jhdm1a, Ndy2	225876
KDM2B	CXXC2	Fbxl10, Jhdm1b, Ndy1	30841
FBXL19	–	–	233902
CFP1	CXXC1	Cgbp, Phf18	74322
DNMTL	CXXC9	Met1	13433
MLL1	CXXC7	All1, Htrx1, Kmt2a	214162
MLL2	–	Wbp7, Kmt2b	75410
MBD1	CXXC3	Pcm1	17190
TET1	CXXC6	Lcx	52463
TET3	CXXC10	–	194388
IDAX	CXXC4	–	319478
CXXC5	CXXC5	–	67393

The TET protein family member TET3 (*Xenopus tropicalis*) is involved in the hydroxylation of the 5-methylcytosine leading to the epigenetic reprogramming [16]. The crystal structure of the CXXC domain together with various methylated or non-methylated CpG dsDNAs revealed that although TET3-CXXC binds preferentially to non-methylated CpGs, the binding to hemi-methylated or methylated CpGs can still occur at lower affinities [17]. These findings suggest that the binding preference of CXXC bearing protein family to methylated and/or non-methylated CpG dinucleotides may yet be incomplete and needs to be characterized for each member.

The ability of the CXXC domain containing proteins to bind to CpG dinucleotides of CpG islands is suggested to be crucial for the recognition of target gene,

transactivation and myeloid differentiation [12]. Cytosine methylation defines a major epigenetic event of DNA modifications and mostly occurs on CpG dinucleotides that can lead to the transcription inactivation or stable silencing of gene expressions [12]. Non-methylated CpG islands are predominantly found in the transcriptionally active regions [18] and the promoter regions of the actively expressed genes are hypomethylated [19].

1.2.2. The CXXC Domain

According to homology studies, due to the presence of a carboxyl-terminally located CXXC domain (zinc-finger, ZF-CXXC domain) (Figure 4), CXXC5 is suggested to be a non-methylated CpG binding protein [20]. In the light of the alignment of the CXXC domain containing proteins, the consensus sequence for CXXC domain is defined as the $C\text{-}X_2\text{-}C\text{-}X_2\text{-}C\text{-}X_{4\text{-}5}\text{-}C\text{-}X_2\text{-}C\text{-}X_2\text{-}C_{9\text{-}15}\text{-}C\text{-}X_4\text{-}C$, which is also present in CXXC5 [21]. CXXC5 also contains a nuclear localization signal juxtaposed to the CXXC domain [21].

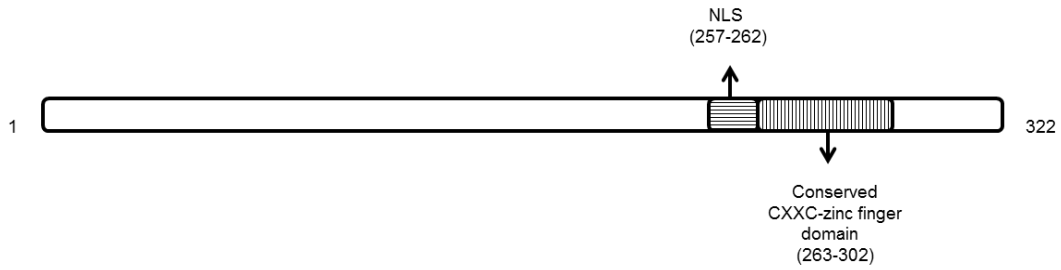


Figure 4: Schematic representation of CXXC5 in which the nuclear localization signal is adjacent to the CXXC domain

The CXXC domains of IDAX (CXXC4) and CXXC5 share a high degree of amino-acid homology. Although the structure of IDAX is yet unknown, it appears that IDAX modulates the protein levels of *TET2* (Ten-eleven translocation methylcytosine dioxygenase 2) by regulating *TET2* expression [22]. TET family proteins TET1, 2 and 3, are also members of ZF-CXXC domain protein family. Most animal species have only one TET orthologue, in jawed vertebrates there are

three paralogous of TET, likely evolved from a genomic triplication [22]. While the CXXC domain is present in *TET1 (CXXC6)* and *TET3 (CXXC10)*, *TET2* seemed to have lost the CXXC domain due to a chromosomal inversion split leading to the transcription of the catalytic and the CXXC domain of the TET2 protein from two distinct segments and in the opposite direction on the same chromosome [23]. IDAX appears to be the ancestral CXXC domain of TET2 [22]. It seems that IDAX is capable of down-regulating TET2 protein levels and that the CXXC domain of IDAX is critical for the decrease in TET2 levels [22]. It was shown that TET2 levels are also decreased in the presence of CXXC5 [22]. This suggests that CXXC5 could also be critical in the transcriptional regulation of *TET2* in physiology and pathophysiology of breast tissue.

Studies on the structural basis of the binding of IDAX-CXXC domain, which is also conserved in CXXC5, indicate that IDAX binds to various non-methylated CpG dinucleotides surrounded with varying non-specific nucleotides. Moreover, TGHQ amino-acid residues in the IDAX-CXXC domain appears to be responsible for the binding of the domain to DNA [13].

1.2.3. Studies on *CXXC5* and Its Protein Product

There are limited studies on *CXXC5* and its protein product. Consequently, the function of CXXC5 is yet unclear. However, the available studies suggest that CXXC5 is involved in cellular proliferation, metabolism, differentiation and death. These studies are summarized below.

CXXC5 is expressed in various tissues at various levels. CXXC5 protein is localized mainly in the nucleus; however, it can be co-localized in the cytoplasm by binding to Dishevelled proteins (Dvl) [24], [25]. Dvl proteins are cytoplasmic scaffold proteins that mediate Wnt/β-Catenin signal transduction. It was observed that CXXC5 binds to intermediary proteins in the Wnt/β-Catenin signaling through its CXXC domain [24], [25]. Dvl proteins serve a connection between the Frizzled WNT receptor and the downstream molecules of the Wnt/β-Catenin signaling pathway [24]. It is suggested that CXXC5 may play a role as a negative regulator by interacting with Dvl proteins in the Wnt/β-Catenin signaling pathway which

contributes to the regulation of stem cell pluripotency, cell proliferation and differentiation as well as the initiation and progression of the malignancies of kidney tissue [24], [25]. The transcription of *CXXC5* is activated by *WT1* which leads to the inhibition of the Wnt/β-Catenin signaling pathway [25]. Both *WT1* and Wnt/β-Catenin signaling pathways are required for the development of the mammalian kidney tissue [25] and deregulations of those pathways leads to tumor progression. By participating in the inhibition of Wnt/ β-Catenin signaling, it was thus suggested that CXXC5 is a tumor suppressor [25].

CXXC5 is also reported to be associated with cytokine-induced myelopoiesis of hematopoietic stem cells (normal CD34+ progenitors) [21]. *CXXC5* is a retinoic acid responsive gene [26] and it is also involved in the differentiation of leukemic cells in response to retinoic acid [21]. These findings suggest that the expression of *CXXC5* is important for the myelopoiesis of normal and malignant hematopoietic cells.

CXXC5 expression appears to also be important for the differentiation of endothelial cells. CXXC5 was suggested to have a role as a transcription factor in the expression of *Flk-1* (Vascular endothelial growth factor receptor, Flk-1/KDR) of the BMP4 (Bone morphogenic protein 4) signaling pathway [27]. By regulating the *Flk-1* expression, CXXC5 aids to vessel formation and angiogenesis [27].

CXXC5 was reported to be involved in the DNA-damage induced p53 activation. The transcription factor p53 has an anti-proliferative role in cells. p53 acts as a tumor suppressor protein by causing cell cycle arrest, activating DNA repair mechanisms, cellular senescence and/or apoptosis [28]. Anomalies in the DNA structure such as nicks and double strand breaks (DSBs) are major factors that trigger the DNA damage response signaling pathway (DDR) [29]. The alarm generated by DSBs is conveyed primarily with ataxia telangiectasia (A-T) mutated (ATM), which is one of the phosphoinositide 3-kinase-like protein kinases [30], [31]. ATM phosphorylates downstream target proteins that include p53. This p53 phosphorylation leads to the activation of repair mechanisms, checkpoint arrest and

apoptosis [28]. Studies suggest that CXXC5 through the CXXC domain interacts with ATM and participates in the phosphorylation of p53 [28].

Wang and colleagues suggested that CXXC5 may have a role in tumor necrosis factor α (TNF-α) induced apoptosis [32]. Apoptosis is a programmed cell death that can be mediated by two different pathways: intrinsic and extrinsic. In the intrinsic pathway, apoptosis is initiated from the mitochondria in response to DNA-damage [33] and p53 is also involved [28]. In the extrinsic pathway, on the other hand, apoptosis is induced through death-receptors located on the cell surface by pro-apoptotic ligands [34]. TNF-α, which is a cytokine, is known to activate the extrinsic pathway [32]. CXXC5 and Smad3 interaction is proposed to induce the initiation of apoptosis through TNF-α [32]. In human, eight Smad proteins are present and five of them (Smad1,2,3,5 and 8), named as receptor-regulated Smads (RSmads), serve as substrates for TGF-β family receptors [35], [36]. The nuclear accumulation of the RSmads-Smad4 complex is mediated by phosphorylation events [37]. The phosphorylated RSmad-Smad4 complex contributes gene transcriptions by interacting with other cofactors [36]. CXXC5 enhances Smad3 phosphorylation and the translocation of the SMAD complex by binding to Smad2, 3 and 4 [32]. TNF-α activity, which is promoted by the RSmad-Smad4 complex, was shown to be induced by CXXC5 and Smad3 interactions [32].

In a yeast one-hybrid study, it was suggested that CXXC5 is an Oxygen Responsive Element (ORE) binding protein. It was reported that CXXC5 can bind to an ORE in the promoter region of *COX4I2* by DNA binding assay [38]. In cells, the oxidative phosphorylation leads to the energy production by transferring of electrons through several electron acceptors. The ultimate electron acceptor is O_2, and the cytochrome c enzyme (COX) catalyzes H_2O production from O_2 [39]. COX has both nuclear and mitochondrial subunits and the activity of the enzyme is generated by COX subunit assembly. The first step of COX subunit assembly is the association of COX1 and COX4 [39]. The COX activity is regulated through matrix/cytosolic ATP/ADP ratios and also interactions with COX4 [39]. COX4 have two isoforms, COX4I1 (tissue-specific, lung) and COX4I2. According to studies, CXXC5 binds to the COX subunit *COX4I2* promoter at high levels of oxygen and represses the

expression of *COX4I2*. However, at low concentrations of oxygen, CXXC5 dissociates from the promoter and the synthesis of COX4I2 increases [38]. Based on these findings, CXXC5 is suggested to be an important transcription factor for the adaptation of cells against drastic changes in oxygen levels.

Collectively, these aforementioned studies suggest that CXXC5 as both nuclear and cytoplasmic protein is involved in cell proliferation, differentiation, metabolism and death as a transcription factor and/or an interacting protein partner.

1.2.4. *CXXC5* in Disease State

The expression of *CXXC5* appears to be de-regulated in solid tumors and leukemia. De-regulation of the *CXXC5* levels in different cancer types suggests that the *CXXC5* expression may have an important role in the initiation and progression of the pathology.

1.2.4.1. Acute Myeloid Leukemia and Myelodysplasia Syndrome

The 5q31.2 chromosomal location of *CXXC5*, found 20 kb downstream of the distal telomeric marker region (D5S594), also harbors several tumor suppressor genes [21], [40]. This locus is a marker for patients with myelodysplasia syndrome (MDS) and acute myeloid leukemia (AML), [41], [42]. AML is a disease caused by the cytogenetic abnormalities that lead to aberrant cell proliferation, death and differentiation [10]. MDS patients suffer from an insufficient hematopoiesis, caused by the trilineage dysplasia [43]. The chromosome locus harboring also *CXXC5* is often deleted in MDS and AML patients [21]. It is therefore suggested that either the deletion of promoter regulatory regions of *CXXC5* or the gene itself contributes to disease state [21].

1.2.4.2. Acute Promyelocytic Leukemia

The expression of *CXXC5* is required for the terminal maturation of premyelocytic leukemia cells [21] leading to acute promyelocytic leukemia (APL). APL is a rare subtype of AML and it is characterized by the translocation of the retinoic acid receptor alpha gene [44].

1.2.4.3. Breast Cancer

Breast cancer is mainly originated from cells that line the ducts and/or the lobules of the breast. High expression levels of *CXXC5* is observed in advanced breast tumors, that also express ERα [20].

1.3. Aim of the Current Project

E2 exerts its proliferative effect in breast tissue primarily through the transcription factor ERα. The E2-ERα complex mediates gene expressions via ERE-dependent or ERE-independent signaling pathways. In a microarray study performed in our laboratory, the expression of *CXXC5* was found to be enhanced by E2-ERα through the ERE-dependent signaling pathway. According to the homology modeling, CXXC5 is a member of the ZF-CXXC domain family proteins that bind to the non-methylated CpG islands. Based on the limited number of studies, CXXC5 is suggested to be involved in the regulation of cellular proliferation, differentiation and death as a transcription factor and/or interacting protein partner.

The aim of my thesis project is to characterize the expression of *CXXC5* and the synthesis of its protein product in model cells derived from breast cancer to provide a foundation for further studies on the structural and functional features of CXXC5.

CHAPTER 2

MATERIALS AND METHODS

2.1. Bioinformatics Tools

The genomic DNA (gDNA) nucleotide, messenger RNA (mRNA) nucleotide for *CXXC5* and the protein sequence for CXXC5 were obtained from Genome Browser database at the University of California Santa Cruz with the purpose of performing bioinformatics analysis. To analyze the charge status of the CXXC5, we used the ProtParam web tool [45].

2.1.1. Family/Domain/Motif Analysis

The pfam [46] and ELM [47] bioinformatics tools were used for the analysis of motifs and functional domains of CXXC5.

2.1.2. Prediction of Post-Translational Modifications

To analyze the putative post-translational modifications of CXXC5, we used the NetPhos [48] and NetAcet web tools [49].

2.1.3. Prediction of Potential ERE Sequences

The dragon ERE finder version 3 (System for Identification and Interactive Analyses of Estrogen Response Elements in DNA Sequences) [50] was used for the prediction of potential ERE sequences in the *CXXC5* gene locus.

2.1.4. Homology Modeling

The Phyre2 program was used to predict the protein structure of CXXC5 [51]. For homology modeling studies the Modeller [52] and Chimera 1.8.1 [53] programs were used.

2.2. Cell lines and treatments

2.2.1. Cell lines and maintenance

MCF7 cell line was kind gift of Dr. A. Elif Erson-Bensan (Middle East Technical University, Ankara, Turkey). MCF7 cells were grown in phenol-red free, high glucose (4.5g/L) containing Dulbecco's modified Eagle's medium (DMEM, Lonza, Belgium, BE12-917F) supplemented with 8% fetal bovine serum (FBS, Biochrom AG, Germany S0115), 0.5% L-Glutamine (Lonza, Belgium, BE17-605E), and 1% Penicillin/Streptomycin (P/S, Lonza, Belgium, DE17-602E).

2.2.2. Hormone Treatments

To determine estrogen effects on the expression of *CXXC5*, we used 17β-estradiol (E2) (Sigma-Aldrich, St. Louis, MO, USA). To test whether the *CXXC5* expression is ER-dependent, we utilized a complete ERα antagonist, Imperial Chemical Industries 182,780 (ICI, Tocris Biosciences, MN, USA).

To test the effect of E2 on *CXXC5* expression, we used ERα-positive MCF7 cells. The hormone treatment experiments were carried out with phenol-red free DMEM supplemented with 8% charcoal-dextran stripped FBS, CD-FBS, (Appendix D), 0.5% L-Glutamine, 1% Penicillin/Streptomycin. Cells were grown for 48 hours in steroid hormone free medium. Cells were then treated with a physiological level, 10^{-9} M (1 nM), of E2 or % 0.01 ethanol (as vehicle control) for 3 and 24 hours. In addition, MCF7 cells were treated with 100 nM ICI in the absence or presence of 1 nM E2 for 24h.

2.3. Expression Analysis

RNA isolation, quantification, cDNA synthesis and RT-qPCR analysis were performed according to MIQE Guidelines and checklist (Appendix F).

2.3.1. Total RNA Isolation, Quantification and Integrity Assessment

For total RNA isolation, cells grown in CD-FBS medium for 48h were subjected to hormone treatment. Cells were then collected by trypsinization, washed with PBS for two times and pelleted. Pellets were stored at -80^{o}C until total RNA isolation.

Total RNA was isolated by the Quick-RNA Miniprep Kit (ZymoResearch, Irvine, CA, USA) with on-column DNaseI treatment. Manufacturer's instructions were followed during the RNA isolations.

NanoDrop 2000 (Thermo Scientific, CA, USA) was used for the assessment of RNA concentration and purity. The evaluation of RNA sample purity was determined at A260/280 and A260/230 ratios.

2.3.2. DNA Contamination Control in Total RNA Samples by PCR

To examine potential genomic DNA contaminations in total RNA samples, we used gene specific primers for *GAPDH* (Glyceraldehyde 3-phosphate dehydrogenase). These primers were: GAPDH_FP: 5'-GGGAGCCAAAAGGGTCATCA-3' and GAPDH_REP: 5'-TTTCTAGACGGCAGGTCAGGT-3' (product size: 409 bp). PCR reaction was performed as described in Table 2.

Table 2: PCR reaction conditions

	^{o}C	Duration	Number of Cycles
Initial Denaturation	95	3 min.	1
Denaturation Annealing Extension	95 65 72	30 sec. 30 sec. 1 min.	40
Final Extension	72	10 min.	1
Hold	+4	∞	1

2.3.3. cDNA Synthesis

The RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific, CA, USA) was used to synthesize cDNA (20 µl) from total RNA samples. 500 ng total RNA and oligo-dT primers were used in PCR reactions, which were assembled as described in Table 3. Reactions were carried out at 42^{o}C for one hour and terminated after five-minute incubation at 70^{o}C. Samples were then stored at -80 ^{o}C.

Table 3: Reaction conditions of cDNA synthesis

Components	Amount
Total RNA	500 ng
5X Reaction Buffer (250 mM Tris-HCl pH 8.3, 250 mM KCl, 20 mM $MgCl_2$, 50 mM DTT)	4 µl
Oligo$(dT)_{18}$ Primers (100µM 0.5 µg/µl 15 A260 u/ml)	1 µl
10mM dNTP Mix	2 µl
RiboLock RNase Inhibitor (20U)	1 µl
RevertAid M-MulVReverse Transcriptase	1 µl
Nuclease-free Water	Up to 20 µl

2.3.4. *pS2/TFF1* Expression Analysis

TFF1 (trefoil factor 1, *pS2*) is a well-documented estrogen responsive gene regulated by E2-ERα through the ERE-dependent signaling pathway. We therefore used the expression of *TFF1* for E2 and ER responsiveness as control. *TFF1* specific primers pS2/TFF1_FP: 5'-TTGTGGTTTTCCTGGTGTCA-3' and pS2/TFF1_REP: 5'- CCGAGCTCTGGGACTAATCA-3' (product size: 209 bp) were used in RT-qPCR experiments. RT-qPCR was carried out as described in Table 4.

Table 4: RT-qPCR conditions

	°C	Duration	Number of Cycles
Initial Denaturation	94	10 min.	1
Denaturation	94	30 sec.	40
Annealing	55	30 sec.	
Extension	72	30 sec.	
Melting	50-99	-	1

2.3.5. RT-qPCR Analysis of *CXXC5* Expression

The SYBR® Green Mastermix (Roche Applied Science, Switzerland) was used for RT-qPCR in 20 µl reactions using the Rotor Gene 6000 PCR system (Corbett, Qiagen, Germany). We used 300 nM cDNA specific *CXXC5* primers during the RT-qPCR. These primers were: CXXC5_FP: 5'-CGGTGGACAAAAGCA

ACCCTAC-3' and CXXC5_REP: 5'- CGCTTCAGCATCTCTGTGGACT-3' that produce a PCR product in the size of 145 bp.

RPLP0, Ribosomal protein, large, P0, and the homolog of Pumilio, Drosophila, 1, *PUM1*, expressions were shown to be effectively used for the normalization of qPCR results in breast carcinoma samples [54]. Primers specific for *RPLP0* were: RPLP0_FP: 5'- GGAGAAACTGCTGCCTCATA -3' and RPLP0_REP: 5'- GGAAAAAGGAGGTC TTCTCG -3'. *PUM1* specific primers were: PUM1_FP: 5'- AGTGGGGGACTAGGCGTTAG -3' and PUM1_REP: 5'- GTTTTCATCACTGTCTGCAT CC -3'. For PCR, the initial denaturation was at 94^{o}C for 10 minutes followed by 40 repeats of 94^{o}C for 30 seconds, 60^{o}C for 30 seconds, and 72^{o}C for 30 seconds cycles.

The relative quantification of reaction efficiency was assessed by ΔΔCq. All RT-qPCR reactions were carried out with three independent biological replicates in triplicates. EtOH treated samples were set to one (1) for each time point, and E2 and ICI treated samples were compared to EtOH samples at every time point. The results were analyzed and graphed with the GraphPad Prism version 6 for Windows (GraphPad Software, San Diego California USA) using one-tailed, paired test with 95% confidence interval.

MIQE guideline was followed in RT-qPCRs (Appendix F). Analyses of melt curves were performed for each reaction.

2.4. Cloning of the Open Reading Frame of *CXXC5*

pMigR1 that bears the open reading frame of *CXXC5* was kindly provided by Dr. F. Pendino, INSERM UMRS-1007, Paris, France. To ease the cloning of this *CXXC5*-ORF into various expression vectors, we re-cloned the ORF using primers: FP_XNmetN_CXXC5: 5'- CGCATATACTCGAGATTACCATGGAAGCTAGC ATGTCAGCCTCGGCGGTGGCTC -3' and REP_EpolyaB_CXXC5: 5'- CGC ATGGGATCCTTTATTAGAATTCCTGAAACCACCGGAAGGCGG -3' that contain new restriction enzyme sites. The generated new ORF was inserted into pBSKS(-), pcDNA3.1(-), and pet28a-6xHis vectors.

2.4.1. Cloning Primers

Two cloning primers that harbor restriction enzyme sites were used for cloning of the *CXXC5*-ORF. The forward cloning primer contains XhoI, NcoI and NheI enzyme restriction sequence and represented red, light blue and purple respectively. The forward primer is FP_XNmetN_CXXC5: 5'- CGCATATACTCGAGATTACC ATGGAAGCTAGCATGTCAGCCTCGGCGGTGGCTC -3'. NcoI restriction site together with an initial A nucleotide serves as the Kozak sequence. On the other hand, the reverse cloning primer contains EcoRI, BamHI restriction sites and polyA signaling sequence in the middle, represented as green, blue and red respectively. This reverse primer is: REP_EpolyaB_CXXC5: 5'- CGCATGGGATCCTTTAT TAGAATTCCTGAAACCACCGGAAGGCGG -3'. The restriction enzyme sites were chosen for the purpose of generating different constructs using only two set of cloning primer. At the end both the wild type and N-terminally Flag-tagged (GATTACAAGGATGACGACGATAAG) constructs were obtained.

2.4.2. Cloning of the Wild-Type *CXXC5* into the pBS-KS Vector

For cloning of the ORF of *CXXC5*, PCR was performed by 500 nM cloning primers and 50 ng of pMigR1-CXXC5 as the template. Phusion High-Fidelity DNA Polymerase (Thermo Scientific, CA, USA) was used with GC buffer and PCR conditions are described in Table 5.

Table 5: Cloning PCR conditions

	°C	Duration	Number of Cycles
Initial Denaturation	98	2 min.	1
Denaturation Annealing/ Extension	98 72	10 sec. 30 sec.	40
Final Extension	72	10 min.	1
Hold	+4	∞	1

PCR product was subjected to PCR purification using the QIAquick PCR Purification Kit (Qiagen, Germany). The PCR product and pBS-KS(-) vector were double digested with restriction enzymes of XhoI (Thermo Scientific, CA, USA)

and BamHI (Thermo Scientific, CA, USA) in Buffer B (Roche Applied Science, Switzerland) for two hours at 37°C. To avoid self-circulation of the vector, FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific, CA, USA) treatment was performed for one hour at 37°C. After the double digestion, the PCR product and the vector were run on 1% agarose gel. The DNA fragments were recovered with the Zymoclean™ Gel DNA Recovery Kit (Zymo Research, CA, USA) and ligated using the Rapid DNA Ligation Kit (Thermo Scientific, CA, USA) for 15 minutes at room temperature. Ligation reactions were transformed into the transformation competent (chemically-compotent) XL-1 Blue cells (Agilent-Stratagene, CA, USA). 10 µl of the ligation mixture was added to competent cells (90 µl) and incubated on ice for one hour followed by a heat-shock for 45 seconds at 42°C. After heat-shock, cells were incubated on ice for two minutes and 450 µl LB was added onto the cells. Cells were then placed into the shaker for one hour at 37°C. Cells were centrifuged for five minutes at 3000 rpm. The most of supernatant was removed and the pellet was resuspended in the remaining liquid. Resuspended cells were inoculated on LB-Agar plates containing 100 µg/ml ampicillin, because of the ampicillin resistance gene present in the pBS-KS (-) vector. The next day, selected colonies were screened with colony PCR with cloning primers to confirm that the plasmid bearing an insert was taken by the colony. After confirmation of the insert presence, selected colonies were grown in 100 µg/ml ampicillin containing LB for overnight at 37°C. 1.5 ml bacterial culture was used for plasmid isolation by using the QIAprep Spin Miniprep Kit (Qiagen, Germany). Plasmids were sequenced at the METU Central Laboratories.

2.4.3. Cloning of the Flag-Tagged *CXXC5* into the pBS-KS Vector

Flag (DYKDDDDK) is a small peptide broadly used for the protein detection and purification (Sigma-Aldrich). We used a pBS-KS (-) Flag vector harboring the flag-tag sequence between NcoI and NheI restriction enzyme sites to obtain the Flag-*CXXC5* cDNA. pBS-KS (-)-*CXXC5* without Flag-tag were double digested with the restriction enzymes of NheI (Thermo Scientific, CA, USA) and BamHI (Thermo Scientific, CA, USA) in Tango Buffer (Thermo Scientific, CA, USA) for two hours at 37°C. Products were then subjected to cloning using FastAP treatment for the

vector. DNA gel recovery for DNA fragments, ligation, transformation, PCR control and plasmid isolation were performed as described in Section 2.4.2.

2.4.4. Cloning of the Wild-Type and Flag-Tagged *CXXC5* into the Mammalian Expression Vector pcDNA3.1 (-)

pBS-KS (-)-*CXXC5*-and pBS-KS (-)-Flag-*CXXC5* together with pcDNA3.1(-) vector were double digested with restriction enzymes, XhoI (Thermo Scientific, CA, USA) and BamHI (Thermo Scientific, CA, USA), in Buffer B (Thermo Scientific, CA, USA) for two hours at 37^{o}C and PCR column purified. Products were then subjected to cloning. DNA gel recovery for DNA fragments, ligation, transformation, PCR control and plasmid isolation were performed as described in Section 2.4.2.

2.4.5. Cloning of the Wild-Type *CXXC5* into the Bacterial Expression Vector pet28a-6xHis

To obtain a His-tagged CXXC5, the ORF of *CXXC5* was cloned into the pet28a-6xHis vector using cloning primers FP_BamHI_NcoI_Flag_NheI and REP_CXXC5_EcoRI_polyA_XhoI given in Table A.1. pcDNA3.1(-)-*CXXC5* vector (for insert) and pet28a-6xHis vector were double digested with the restriction enzymes of NheI (Thermo Scientific, CA, USA) and BamHI (Thermo Scientific, CA, USA) in Buffer 3.1 (New England Biolabs, NEB, MA, USA) for two hours at 37^{o}C. Products were then subjected to cloning following the FastAP treatment of the vector. DNA gel recovery for DNA fragments, ligation, transformation, PCR control and plasmid isolation were performed as described in Section 2.4.2. The pet28a(+) bacterial expression vector contains kanamycin resistant gene, allowing the selection to be carried out with 30 µg/ml kanamycin.

2.4.6. Cloning of Truncated and Mutant Forms of *CXXC5* cDNA

The CXXC5 protein harbors a sequence of threonine, glycine, histidine, and glutamine amino acids (TGHQ) at the carboxyl-terminus that are considered to be critical residues for DNA binding [13]. To ensure that this is indeed the case, we mutated TGHQ sequence to AAAA with an overlapping PCR. Mutations were introduced by two specific primers that replace TGHQ with alanine. The red color

indicates the alanine nucleotide sequences in the primers; CXXC5_Ala Mutant_FP: 5'- AATCGAAAGGCTGCCGCAGCGATTTGCAAATT CAGAAAATGTGAG -3' and CXXC5_Ala Mutant_REP: 5'- TTTGCAAATCGCTGCGGCAGC CTTTCGATTCCTACAACTGCTGCA -3'. For the full length mutant form of CXXC5, the cloning primers of *CXXC5* cDNA which was given in Section 2.4.2 and the mutation primers were used in overlapping PCR. Firstly, two independent PCR reactions were performed as the first round PCR. Conditions of overlapping PCR were given in Table 6. FP_XNmetN_CXXC5 cloning and Mutant_REP primers were used for one PCR reaction; Mutant_FP and REP_EpolyaB_CXXC5 cloning primers were used for another. The products of these PCRs were run on 1% agarose gel and recovered after the assessment of their sizes as described in Section 2.4.2. The purified PCR products (800 and 150 bp) at 1:5 ratios were used as the template for the second round of the overlapping PCR with the FP_XNmetN_CXXC5 and REP_EpolyaB_CXXC5 cloning primers respectively. The generated PCR product was run on a 1% agarose gel and the expected size of PCR product (950 bp) was recovered. The purified PCR product was double digested and cloned into both pBS-KS (-) and pcDNA3.1 (-) as described in Section 2.4.2 and 2.4.4. The C terminus of CXXC5 contains the CXXC domain which is associated with the CpG DNA binding ability of the protein. To ensure that the N terminus of CXXC5, which has no predictable functional domain, does or does not alter the ability of the CXXC domain to interact with DNA, we also cloned the sequences for the N terminus of CXXC5 alone into the pcDNA3.1(-) expression vector. The N-terminus of CXXC5 was obtained with FP_XNmetN_CXXC5 and CXXC5_N Term_REP: 5' CGCATGGGATCCTTTATTAGAATTCCATGCCGC AGCGTTTCCGCTT -3'. Wild-type C-terminus of CXXC5 was obtained with CXXC5_C Term_FP: 5'- CGCATATACTCGAGATTACCATGGAAGCTAGCA GCTCCGGCAAGAAGAAGCGG -3' and REP_EpolyaB_CXXC5. For the mutant_C-Terminus of CXXC5, the full length mutant *CXXC5* cDNA was used as the template together with CXXC5_C Term_FP and REP_EpolyaB_CXXC5 primers. All PCR products were cloned into both pBS-KS (-) and pcDNA3.1 (-) as described in Section 4.2.4. All generated constructs were sequenced at the METU Central Laboratories.

Table 6: Overlapping PCR conditions

	°C	Duration	Number of Cycles
Initial Denaturation	98	2 min.	1
Denaturation Annealing/ Extension	98 72	10 sec. 30 sec.	40
Final Extension	72	10 min.	1
Hold	+4	∞	1

2.5. Biochemical and Cellular Assays

2.5.1 Transfection

To perform immunocytochemistry (ICC) experiments, MCF7 cells were transiently transfected with the mammalian expression vector pcDNA3.1(-) bearing the *CXXC5* cDNA. To accomplish ICC, absolute EtOH sterilized round coverslips were placed into 12-well tissue culture plates and exposed to UV light for 60 min. $6x10^4$ MCF7 cells/well were then seeded onto cover slips in growth medium. 48 hours after the seeding, cells were transfected with transfection complex containing one µg DNA, two µl of Turbofect transfection reagent (Thermo Scientific, CA, USA), and 100 µl phenol red free medium. After 30-minute incubation at room temperature for the transfection complex formation, the mixture was applied onto cells. Four hours later, transfection medium was removed and cells were refreshed with growth medium. 36 hour after transfection, cells were fixed and ICC was performed.

For WB, MCF7 cells were transiently transfected as described above. 36 hours after transfection, cells were trypsinized and subjected to protein isolation.

2.5.1.1. siRNA Transfection

The FlexiTube GeneSolution GS51523 for *CXXC5* (Qiagen) was used for siRNA knock-down studies. There are 4 different siRNAs; three of which (siRNA2, 7 and

10) target the *CXXC5* transcript and one (siRNA9) targets the 3'UTR of the *CXXC5* transcript. The target sequences of siRNAs are given in Table 7.

Table 7: The siRNA sequences and targeting regions of *CXXC5*

siRNAs	siRNA sequence (5' to 3')	Target
Hs_CXXC5_2	CAGCAGTTGTAGGAATCGAAA	In the ORF of *CXXC5*
Hs_CXXC5_7	CAAGAGCGGTATCATCAGTGA	In the ORF of *CXXC5*
Hs_CXXC5_9	AGGGATTCGGGCGAAGACAAA	In the 3'UTR of *CXXC5*
Hs_CXXC5_10	TCAGATTTGCAAATTCAGAAA	In the ORF of *CXXC5*

FlexiTube siRNAs were received in lyophilized forms and they were dissolved in RNase free water to 10µM concentration and stored at -20^{o}C. The AllStars negative control siRNA (Qiagen, Germany) is a validated non-silencing siRNA that has no homology to any known mammalian gene sequence and it was used as a negative control in siRNA transfection experiments. The HiPerFect transfection reagent (Qiagen, Germany) was used with a Fast-Forward transfection protocol as follows: 1.2 µl siRNA (from 10 µM main stock) was added to 100 µl phenol red free medium. Mixture was then received 6 µl of HiPerFect transfection reagent. The mixture was vortexed briefly and incubated for 20 minutes for the transfection complex formation. Meantime, we seeded 16x10^{5} MCF7 cells/well into wells of a 12-well cell culture plate. The transfection mixture (Table 8) was then added drop-wise directly onto cell suspension immediately after placing it into wells. 24 hours after transfection, cells were trypsinized and collected for the total RNA isolation. The total RNA isolation, genomic DNA contamination control, cDNA synthesis

and qRT-PCR experiments for the analysis of the *CXXC5* expression levels were performed as described in Sections 2.3.1, 2.3.2, 2.3.3, 2.3.5.

For the protein isolation from siRNA treated cells, $32x10^4$ MCF7 cells/well were seeded into 6-well cell culture plate. The transfection ratios for 6-well culture are given in Table 8 and transfection was carried out as described above. 2.4 µl siRNA (from 10 µM main stock) and 12 µl of HiPerFect reagent were mixed with 100 µl phenol red free medium.

Table 8: siRNA transfection mixture

Culture Format	Volume of cell suspension (µl)	Volume of diluted siRNA (µl)	Volume of HiPerFect Reagent (µl)	Final siRNA concentration (nM)
12-well plate	1100	100	6	10
6-well plate	2300	100	12	10

2.5.2. Western Blot Analysis (WB)

2.5.2.1. Nuclear and Cytoplasmic Protein Isolation

To obtain nuclear and cytoplasmic proteins, we used the NE-PER system (Nuclear and Cytoplasmic Protein Extraction Kit; Thermo Scientific, CA, USA) with CERI, CERII and NER protein extraction buffers as instructed by the manufacturer. For two combined wells of a 6-well culture plate of 80-90% confluent cells, the ratios of CERI:CERII:NER buffers were 100:5.5:50 µl. Protease-inhibitor (Roche Applied Science, Switzerland) was added to the CERI and NER buffers. Isolated proteins were stored at -80^{o}C. The concentration of the proteins was assessed with the Quick Start Bradford Protein Assay (Bio-Rad, CA, USA).

2.5.2.2. Total Protein Isolation

The M-PER reagent (Mammalian Protein Extraction Reagent; Thermo Scientific, CA, USA) was used for the total protein isolation. For two combined wells of a 6-well cell culture plate of 80-90% confluent cells, we used 100 µl M-PER containing protease inhibitors (Roche Applied Science, Switzerland). Isolated proteins were

stored at -80^{o}C. The concentration of the proteins was assessed with the Quick Start Bradford Protein Assay (Bio-Rad, CA, USA).

2.5.2.3. Western Blotting

Extracted proteins were denatured in 6X Laemmli buffer (Appendix H) for five minutes at 95^{o}C. Denatured proteins were separated on a 10% SDS-PAGE and transferred onto a PVDF (Polyvinylidene fluoride, Roche Applied Science, Switzerland) membrane via wet-transfer. For the CXXC5 (Abcam, ab106533), Hdac1 (Abcam, ab19845) and β-actin (Abcam, ab8227) antibodies, membranes were blocked in 5% skim milk in 0.05% TBS-T (Tris Buffered Saline- Tween) for one hour at room temperature. The membrane was blocked in 5% BSA in 0.1% TBS-T for overnight at 4^{o}C for the Flag-M2 (Sigma Aldrich, MO, USA) antibody. The ab106533 CXXC5, HDAC1, β-ACTIN and Flag-M2 antibody dilutions for the protein detection were 1:500, 1:5000, 1:15000 and 1:1000 respectively. Primary antibody incubations were performed for one hour at room temperature for all antibodies. The secondary goat anti-rabbit horseradish peroxidase, HRP, conjugated (Santa Cruz Biotech., CA, USA) antibody was used for the ab106533, Hdac1 and β-actin antibodies. The secondary anti-mouse (horseradish peroxidase; HRP conjugated and raised in goat, Santa Cruz, Biotech., CA, USA) antibody was used for the Flag-M2 antibody. Secondary antibody dilutions were 1:2500. The secondary antibody incubation was performed for one hour at room temperature. For the visualization of protein bands, Clarity Western ECL Substrate (Bio-Rad, MO, USA) was used. For imaging of developed membranes, we used the ChemiDoc^{TM} MP system (Bio-Rad, MO, USA).

2.5.3. Immunocytochemistry (ICC)

Fixation and permeabilization: 36 hours after transfection, cells were washed with PBS for three times. 2% paraformaldehyde in PBS was added onto cells and waited for 30-minutes at room temperature for fixation. Cells were then washed with PBS for three times. 0.4% Triton-100X in PBS was added and incubated for 10 minutes at room temperature for permeabilization.

Blocking: Fixed cells were washed with PBS three times. 10% Bovine Serum Albumin (BSA) in PBS as the blocking agent for non-specific staining was used for the Flag-M2 (Sigma Alrich, MO, USA) and LaminB1 (ab16048, Abcam Inc., MA, USA) antibodies for one hour at room temperature by gently shaking. We used 10% Normal Goat Serum (NGS) in PBS for blocking of the ab106533 CXXC5 and LaminA (ab8980, Abcam) antibodies.

Primary Antibody: The Flag-M2 (1:500) and LAMINB1 (1:200) antibodies were diluted in 3% BSA in PBS and the ab106533 CXXC5 (1:250) and LAMINA (1:200) antibodies were diluted in 2% NGS in PBS. The antibody was added onto cells and cells were incubated for two hours at room temperature by gentle shaking.

Secondary Antibody: After the primary antibody incubation, cells were washed with PBS for three times and secondary antibodies were added. An AlexaFluor® 488 conjugated goat anti-mouse (ab150113, Abcam) secondary antibody (1:1000) was diluted in 3% BSA in PBS for Flag-M2. An AlexaFluor® 488 conjugated goat anti-rabbit (ab150077, Abcam) secondary antibody (1:1000) was diluted in 2% NGS in PBS for the ab106533 antibody. An AlexaFluor® 647 conjugated goat anti-rabbit (ab150083, Abcam) secondary antibody (1:250) was diluted in 3% BSA for LaminB1 antibody. An AlexaFluor® 647 conjugated goat anti-mouse (ab150119, Abcam) secondary antibody (1:250) was diluted in 2% NGS in PBS for LaminA antibody. Cells were incubated for 30-minutes at room temperature by gentle shaking.

Nuclei Staining: Cells were washed for three times with PBS. 1:1000 diluted DRAQ5^{TM} (Thermo Scientific, CA, USA), which is a fluorescent probe for a far-red DNA counter-staining, in PBS was added and incubated for 10 minutes at room temperature. Cells were then washed three times with PBS and 300 nM DAPI solution for DNA staining was added followed by 10 minute incubation at room temperature. Cells were then washed and subjected to both fluorescence (Leica, in the laboratory of Dr. Çağdaş Son, Middle East Technical University, Ankara, Turkey) and confocal microscopy (Leica, in the laboratory of Prof. Dr. Vasıf Hasırcı, Middle East Technical University, Ankara, Turkey).

2.5.4. Recombinant Protein Expression

2.5.4.1. Overexpression of the Histidine-Tagged CXXC5 in E.coli

The pet28a-6xHis-CXXC5 construct and pet28a-6xHis vector (empty vector) were transformed into BL-21(DE3) bacterial strain (in Section 2.4.5) and cells were grown for overnight. The overnight culture was 1:100 diluted into 25 ml LB 30µg/ml kanamycin containing medium. Cultures were incubated at 37^{o}C by shaking until the OD reached to 0.6 OD_{600}. After the desired OD, 1 ml uninduced culture was pelleted and stored at -80^{o}C for downstream analysis. The remaining 24 ml culture was induced with 1mM IPTG (Isopropyl β-D-1-thiogalactopyranoside) to final concentration and the culture was incubated at 30^{o}C for six hours. At the end of the incubation, the induced culture was spun down and cells were pelleted at 3000 rpm for 15 minutes for protein analysis and stored at -80^{o}C. For the optimization of the induction, different induction conditions (Table 9) were used.

Table 9: Different induction conditions for the pet28(a)-6xHis-Flag-CXXC5

Duration	Temperature	IPTG
Overnight	15^{o}C	1 mM
Overnight	20^{o}C	1 mM
6h	30^{o}C	1 mM
4h	37^{o}C	1 mM

2.5.4.2. Purification of 6xHis-tagged CXXC5

Bacterial pellet was lysed with 630 µl lysis buffer supplemented with 70 µl lysozyme (10 mg/ml) in PBS. After 30-minute incubation on ice, samples were sonicated at 90 watt for 6 times of 10 seconds pulse/five seconds resting cycles and centrifuged at 12.000 g for 20-minute at 4^{o}C. Supernatants were collected and loaded onto the Ni-NTA Spin column (Qiagen, Germany) as per instruction of the

manufacturer. The contents of buffers used for the purification are given in Table 10.

Table 10: Buffers for Ni-NTA Spin Kit

Buffers (pH 8.0) \Contents	NaH_2PO_4	NaCl	Imidazole	$ZnCl_2$
Lysis Buffer	50 mM	300 mM	10 mM	-
Wash Buffer	50 mM	300 mM	20 mM	-
Elution Buffer	50 mM	300 mM	500 mM	50 µM

2.5.4.3. Whole Cell Extract (WCE)

Cells, $16x10^4$ cells/well, were seeded into 6-well cell culture plates. Transient transfection with pcDNA3.1 vector (empty vector) and pcDNA3.1-Flag-ERα constructs were carried out as described in Section 2.5.1. 36 hours after transfection, cells were trypsinized and pelleted. Pellets were washed two times with cold PBS. Re-suspensions were re-pelleted and pellets were dissolved in 50 µl WCE buffer as described in Table 11. The samples were incubated on ice for 15-minute and tubes were tapped every five minutes. Samples were then subjected to freeze/thaw cycle (-80^{o}C/ice) for 3 times and centrifuged for 10-minute at 14.000 g. After centrifugation, supernatant was immediately taken out and stored at -80^{o}C.

Table 11: Whole Cell Extract Buffer

Components	Final Concentrations
Tris pH 7.5	40 mM
EDTA	1 mM
NP-40	0.1%
KCl	400 mM
Glycerol	20%

2.5.5. Electrophoretic Mobility Shift Assay

2.5.5.1. Electrophoretic Mobility Shift Assay (EMSA)

For EMSA, 5' end biotin-labeled oligonucleotides depicted in Table 12 were used.

Table 12: DNA sequences of CpG, methylated CpG and non-CpG oligonucleotides

Oligonucleotides	Sequences (5' to 3')
CpG_Up	CTGTCAGTAT CCGG TACAGGTCTA
CpG_Down	TAGACCTGTA CCGG ATACTGACAG
Methylated CpG_Up	CTGTCAGTAT CC^{M}GG TACAGGTCTA
Methylated CpG_Down	TAGACCTGTA CC^{M}GG ATACTGACAG
Non-CpG_Up	CTGTCAGTAT CAGT TACAGGTCTA
Non-CpG_Down	TAGACCTGTA ACTG ATACTGACAG

Oligonucleotides were diluted to 1 µg/µl in Tris-EDTA buffer. Up and down oligonucleotides were combined as described in Table 13 to obtain double-stranded oligonucleotides.

Table 13: Annealing conditions of oligonucleotides

Up_oligonucleotide	10 µl (10 µg)
Down_oligonucleotide	10 µl (10 µg)
TE Buffer	30 µl

Tubes containing the annealing mixture were placed in a heat block at 95^{o}C for five minutes. The heat block was turned off and reactions were allowed to gradually cool to room temperature. As a positive control, we used a 5' biotin-labeled double-stranded DNA that harbors the consensus ERE sequence for ERα-ERE interactions.

Table 14: DNA sequences of the consensus ERE oligonucleotides

Oligonucleotides	Sequences (5' to 3')
ERE-Up	GCGCCGAGTTATCA **GGTCA GAG TGACC** TGGAGATTCCTTA
ERE_Down	ATATAAGGAATCTCCA **GGTCA CTC TGACC** TGATAACTCGG

The oligonucleotides were diluted and annealed as described above. Concentrations of the double-stranded oligonucleotides were measured with NanoDrop 2000 and 40 fmol DNA was used for the experiments.

LightShift® Chemiluminescent EMSA Kit (Thermo, Scientific, CA, USA) was used for EMSA by following the Manufacturer's Instructions. The components of EMSA are given in Table 15.

Table 15: Components of the EMSA reaction

Components	Final Concentrations
Ultrapure Water	-
10X Binding Buffer (20148A)	1X
50% Glycerol (20148F)	2.5%
100mM $MgCl_2$ (20148I)	5mM
1µg/µl Poly (dI:dC) (20148E)	50 ng/µl
1% Np-40 (20148G)	0.05%
Biotinylated DNA	-

20 µl total reaction mixture was incubated for 30-minutes at room temperature. In the middle of the incubation time, one µl of the ab106533 CXXC5 (Abcam) antibody or HC-20 antibody specific for ERα (Santa Cruz Biotech., CA, USA) was

added into the reaction tube to assess the specificity of protein-DNA interaction. The samples were loaded onto a 5% native polyacrylamide gel (Table 16) which was pre-run (100V) for 40-minute. Electrophoresis was performed in 0.5X Tris/Borate/EDTA (TBE) (Appendix H) for 45-minute. Following electrophoresis, reactions were transferred onto a positively charged nylon membrane (Roche Applied Science, Switzerland) for visualization. The membrane was soaked in 0.5X TBE for 10-minute and a gel sandwich was prepared. Transfer was performed in cold (~ 10^{o}C) 0.5X TBE for 30-minutes at 380mA. Transferred DNAs were cross-linked to the membrane via UV transilluminator (markası) which has UV lambs with 312nm wavelength for 15-minute.

Table 16: Gel Retardation Assay 5% Native Polyacrylamide Gel

dH_2O	7.75 ml
30% Polyacyrlamide	1.25 ml
5X TBE	1 ml
10% APS	0.75 ml
TEMED	0.075 ml

2.5.5.2. Detection of the Biotin-labeled DNA

For the detection of DNA-protein interactions, the Chemiluminescent Nucleic Acid Detection Module (Thermo Scientific, CA, USA) was used as directed by Manufacturer. The developed DNA blot was visualized with the ChemiDoc^{TM} MP system (Bio-Rad, MO, USA).

CHAPTER 3

RESULTS AND DISCUSSION

3.1. Bioinformatics Tools

According to the Genome Browser database at the University of California Santa Cruz, *CXXC5* consists of three exons, two of which encode (in Section 1.2., Figure 2) a 322 amino acid long protein. The Ensemble Genome Browser predicts that CXXC5 has a charge of 13.0 with an isoelectric point of 9.5248. The mass of the protein was calculated to be as 32,977.35 g/mol, which approximates 33 kDa. The ProtParam [45] web tool suggests that 8% of the total amino acids of CXXC5 are negatively and 11.5% of the total amino acids are positively charged.

3.1.1. Family/Domain/Motif Analysis

The pfam bioinformatics tool was used for the analysis of putative domains of CXXC5. pfam (servers for UK: http://pfam.sanger.ac.uk/ and for USA: http://pfam.janelia.org/) is a database of protein families and family information, and contains 14,831 manually curated entries [46], [55]. pfam is based on the multiple alignments and hidden Markov model based profiles (HMM-profiles) [56]. pfam-based predictions suggest that due to a CXXC zinc finger domain at the carboxyl-terminus, CXXC5 is a member of the ZF-CXXC protein family. Although we have used various web tools to predict the structural features including the amino-terminus of CXXC5, we failed to obtain a feature for it.

Short linear motifs (SLiMs) within a protein display putative regulatory functions that could act semi-independently from the rest of the protein [57]. SLiMs are targets of various co- or post-translational modifications [57]. The eukaryotic linear motif (ELM, http://elm.eu.org) is a database for short motif determinations [47]. The ELM database consists of more than 200 different motif classes which have

been validated with more than 2400 experiments [47]. ELM results given in Table 17 suggest that CXXC5 can be post-translationally modified.

Table 17: Results of ELM

Elm Name	Elm Description	Positions
CLV_NRD_NRD_1	N-Arg dibasic convertase (NRD/Nardilysin) cleavage site (X-|-R-K or R-|-R-X)	259-261 271-273 283-285 293-295
CLV_PCSK_KEX2_1	Yeast kexin 2 cleavage site (K-R-|-X or R-R-|-X)	259-261 261-263 270-272 271-273
CLV_PCSK_PC1ET2_1	NEC1/NEC2 cleavage site (K-R-|-X)	259-261 261-263
DOC_USP7_1	The USP7 NTD domain binding motif variant based on the MDM2 and P53 interactions	252-256
MOD_CK1_1	CK1 phosphorylation site	251-257
MOD_Cter_Amidation	Peptide C-terminal amidation	255-258
MOD_GlcNHglycan	Glycosaminoglycan attachment site	254-257
MOD_GSK3_1	GSK3 phosphorylation recognition site	248-255
MOD_OFUCOSY	Site for attachment of a fucose residue to a serine	275-281
MOD_PKA_2	Secondary preference for PKA-type AGC kinase phosphorylation	283-289
TRG_ER_diArg_1	The di-Arg ER retention motif is defined by two consecutive arginine residues (RR) or with a single residue insertion (RXR). The motif is completed by an adjacent hydrophobic/arginine residue which may be on either side of the Arg pair	270-272 271-273
TRG_NLS_Bipartite_1	Bipartite variant of the classical basically charged NLS	284-304
TRG_NLS_MonoCore_2	Monopartite variant of the classical basically charged NLS. Strong core version	256-261 257-262 258-263
TRG_NLS_MonoExtC_3	Monopartite variant of the classical basically charged NLS. C-extended version	256-261 257-262
TRG_NLS_MonoExtN_4	Monopartite variant of the classical basically charged NLS. N-extended version	257-264 258-264 259-264

3.1.2. Post-Translational Modification Analysis

3.1.2.1. Prediction of Phosphorylation Sites

Protein phosphorylation is a dynamic and reversible process controlled by protein kinases (addition of phosphates) and protein phosphatases (removal of phosphates) [58]. Phosphorylation specifically occurs at serine, threonine and tyrosine residues of proteins. Protein phosphorylation is a very crucial mechanism in eukaryotes, especially in signal transduction pathways [58]. It has a major role in the regulation of cellular events that include proliferation, differentiation, metabolism and death [48]. For the prediction of putative phosphorylation sites of CXXC5, NetPhos 2.0 was used [48]. 15 serine and 2 threonine amino acids, indicated with above the threshold in Figure 5, throughout the protein were predicted as putative phosphorylation sites.

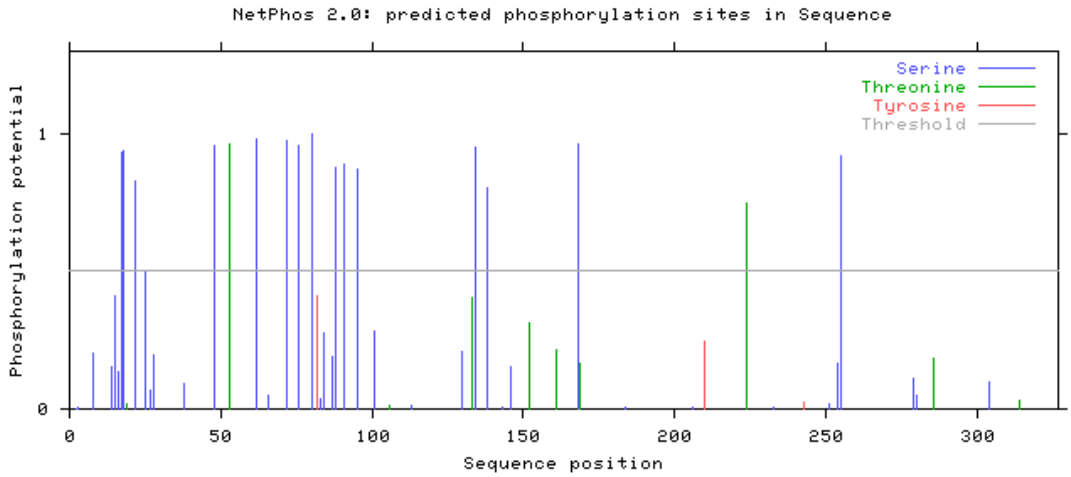


Figure 5: Predicted phosphorylation sites of CXXC5

3.1.2.2. Prediction of N-Acetylation Sites

N-terminal acetylation is a common co-translational covalent modification of proteins in eukaryotes and plays an important role in the synthesis, localization and stability of proteins. About 85% of all human proteins are acetylated at their N-terminus [59]. N-terminal acetylation of primarily serine residue occurs following the cleavage of the first methionine [59]. For the prediction of potential acetylation

sites in CXXC5, we used NetAcet 1.0 [49]. According to the prediction, two serine residues at the N terminus can be acetylated (Figure 6).

```
#
# NetAcet 1.0 prediction results, 1 sequence
#
#
# Sequence              #     Context   Score      Acetylation
# ------------------------------------------------------------
Sequence                2 S   --MSSLG   0.503          YES
Sequence                3 S   -MSSLGG   0.510          YES
```

Figure 6: Predicted acetylated sites of CXXC5

3.1.3. Potential ERE sequences

Based on our previous microarray results [5], *CXXC5* is an E2 responsive gene and regulated by the E2-ERα complex through the ERE-dependent signaling pathway. One conclusion is that the *CXXC5* gene locus contains an ERE(s) that provides responsiveness to E2-ERα. To analyze the potential ERE sequences of *CXXC5*, 10 kb upstream and downstream sequences from the translation initiation codon (ATG, Met -1) and from the translation termination codon (TGA, +1), respectively, of *CXXC5* were subjected to the Dragon ERE finder version 3, System for Identification and Interactive Analyses of Estrogen Response Elements in DNA Sequences (http://datam.i2r.a-star.edu.sg/ereV3/) [50]. Studies using *in vitro* binding assay and ChIP-seq analysis revealed that ERE sequences containing up to three nucleotide changes from the consensus ERE (5'-GGTCAnnnTGACC-3') could serve as DNA binding sites for E2-ER [60]. Using a matrix of maximum two mismatches sequence, a stringent analysis, compared to consensus ERE, we located seven potential ERE sequences in the *CXXC5* gene locus (Figure 7).

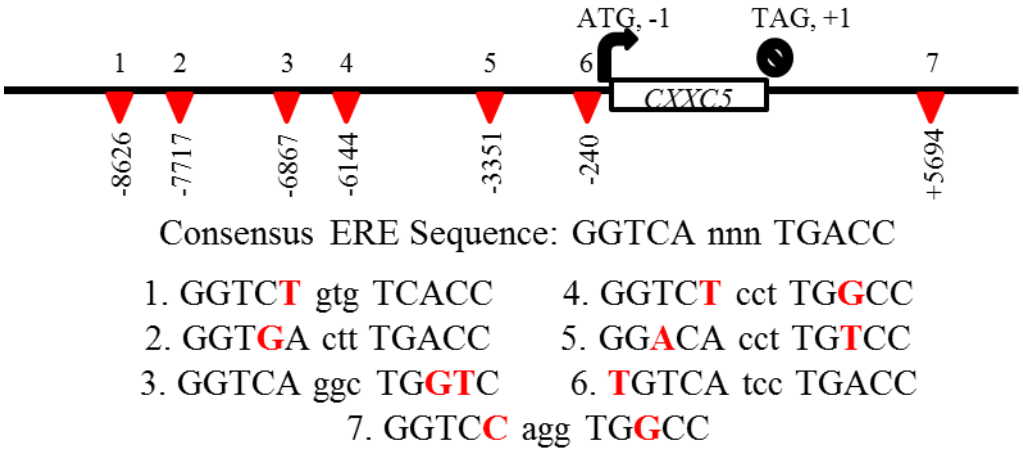


Figure 7: Predicted ERE sequences on the upstream and downstream sequences of the *CXXC5* gene locus

3.1.4. Homology Modeling Studies

CXXC5 is a recently identified protein and its structure/function is yet unclear. To begin to delineate the structural and functional properties of CXXC5, we initially analyzed *in silico* properties of CXXC5. Computational methods are being widely used for the functional and structural modeling of novel proteins. Homology modeling, which is one of these computational approaches, is based on the calculations of the energy within the given amino acid sequence of a query protein and attempt to select the energetically appropriate candidate protein (http://www.sciencedirect.com/science/article/pii/S0022283683713318). The purpose of the calculations is to predict a secondary structure for an unknown protein using a structurally known protein [61]. To predict structural features of CXXC5, we used the Phyre2 server (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index). Phyre2 is an open access software and with the help of the program, secondary structures of the unknown protein can be predicted [51]. Phyre2 also provides homology and analogy information with percentages of confidence and coverage between the query and template protein [62]. The Phyre2 library containing structurally known proteins is screened for structure prediction of query proteins by using profile-profile algorithms [51]. Members of the ten-eleven translocation (TET) gene family including Tet Methylcytosine Dioxygenase 1 and 3 (TET1 and TET3) are DNA

binding proteins and play roles in the DNA methylation process [23]. The conserved C-terminus CXXC domain is considered to be DNA binding domain of TET proteins [22]. Phyre2 results revealed that TET3, which is an active 5mC hydroxylase regulating the 5mC/5hmC status at target gene promoters, has the highest structural similarity to CXXC5 with a 99.5% confidence. However, the coverage of the homolog region encompasses only 16% of CXXC5 (Figure 8). This homology region is a 59 amino acid long stretch located within the CXXC domains of TET3 and CXXC5.

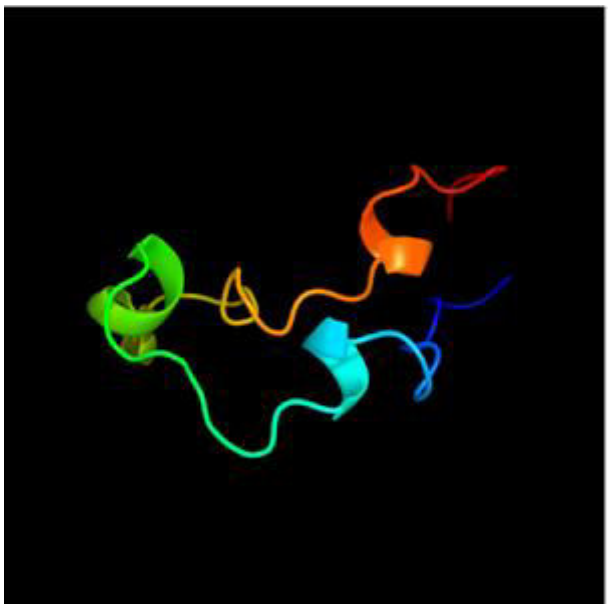


Figure 8: CXXC5 structure prediction by Pyhre2. Pyhre2 uses the amino acid sequence of the query protein for the secondary structure prediction. NP_057547.5 sequence (NCBI, CXXC5 protein) was uploaded onto the computational analysis page. The hit parts of the amino acid sequence within the protein in the Pyhre2 library was used as the template for folding of CXXC5. The 16% CXXC5 amino acid sequence is covered with 99.5% confidence by the template protein was revealed to be TET3 (*Xenopus laevis*).

Pyhre2 results were verified by psi-blast from NCBI using the PDB database (http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE=Proteins&PROGRAM=blastp&RUN_PSIBLAST=on) and TET3 was again revealed the highest hit for the CXXC5 homology modeling.

The tertiary structure of TET3 bound to CpG dinucleotide containing DNA fragment was identified using X-Ray Diffraction method and the data are publicly available in PDB [17]. We used TET3 (PDB code: 4HP3) as the template protein

for the prediction of tertiary structure of the CXXC5-CXXC domain. Modeller [52] was used for the modeling of the CXXC domain of CXXC5, while Chimera 1.8.1 were utilized for the visualization of the prediction [53]. The superimposition of the predicted tertiary structure of the CXXC domain of CXXC5 and that of TET3 indicates a high degree of structural similarity (Figure 9).

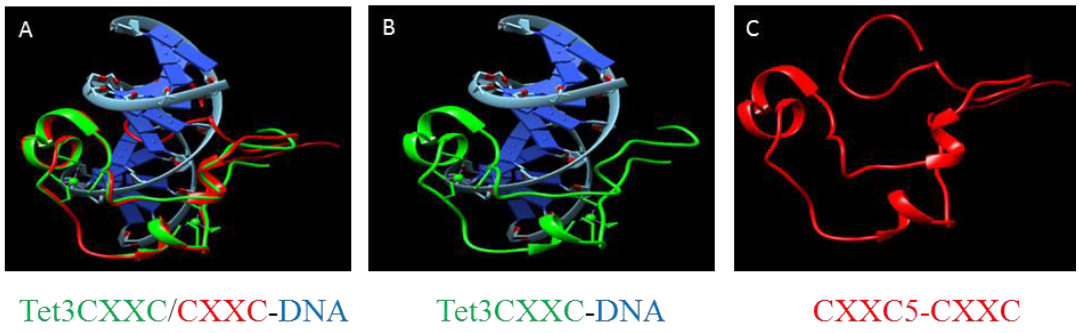


Figure 9: The superimposition of the CXXC domains of CXXC5 and TET3. A) The CXXC domains of CXXC5 and TET3 are superimposed. B) The DNA and TET3 complex is shown. C) The predicted tertiary structure of CXXC domain of CXXC5 is demonstrated.

Although TET3 as other TET proteins shows an overall poor amino-acid homology (16%) to CXXC5, the remarkable degree of similarity (97%) of the DNA binding domain of TET3 to the carboxyl-terminus CXXC domain of CXXC5 suggest that CXXC5 is also a DNA binding protein that binds preferentially to non-methylated CpG dinucleotides.

3.2. *CXXC5* is an Estrogen Responsive Gene

Since our previous studies using microarrays suggest that *CXXC5* is an E2-ERα responsive gene regulated through the ERE-dependent signaling pathway, we wanted to verify with RT-qPCR that the expression of *CXXC5* is indeed mediated by E2 signaling in cells synthesizing ERα. We used the expression of *TIFF1/pS2*, a well characterized E2-ERα responsive gene as a positive control. MCF7 cells, derived from a breast adenocarcinoma and kindly provided by Dr. A. Elif Erson-Bensan (Middle East Technical University, Ankara, Turkey), is an E2 responsive and ERα positive cell line that are extensively used as a model experimental system for ERα-positive breast cancers. To assess the expression of *CXXC5* in response to

E2, MCF7 cells in T-25 culture flasks were maintained in phenol-red free DMEM medium, containing charcoal dextran-treated fetal bovine serum (CD-FBS) to remove endogenous steroid hormones, for 48h. Cells were then treated with 10^{-9} M (1 nM) E2 or with and 0.01 % (v/v) EtOH as control for 3h or 24h. Samples were then subjected to total RNA extractions and RT-qPCR (Figure 10). Reactions were normalized by using the expression of *RPLP0* or homolog of Pumilio, Drosophila, 1 (*PUM1*) genes which are shown to be the most reliable reference genes for normalization of RT-qPCR results in breast carcinomas [54].

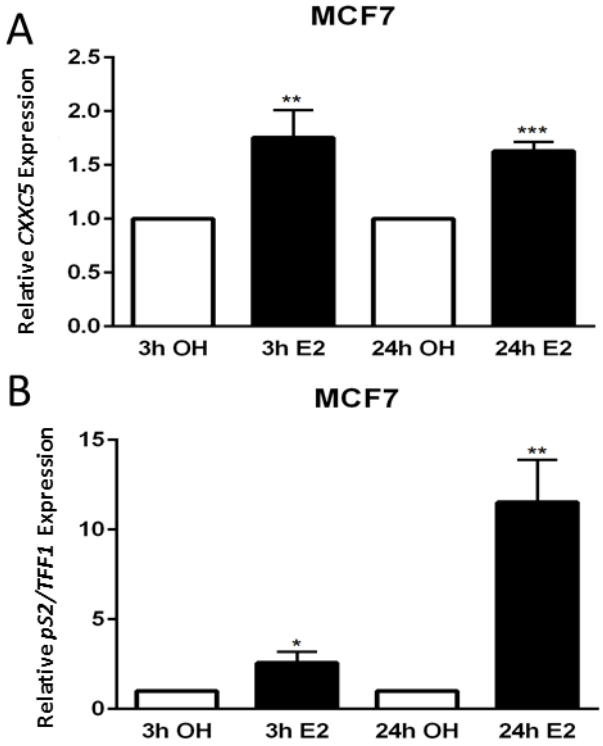


Figure 10: *CXXC5* is an estrogen responsive gene. Total RNA isolation was carried out from MCF7 cells which were grown in T25 flask containing charcoal-dextran treated fetal bovine serum (CD-FBS) for 48h and then treated with E2 (1 nM) or vehicle control EtOH (0.01 %) for 3h or 24h. RT-qPCR was performed three independent times with three technical replicates and the reference gene *RPLP0* was used for the normalization of A) *CXXC5* and B) *pS2/TFF1*. Comparison of the E2 and EtOH treatment within the time points indicates the fold change. *, ** and *** represent significant changes ($P<0.05$; One-tailed, paired t-test with 95% confidence interval) in response to E2 compared to EtOH control within each time point.

For the *CXXC5* expression (amplicon length: 145 bp), the reaction conditions were as follows: Incubation at 94° C for 10 minutes followed by 40 cycles of 30 seconds at 94°C, 30 seconds at 60°C and 30 seconds at 72°C. The reaction condition for

pS2/TFF1 (amplicon length: 209 bp) includes incubations at 94° C for 10 minutes followed by 40 cycles of 30 seconds at 94°C and 30 seconds at 55°C and 30 seconds at 72°C.

For the expression of *RPLP0* (amplicon length: 191 bp) and of *PUM1* (amplicon length: 111 bp), samples were incubated at 94° C for 10 minutes then 40 cycles of 30 seconds at 94°C, 30 seconds at 60°C and 30 seconds at 72°C.

All experiments were conducted with three biological replicates in triplicates. ΔΔCq was used for relative quantification [63]. In E2 treatments, levels in EtOH treated samples were set to one for each time point. In ICI treatments, expression levels in EtOH treated samples were set to one. E2, ICI or E2+ICI treated sample was compared to EtOH treated sample. One tailed paired t-test of mean ± SD, with 95% confidence was performed using GraphPad Prism version 6 for Windows (GraphPad Software, San Diego California USA).

Results revealed that *CXXC5* is up-regulated at 3h and 24h in response E2. To obtain evidence that the E2-mediated modulation of *CXXC5* gene expression is ER-dependent, we used a complete ER antagonist ICI (Imperial Chemical Industries 182,780) [64]. The ICI (100nM) treated samples were subjected to RT-qPCR (Figure 11). Samples were normalized by using the expression of *PUM1*.

We found that E2 response of *CXXC5* at 24h is lost when E2 (1nM) and ICI (100nM) were used together. Results indicate that E2-mediated up-regulation of the *CXXC5* gene expression is ER-dependent. In addition, there was no change in the expression levels of *CXXC5* when cells were treated only with ICI. It demonstrates that ICI has no effect on the basal expression of *CXXC5*.

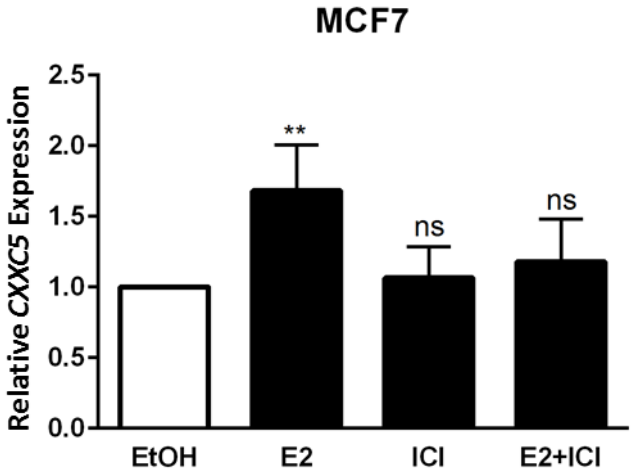


Figure 11: Up-regulation of *CXXC5* in response to E2 is ER-dependent. MCF7 cells were grown in T25 tissue culture flask containing charcoal-dextran treated fetal bovine serum for 48h and then treated with E2 (1 nM), vehicle control EtOH (0.01 %) and/or ICI (100 nM) for 24h. Cells were then subjected to total RNA isolation. RT-qPCR was performed three independent times with three technical replicates and the reference gene *PUM1* was used for the normalization. ** represents significant changes ($P<0.05$; One-tailed, paired t-test with 95% confidence interval) in response to E2 compared to EtOH control. 'ns' denotes non-significance.

3.3. Cloning of the ORF of *CXXC5* into Expression Vectors

To study CXXC5, the ORF of *CXXC5* cDNA was cloned into a high copy number plasmid, pBS-KS(-) and into a mammalian expression vector, pcDNA3.1(-). The ORF of *CXXC5* in an expression plasmid, pMigR1-*CXXC5*, was kindly provided by Dr. F. Pendino, INSERM, Paris, France. The pMigR1-*CXXC5* cDNA was transformed into the XL-1 Blue bacterial strain and two different colonies were chosen for plasmid isolation using the Qiagen Mini Prep plasmid isolation system (Qiagen, Germany). Both bacterial colonies were subjected to PCR using the *CXXC5*-ORF cloning primers to ensure the presence of the cDNA in the isolated colony (in Section 2.4.2). The PCR products were run on a 1% agarose gel (Figure 12).

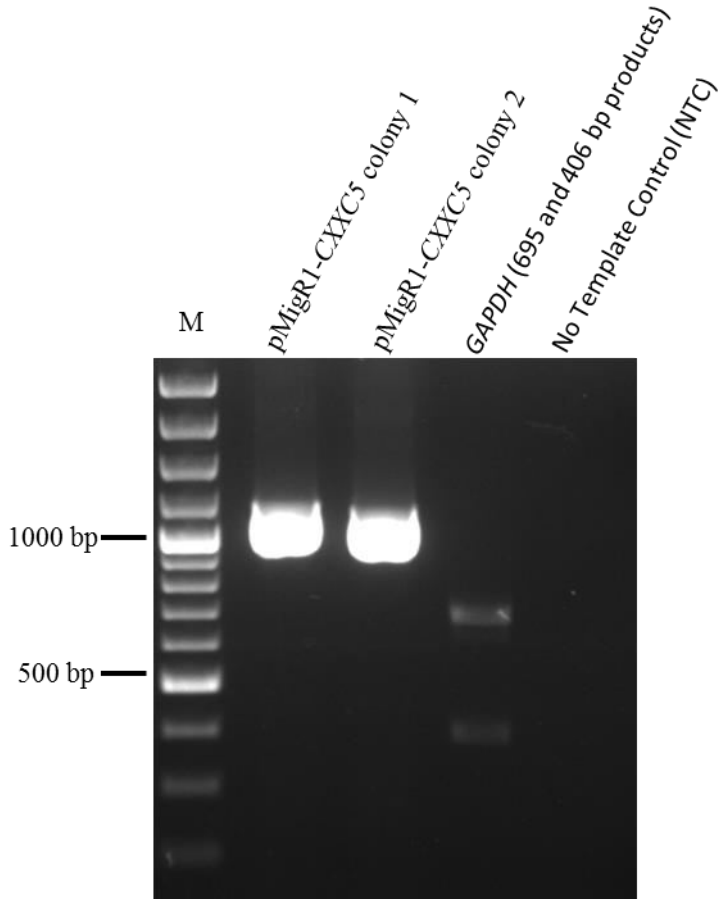


Figure 12: Colony PCR. PCR products of two pMigR1-*CXXC5* plasmids obtained from two different bacterial colonies. M: DNA Ladder (GeneRuler™ DNA Ladder Mix 100-10000 bp, Thermo Scientific, CA, USA)

After the verification that both plasmids bear the *CXXC5*-ORF, one of the PCR products (#1) was subjected to PCR purification and to restriction enzyme digestion (in Section 2.4.3). We also subjected the pBS-KS (-) vector to digestion using the same restriction enzymes (in Section 2.4.3). After the ligation of the *CXXC5*-ORF cDNA fragment into pBS-KS(-) and transformation of the ligated products into the XL1-Blue cells, we selected colonies resistant to Ampicillin. Colonies were then subjected to colony PCR using cloning primers. Plasmid isolation was carried out with one of the colonies positive for pBS-KS(-) bearing the *CXXC5*-ORF cDNA and the isolated plasmid was sequenced at the METU central laboratories. After the sequence verification, the *CXXC5*-ORF cDNA was excised with restriction enzymes and subcloned into the pcDNA3.1.(-) or Flag-pcDNA3.1.(-), which bears

sequences encoding an N-terminally located Flag epitope, linearized with the same restriction enzymes (in Section 2.4.4). This cloning strategy allowed us to have a mammalian expression vector for the expression of the wild-type or the Flag-*CXXC5* cDNA.

3.4. Analysis of the Endogenously or Exogenously Expressed CXXC5 Synthesis in MCF7 Cells

3.4.1. Western Blot (WB)

To examine the synthesis of CXXC5, we initially used MCF7 cell line, in which we had found by RT-qPCR that *CXXC5* is an E2-ERα responsive gene. Of the commercially available limited number of antibodies specific to CXXC5, only the Abcam Inc. antibody (ab106533) reportedly detects a protein band in WB analysis that shows a molecular mass of 33 kDa approximating the calculated mass of CXXC5. This polyclonal ab106533 antibody is raised against a synthetic peptide that corresponds to an amino acid sequence close to the amino terminus of CXXC5 (http://www.abcam.com/cxxc5-antibody-ab106533.html). Based on this information, we purchased and test the efficacy of the antibody to recognize the endogenous as well as the recombinant wild-type and Flag-CXXC5 proteins in WB and ICC.

To examine the efficacy of the ab106533, we initially carried out WB analysis of nuclear (100 µg) and cytoplasmic (100 µg) extracts of MCF7 cells transfected with the mammalian expression vector pcDNA3.1(-) bearing none (as control), the *CXXC5* cDNA or the Flag-*CXXC5* cDNA for 36h. In WB analysis, we used both the Flag-M2 (data not shown) and the ab106533 antibodies. In previous studies when samples from cells synthesizing exogenously introduced *CXXC5* cDNA were loaded equal amounts to those from cells transfected with pcDNA3.1(-) bearing no cDNA as control for endogenous CXXC5, we observed an overshadowing effect of the overexpressed protein on WB membranes that hinders the detection of endogenous proteins. This necessitated the use of lower concentrations of extracts from cells transfected with expression vectors bearing *CXXC5* cDNA (Figure 13).

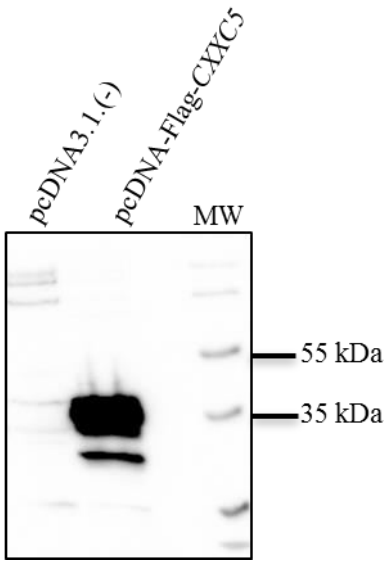


Figure 13: Equal amount of loading results in an overshadowing effect. Equal amounts (100 µg) overexpressed and endogenous proteins loaded onto the 10% SDS-PAGE. The overexpressed sample was rapidly overshadows (within milliseconds) the endogenous protein detection (seconds) and renders the detection of endogenous protein difficult. ab106533 CXXC5 antibody was used. MW: Molecular weight marker.

We found that the overexpressed Flag-CXXC5 (blue star, Panel A in Figure 14) can be detected with both the ab106533 and the Flag-M2 (data not shown) antibodies at comparable levels in the nuclear fraction and the overexpressed wild-type CXXC5 (red star, Panel A in Figure 14) can also be detected in nuclear extract by using ab106533 antibody. It should be noted that low amounts of the overexpressed protein can also be seen in cytoplasmic fractions (Panel B in Figure 14). We also observed that HDAC1, a nuclear protein as loading control for nuclear fractions (conversely, the small yet detectable amounts of β-actin (ab8227) as loading control for cytoplasmic control in nuclear fractions) is, although low, also be detected in cytoplasmic extracts (data not shown). It is therefore likely that the detection of CXXC5 protein in cytoplasmic fractions is due to contaminating nuclear fractions, suggesting to somewhat inefficiency of the protein extraction system (NE-PER, Nuclear and Cytoplasmic Protein Extraction Kit; Thermo Scientific, CA, USA) we used.

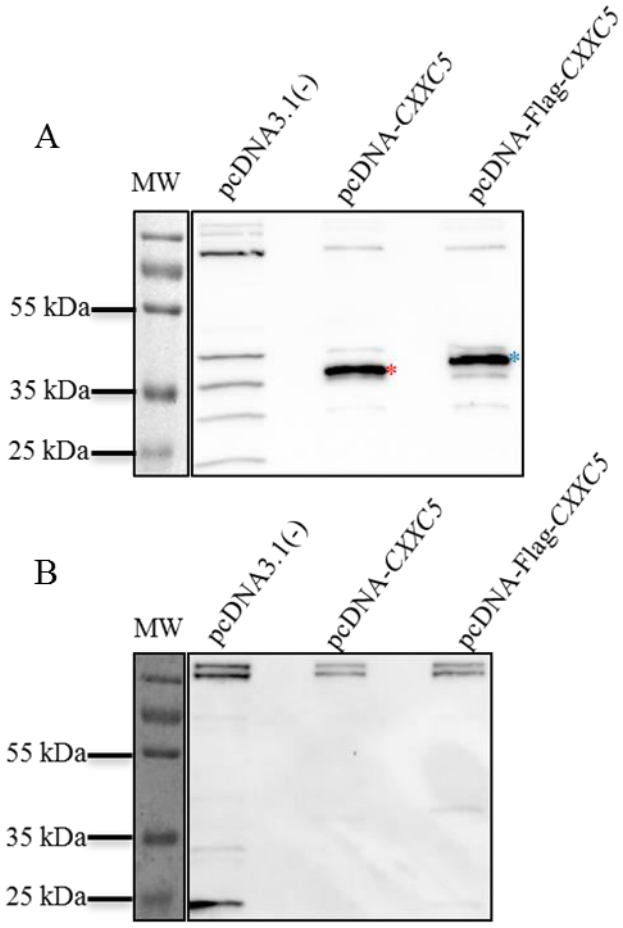


Figure 14: Detection of the endogenous and overexpressed CXXC5 protein in MCF7 cells. For the growth of MCF7 cells, we used 6-well tissue culture plates. 48h after seeding, MCF7 cells were transfected with pcDNA3.1(-) bearing none (as control) (100 µg), the *CXXC5* cDNA (25 µg) or the Flag-*CXXC5* cDNA (25 µg). *A)* Nuclear and *B)* Cytoplasmic lysates were obtained from cells 36h after transfection. Indicated amounts of proteins were loaded onto the 10% SDS-PAGE and WB was performed using the ab106533 antibody. In panel A; red star represents the overexpressed wild-type CXXC5 and blue star denotes the overexpressed Flag-tagged CXXC5. MW: Molecular weight marker. A representative image from two independent experiments is shown.

It should also be noted that the overexpressed wild-type CXXC5 or the Flag-CXXC5 protein shows an electrophoretic migration slower than the expected molecular mass (33kDa) of the endogenous protein in WB as depicted in Figure 14, Panel A. The presence of the N-terminally added Flag epitope is likely the reason of the difference in the migration of the wild-type and Flag-tagged CXXC5 due to approximately 2900 Da molecular mass of the Flag epitope. In addition, results revealed many protein bands with varying molecular masses in the pcDNA3.1(-) (control, as endogenous CXXC5 protein) transfected sample that has no matching

protein bands to the overexpressed wild-type or to the Flag-tagged CXXC5 protein. Under these conditions, it was difficult to identify CXXC5 (Figure 14, Panel A). To decipher which protein band corresponds to the endogenous CXXC5 protein, we decided to use siRNA approach to reduce the intracellular levels of the endogenous protein. For this purpose, we utilized the FlexitubeGene Solution (Qiagen, Germany) containing four siRNAs (the siRNA sequences were given in Section 2.5.1.1 in Table 7), three of which specifically target encoding *CXXC5* transcript and one targets 3'UTR of *CXXC5*. We also used AllStars Negative Control siRNA (Qiagen, Germany) which has no homology to any known sequence (https://www.qiagen.com/tr/products/catalog/assay-technologies/rnai/allstars-negative-control-sirna/?catno=SI03650318) as negative control. Firstly, to ensure that siRNAs indeed reduce the intracellular levels of the *CXXC5* transcript, we subjected total RNA samples from MCF7 cells transfected with a *CXXC5* specific siRNAs or the control siRNA to RT-qPCR (Figure 15). For the normalization of results, we used the expression of *RPLP0*. We found that siRNA2 or siRNA10 effectively repressed whereas siRNA7 or siRNA9 reduced the intracellular levels of the *CXXC5* transcript compared to levels observed in untransfected (or AllStars Negative Control siRNA transfected) cells.

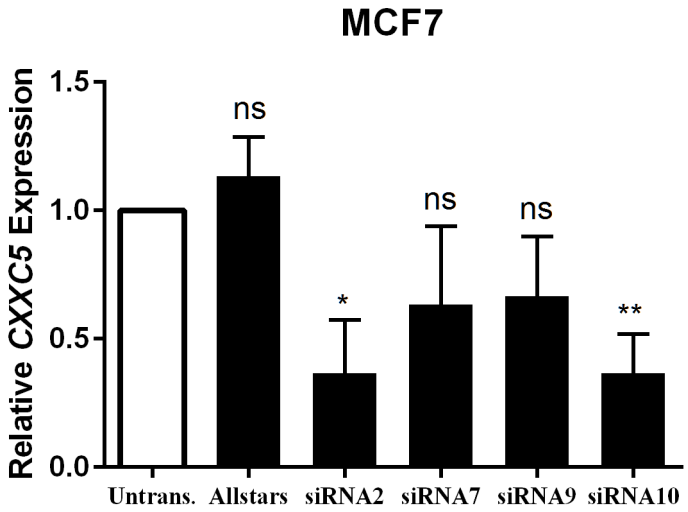


Figure 15: Transfection of siRNA specific to the *CXXC5* transcipt into MCF7 cells. Total RNA was isolated from the untransfected, siRNA2, 7, 9, 10 or control siRNA transfected MCF7 cells 24h after the transfection. Transfection was carried out with cell suspensions in wells of 12-well culture plates. *CXXC5* transcript levels in specific siRNA or control siRNA treated samples were compared to levels of the *CXXC5* transcript in untransfected sample. Results are the mean± SD of three technical replicates conducted at three independent times. * and ** denote the significant changes ($P<0.05$; One-tailed, paired t-test with 95% confidence) in the presence of siRNA2 and 10. 'ns' indicates non-significance.

Based on RT-qPCR results of siRNA transfections and since the overexpressed CXXC5 was observed in nuclear fractions, we carried out WB analysis of nuclear extracts (100 µg) of MCF7 cells transfected with siRNA2, 9, 10 or the control siRNA for 24h. To examine the influence of the siRNA transfection on the putative endogenous CXXC5, we first carried out WB using the protein extracts from the negative control Allstars siRNA transfected cells. We did not observe any chages in the levels of protein bands when cell extracts from Allstar siRNA control were compared to those of untransfected cells as depicted in Figure 16.

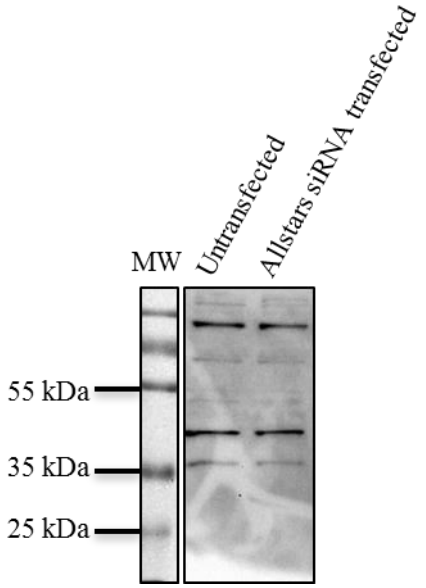


Figure 16: Transfection of Allstars negative siRNA followed by WB. MCF7 cells were transfected without or with Allstars negative siRNA in suspension in wells of 6-well culture plates. 24h after transfection, cells were subjected to the nuclear protein isolation. 100 µg nuclear extracts loaded on 10% SDS-PAGE. WB was performed with the ab106533 antibody. MW: Molecular weight marker is shown. A representative image from two independent experiments is shown.

After the verification of the control siRNA did not alter the levels of protein bands, we transfected MCF7 cells without or with siRNA2, 9 and 10. Nuclear fractions (100 µg) were subjected to WB together with pcDNA3.1(-) bearing none (as a control, endogenous) or the wild-type-*CXXC5* cDNA (Figure 17). We found that the transfection of cells with siRNA10 completely and siRNA2 and 9 partially reduced the detection of a protein band with a molecular mass close to that of CXXC5. Nevertheless, the predicted CXXC5 protein shows again a faster electrophoretic migration compared to the overexpressed wild-type CXXC5 band (The difference between the red labeled and blue labeled proteins). The reason of the migratory differences between the endogenous and exogenously introduced CXXC5 in WB are not clear. The overexpression of CXXC5 could lead to an aberrant post-translational modification(s) (PTM) that alters the mass and/or the electrophoretic migration of the protein compared to those of the endogenous protein.

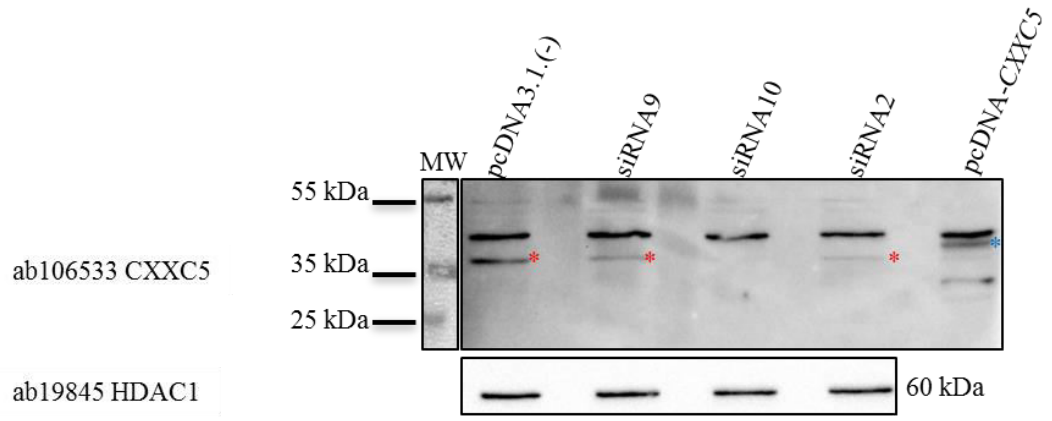


Figure 17: Deciphering of the endogenous CXXC5 in MCF7 cells. MCF7 cells in suspension were transfected with siRNA2, 9, and 10 in wells of 6-well culture plates. We also transfected cells with pcDNA bearing none (as endogenous control) or the wild-type-*CXXC5* cDNA as overexpression control. 24h after siRNA transfection, the nuclear protein isolation was carried out. 100 µg proteins from cells transfected with siRNA or pcDNA3.1(-) and 25 µg protein of overexpression control sample were loaded on 10% SDS-PAGE. WB was performed with the ab106533 antibody. For loading control, the Histone deacetylase 1 (Hdac1) antibody was used (Since overexpressed CXXC5 was loaded as 25 µg, the Hdac1 result of was not shown). MW: Molecular weight marker is shown. Red stars indicate the endogenous CXXC5 and the overexpressed CXXC5 is denoted with blue star. A representative image from two independent experiments is shown.

Based on these results, we think that a protein with an apparent molecular mass of 37 kDa is the endogenous CXXC5 protein. We subsequently examined the synthesis of CXXC5 in various cell lines derived breast carcinomas to examine and to select cell models synthesizing low or high levels of CXXC5 for further studies on CXXC5 functions. Total protein extracts, in the amount of 100 µg, of each cell lines, kindly provided by Dr. Begüm Akman of the Erson-Bensan laboratory, Biological Sciences, METU, Ankara-Turkey (Table 18), and equal amount of the overexpressed wild-type CXXC5 protein as a control were subjected to WB analysis. Since total protein extracts were used in the screening of cell lines, we selected the detection of the cytoplasmic structural protein β-actin (ab8227) as the loading control.

Table 18: Triple negative and ER-positive cell lines

Triple Negative Cell Lines	ER Positive Cell Lines
Cal-51	JIMT-1
HDQ-Q1	BT474
MDA-MB-453	EFM-19
MDA-MB-231	MCF7
HCC1143	SKBR3
HCC1937	CAMA-1
Cal-85-1	

We found that CXXC5 was synthesized at varying levels in triple negative [that do not express *ER*, Progesteron Receptor (*PR*) or the human epidermal growth factor receptor 2 (*HER2*) gene] (Figure 18) or in ER-positive (Figure 19) breast cancer cell lines.

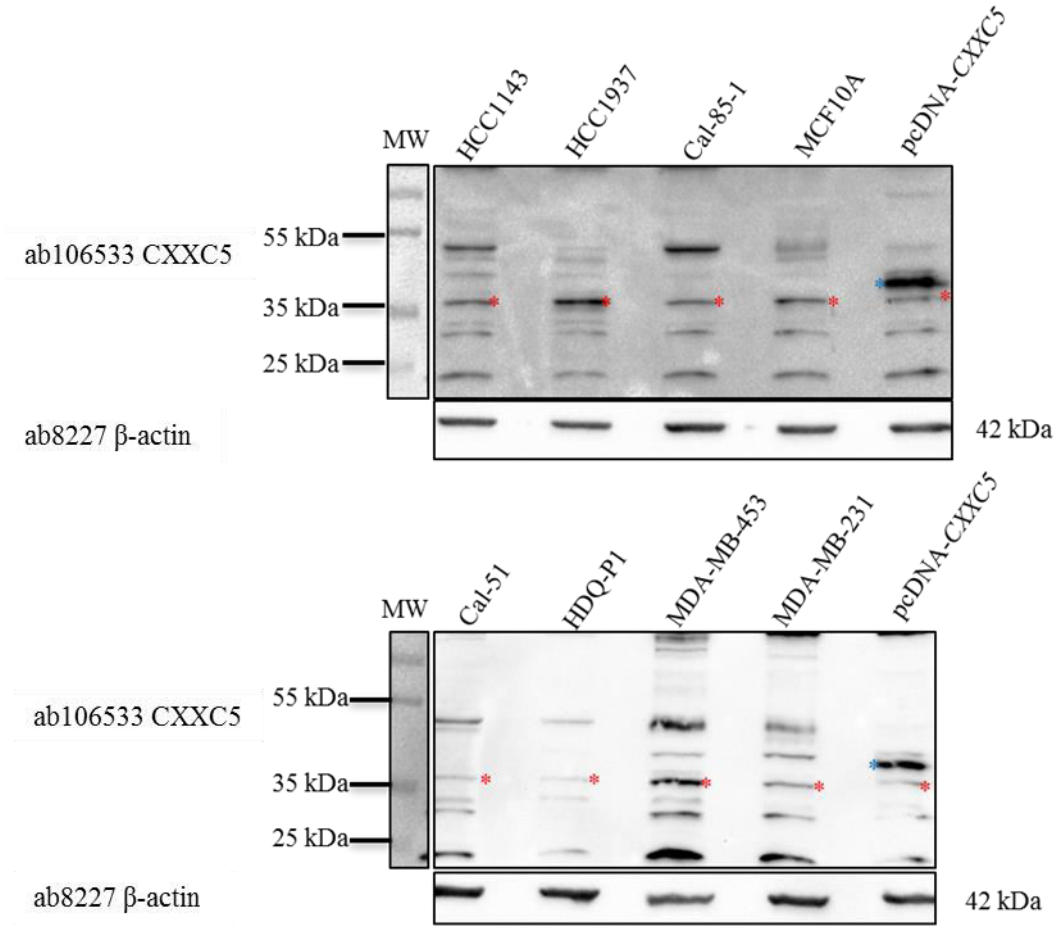


Figure 18: Screening of triple negative breast cancer cell lines. Total proteins, 100 µg, of triple negative cell lines and of total protein from MDA-MB-231 cells transfected with pcDNA3.1(-)-*CXXC5* were loaded onto 10% SDS-PAGE. WB was carried out using the ab106533 antibody. Protein bands corresponding to the endogenous CXXC5 are shown in red star and the overexpressed wild-type CXXC5 protein is shown with blue star. Cell lines are indicated. The β-actin (ab8227) antibody was used to detect β-actin (42 kDa) as loading control.

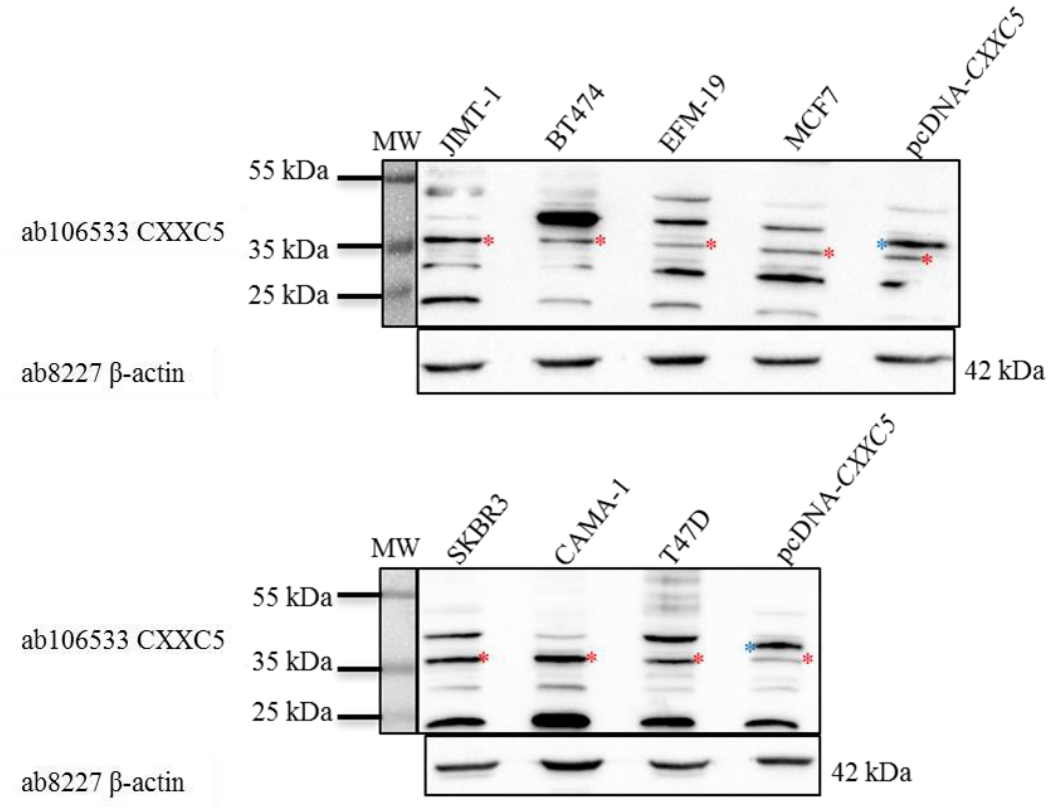


Figure 19: Screening of ER-positive breast cancer cell lines. 100 µg of total proteins of ER positive cell lines and of total protein from pcDNA3.1(-)-*CXXC5* transfected MDA-MB-231 cells were loaded onto 10% SDS-PAGE. WB was carried out by using the ab106533 antibody. Protein bands corresponding to the endogenous CXXC5 are shown in red star and the overexpressed wild-type CXXC5 protein is shown with blue star. Cell lines are indicated. The β-actin antibody (ab8227) was used to detect β-actin (42 kDa) as loading control**.**

3.4.2. Immunocytochemistry (ICC)

CXXC5 with the amino-terminally located Flag epitope was previously shown to localize in the nucleus using a Flag antibody in NB4 cell line, derived from leukemia blast cells of a human acute promyelocytic leukemia [21]. To recapitulate this finding and to verify our WB findings with which we detect the endogenous and overexpressed in the nuclear fractions in MCF7 cells, we transiently transfected cells with pcDNA3.1(-) bearing none or the Flag-*CXXC5* cDNA and subjected them to immunocytochemistry (ICC) using both the ab106533 and the Flag-M2 antibodies. The overexpressed CXXC5 was detected in the nucleus with either antibody (Figure 20) with no gross difference between the detection efficacies of two antibodies (Panel A and D). That no staining with the Flag-M2 antibody in cells transfected with the vector bearing no cDNA as control (data not shown and also untransfected cells of the cell population) was observed suggests that the nuclear staining is specific to the Flag-CXXC5 protein.

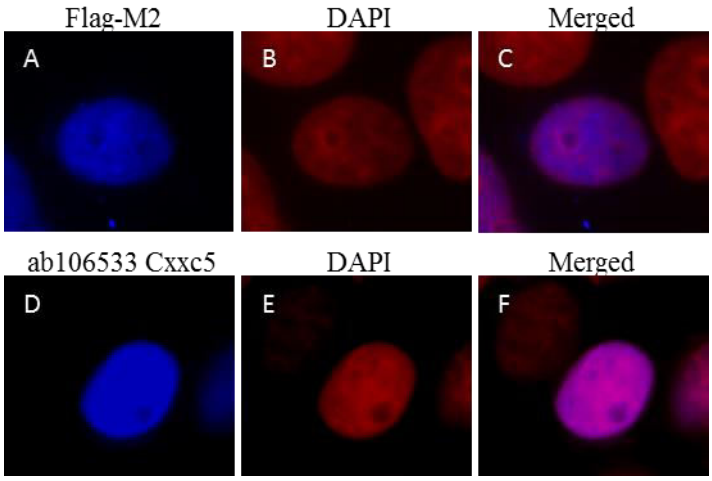


Figure 20: ICC for the overexpressed Flag-tagged CXXC5. MCF7 cells were grown on coverslips in 12-well culture plates in growth medium containing 8% fetal bovine serum (FBS). 48h later, pcDNA3.1(-)-Flag-*CXXC5* was transiently transfected into cells. 36h after transfection, cells were fixed with 2% paraformaldehyde in PBS and permeabilized with 0.4% Triton-X100 in PBS. For Flag-M2 antibody, 10% bovine serum albumin (BSA) was used for blocking for one hour. Cells were then incubated with the Flag-M2 antibody (1:250) in PBS containing 3% BSA for two hours. After the primary antibody incubation, an Alexa Fluor®-488 (green channel) conjugated goat anti-mouse secondary antibody (1:1000) in PBS containing 3% BSA was applied for 30-minutes. For the ab106533 antibody, we used 10% normal goat serum (NGS) for blocking for one hour. Cells were then incubated with the ab106533 antibody for two hours (1:250) in PBS containing 2% NGS. After the primary antibody incubation, an Alexa Fluor®-488 (green channel) conjugated goat anti-rabbit (1:1000) secondary antibody in PBS containing 2% NGS was applied for 30-minutes. For nuclei staining, DAPI (blue channel) was used. Images were obtained with aforementioned channels; then to emphasize the overlapping regions artificial coloring was used to display the overlapping areas. The Flag-CXXC5 staining is in blue (with both antibodies) and DAPI staining in red. A representative image from two independent experiments is shown. *A)* Channel for the Flag-M2 antibody *B)* Channel for DAPI staining *C)* Merged image of the Flag antibody and DAPI staining *D)* Channel for the ab106533 staining *E)* Channel for DAPI *F)* Merged image of the ab106533 and DAPI staining.

In ICC experiments, we also used Flag-ERα that bears an amino-terminally located Flag epitope as positive control for the Flag antibody staining and intracellular location. Flag-ERα, like the wild-type ERα, is localized in the nucleus [5]. By this way, we would confirm that the Flag-M2 antibody indeed recognize the Flag-epitope in ICC. pcDNA3.1(-) bearing none or the Flag-ERα cDNA was transiently transfected into MCF7 cells. 36h after transfections, cells were subjected to ICC using the Flag-M2 antibody (in Figure 21). As expected, the Flag antibody detected the Flag-ERα in the nuclei of cells transfected with the expression vector bearing the Flag-ERα cDNA but not with the vector bearing no cDNA (data not shown and untransfected cells in the cell population).

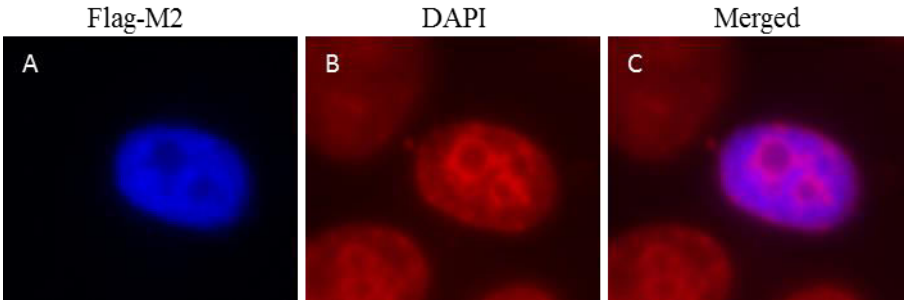


Figure 21: ICC for the overexpressed Flag-tagged ERα in MCF7 cells. MCF7 cells were grown on coverslips in 12-well culture plates in growth medium containing 8% fetal bovine serum. 48h later, the pcDNA3.1(-)-Flag-ERα transfection was performed. 36h after transfection, cells were fixed with 2% paraformaldehyde in PBS and permeabilized with 0.4% Triton-X100 in PBS. 10% bovine serum albumin (BSA) in PBS was used for blocking for one hour. Cells were then treated with the Flag-M2 antibody for two hours (1:250 in PBS with 3% BSA). After the primary antibody incubation, an Alexa Fluor®-488 (green channel) conjugated goat anti-mouse secondary antibody (1:1000 in PBS with 3% BSA) was applied for 30-minutes. For the nuclei staining, DAPI (blue channel) was used. Images were obtained with aforementioned channels; then to emphasize the overlapping regions, the artificial coloring was used to display the Flag-ERα staining in blue and DAPI staining in red. A representative image from two independent experiments is shown. *A)* Channel for the anti- Flag antibody *B)* Channel for the DAPI staining *C)* Merged image of the Flag antibody and DAPI staining.

Since, our RT-qPCR results suggest that *CXXC5* is an E2 and ERα responsive gene (Figure 10) and the ab106533 antibody recognizes the overexpressed CXXC5 in ICC of transfected MCF7 cells, we wanted to test whether we could also detect the endogenous protein in MCF7 cells. To examine this issue, we subjected untransfected MCF7 cells to ICC using the ab106533 antibody (Figure 22). Although the amount of protein was low, as we needed longer exposure times compared to cells transfected with the Flag-*CXXC5*-cDNA bearing expression vectors, we were able to detect CXXC5 in the nuclei of MCF7 cells as well.

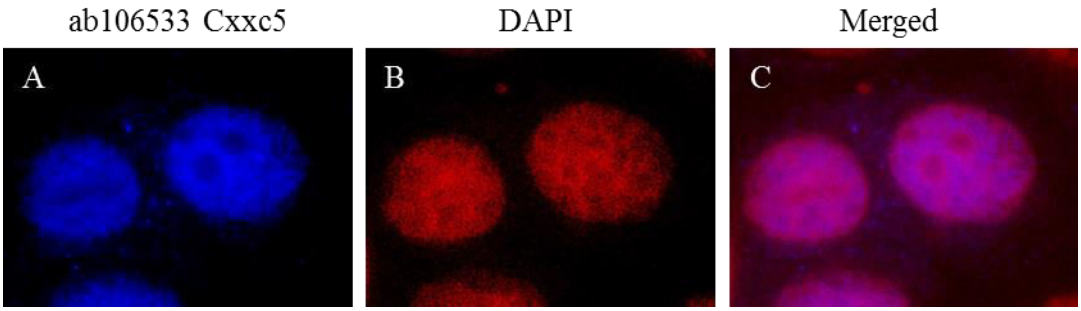


Figure 22: ICC for the endogenous CXXC5 in MCF7 cells. MCF7 cells were grown on coverslips in 12-well cell culture plates in growth medium containing 8% fetal bovine serum. Cells were fixed with 2% paraformaldehyde in PBS and permeabilized with 0.4% Triton-X100 in PBS. 10% normal goat serum (NGS) in PBS was used for blocking for one hour. Cells were then treated with the CXXC5 antibody (ab106533) for two hours (1:250 in PBS with 2% NGS). After the primary antibody incubation, an Alexa Fluor®-488 (green channel) conjugated goat anti-rabbit secondary antibody (1:1000 in PBS with 2% NGS) was applied for 30-minutes. For nuclei staining, DAPI (blue channel) was used. Images were obtained with aforementioned channels; then to emphasize the overlapping regions artificial coloring was used to display the CXXC5 staining in blue *(A)* and DAPI *(B)* staining in red *C)* Merged image of the Flag antibody and DAPI staining. A representative image from two independent experiments is shown.

Based on these results, we conclude that the ab106533 antibody recognizes both exogenously introduced and endogenously synthesized, albeit at low levels, CXXC5 in the nuclei of MCF7 cells by ICC.

During the fluorescent microscopy imaging, we also observed unique distributions of CXXC5 in the nuclei of dividing cells. Unlike the interphase cells which show an even distribution in the nucleus, CXXC5 in dividing cells shows a nuclear location that does not overlap with DNA (Figure 23). This observation was further examined using confocal microscopy in cells at various stages of cell division.

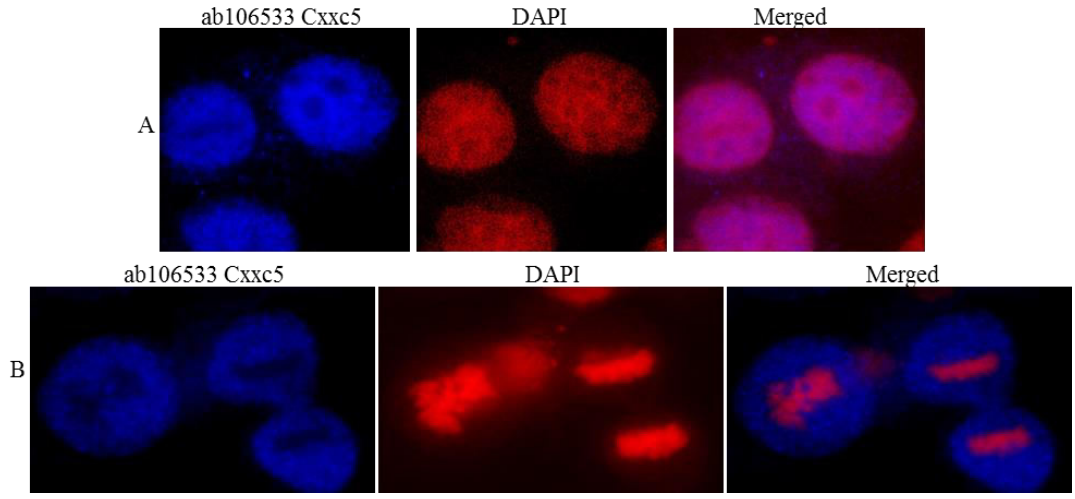


Figure 23: Alterations in the subnuclear localization of the endogenous CXXC5 in dividing cells. MCF7 cells were grown on coverslips in 12-well cell culture plates in growth medium containing 8% fetal bovine serum. Cells were fixed with 2% paraformaldehyde in PBS and permeabilized with 0.4% Triton-X100 in PBS. 10% normal goat serum (NGS) in PBS was used for blocking for one hour. Cells were then treated with the CXXC5 antibody (ab106533) for two hours (1:250 in PBS containing 2% NGS). After the primary antibody incubation, an Alexa Fluor®-488 (green channel) conjugated goat anti-rabbit secondary antibody (1:1000 in PBS with 2% NGS) was applied for 30-minutes. For the nuclei staining, DAPI (blue channel) was used. Images were obtained with aforementioned channels; then to emphasize the overlapping regions, the artificial coloring was used to display CXXC5 staining in blue and DAPI staining in red. A representative image from two independent experiments is shown. *A)* CXXC5 shows an even nuclear distribution in interphase MCF7 cells. *B)* However, in MCF7 cells at metaphase and late-anaphase/telophase, the CXXC5 staining does not overlap with DAPI staining.

3.4.2.1. Confocal Microscopy Studies

To examine the endogenous CXXC5 staining at various stages of cell division, we performed ICC experiments using confocal microscopy. Images were taken with the collaboration of Gizem Güpür of the Muyan laboratory (MSc, Department of Biology, METU, Ankara-Turkey) and Menekşe Ermiş of the Vasıf Hasırcı laboratory (MSc, Department of Biotechnology, METU, Ankara-Turkey). The principle of confocal microscopy is based on the 3D reconstruction. Signals obtained at segments are gathered to construct the image with the help of a software program. ICC was carried out using untransfected MCF7 cells and the ab106533 antibody. The cell population was then scanned to locate a cell at a distinct stage of cell division. Cells in interphase showed a CXXC5 staining that has an even nuclear distribution. On the other hand, CXXC5 at various stages of cell division

encompassing metaphase, late anaphase/telophase and telophase shows a CXXC5 nuclear staining that does not overlap with DNA stained with DRAQ5 (Figure 24).

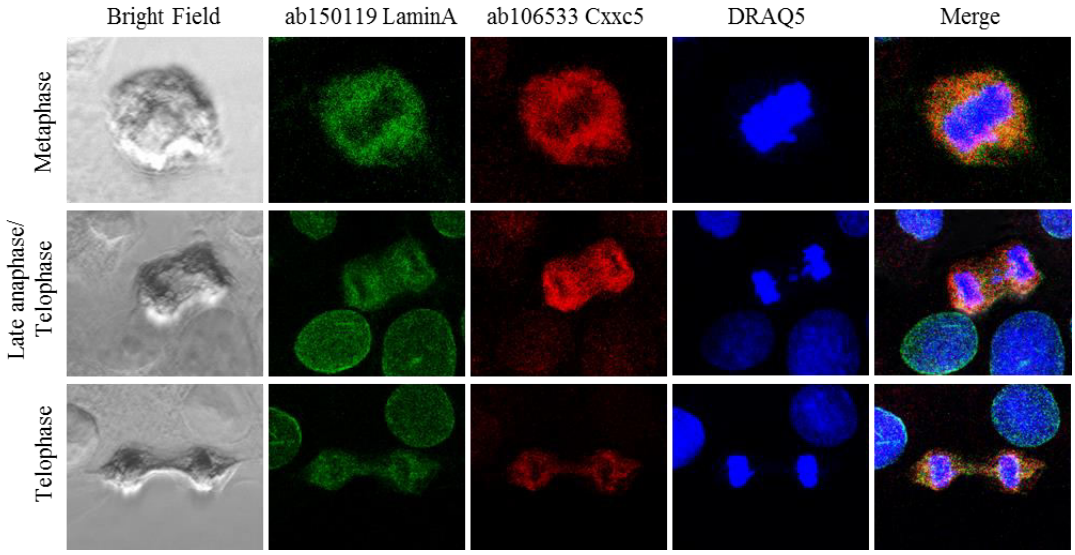


Figure 24: CXXC5 distribution at different stages of cell division in MCF7 cells. MCF7 cells were grown on coverslips in 12-well cell culture plates in growth medium containing fetal bovine serum. Cells were fixed with 2% in PBS paraformaldehyde and permeabilized with 0.4% Triton-X100 in PBS. 10% normal goat serum (NGS) in PBS was used for blocking for one hour. Cells were then treated with the ab106533 antibody for two hours (1:250 in PBS with 2% NGS) and LaminA (1:200 in in PBS with 2% NGS) for another two hours. After the ab106533 antibody incubation, an AlexaFluor® 488 conjugated goat anti-rabbit secondary antibody (ab150077, Abcam) (1:1000 in PBS with 2% NGS) was applied for 30-minutes for the ab106533 antibody. Similarly, 30-minutes incubation was performed with an AlexaFluor® 647 conjugated goat anti-mouse secondary antibody (ab150119, Abcam, anti-LaminA antibody) (1:250 in PBS with 2%NGS) for the LaminA antibody. For the nuclei staining, DRAQ5 (blue) was used.

Z-stack images of the metaphase cell, for example, shown at the first line in Figure 24, were depicted in Figure 25. With the help of Z-stack images, captured cells can also be analyzed at various rotations with 90^{o} angles. In Z-stack imaging, the metaphase cell was observed in different rotational angles and it provides information about signals coming from different sides of the cell. According to Z-stack images of a metaphase cell, the CXXC5 signal does not overlap with the DNA staining.

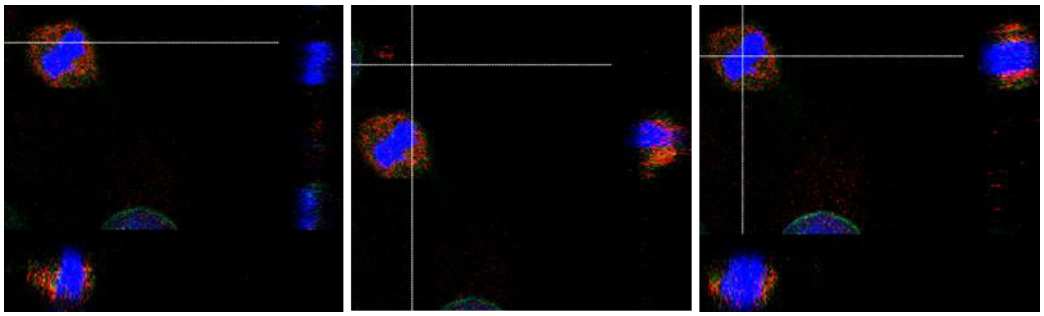


Figure 25: Z-stack images of a metaphasic MCF7 cell. Red staining indicates CXXC5 and DRAQ5 (blue) denotes DNA.

It should also be noted that the CXXC5 staining in interphase was considerably lower than that of dividing cells. In Figure 26, an interphase cell shows low amounts of CXXC5 compare to that of the early metaphase cell. Although it is yet unclear, this raises the possibility that the expression and/or synthesis of CXXC5 is cell-cycle dependent. We are currently exploring this possibility in MCF7 cells subjected to cell cycle synchronization.

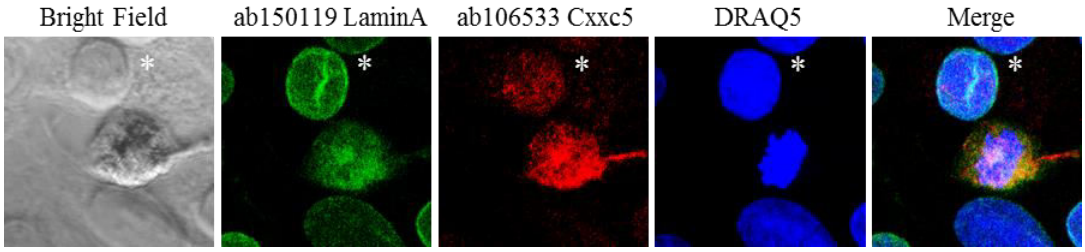


Figure 26: CXXC5 levels in a cell at interphase. Low levels of CXXC5 in cell at interphase (white star) compared to a cell at early metaphase.

3.4.3. Protein Expression in *E. coli*

3.4.3.1. Expression of the 6xHis-Flag-CXXC5 Protein

For a detailed analysis of biochemical and functional features of CXXC5, for which very little is known, we wanted to explore the possibility of generating a recombinant CXXC5 protein. Previous studies have shown that the isolated CXXC domain from some members of the ZF-CXXC domain family proteins including TET3 and IDAX (CXXC4) binds to non-methylated CpG dinucleotides *in vitro* and *in situ* [17], [22]. Our homology modeling also suggests that CXXC5 is a non-methylated CpG dinucleotide binding protein that interacts with DNA through the highly conserved CXXC domain. We therefore wanted to test this prediction as a prelude to the structural-functional properties of CXXC5 using recombinant proteins. To accomplish this, various CXXC5 proteins as depicted in Figure 27 were designed.

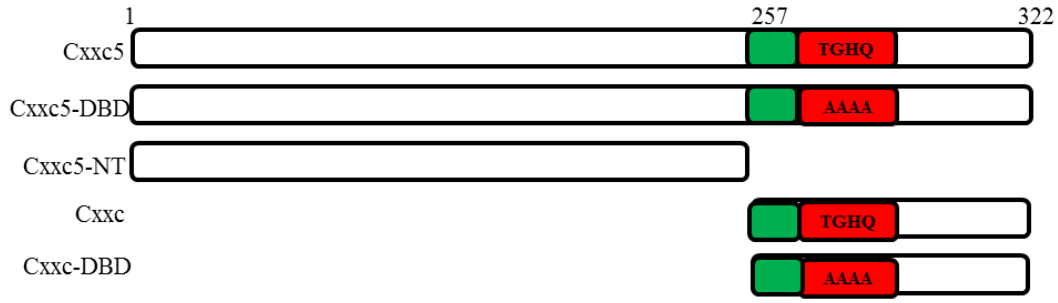


Figure 27: Schematics of truncated and mutated forms of CXXC5.

To generate DNA binding defective mutant of CXXC5 ($CXXC5_{DBD}$), we used an overlapping PCR approach [65]. In this mutant, TGHQ residues, shown to be critical for the DNA binding of IDAX and share a high amino acid identity with the CXXC domain of CXXC5, [22], was changed to alanine residues (AAAA) in the CXXC5-CXXC domain. The structural features of the amino-terminus region of CXXC5 are unkown and our *in silico* analyses fail to assign structural features to it. Nevertheless, our secondary and tertiary predictions suggest that the amino-terminus of CXXC5 may have several α-helical regions that could be important for intra-molecular interactions potentially altering the ability of the protein to bind to

DNA. Based on this, we also generated with PCR a fragment of *CXXC5*-cDNA encoding the amino-terminus domain (residues 1 through 256, CXXC5-NT) using the wild-type *CXXC5*-cDNA as the template. To produce a CXXC domain (257-322) without (CXXC) or with mutations that prevents DNA binding ($CXXC_{DBD}$), we used the wild-type *CXXC5*-cDNA or the $CXXC5_{DBD}$-cDNA as template, respectively. We envisioned that CXXC5 and CXXC, but not the DNA binding defective counterparts ($CXXC5_{DBD}$ and $CXXC_{DBD}$), nor the CXXC5-NT, would bind to DNA. In case of differences in DNA affinities of CXXC5 and CXXC, to be determined by competitive binding with electrophoretic mobility shift assay (EMSA), we would conclude that the amino-terminus of CXXC5 affects the ability of CXXC to bind to DNA through an intramolecular interaction.

Following the verification of amino acid sequence fidelity with nucleotide sequencing of the constructs, each construct was cloned into the pet28a(+) vector to contain an amino-terminally located 6xHis tag for purification followed by the Flag tag for protein detection in collaboration with Gamze Ayaz of the Muyan Laboratory (MSc, Biological Sciences, METU, Ankara-Turkey). pET28(a) bearing none or a cDNA was then transformed into chemically competent BL-21(DE3) *E. coli* cells suitable for transformation and protein expression.

To ensure that we could obtain a soluble CXXC5 protein, we initially set up optimization experiments at various conditions that included different incubation temperatures, induction duration and/or induction temperatures (Table 9). Based on the results, we selected to use 1mM IPTG to induce recombinant protein synthesis in bacterial cells when the OD_{600} of cell culture reached to 0.6 for six hours at $30^{o}C$. Bacterial cells without or with IPTG induction were pelleted. Pellets were resuspended in 200 µl 1X Laemmli buffer containing 5% β-mercaptoethanol and boiled. 20 µl suspensions were subjected to 10% SDS-PAGE followed by Commassie Brilliant Blue staining of the gel. Results revealed that the synthesis of 6xHis-Flag-CXXC5 protein can be induced by 1mM IPTG in bacteria (Figure 28).

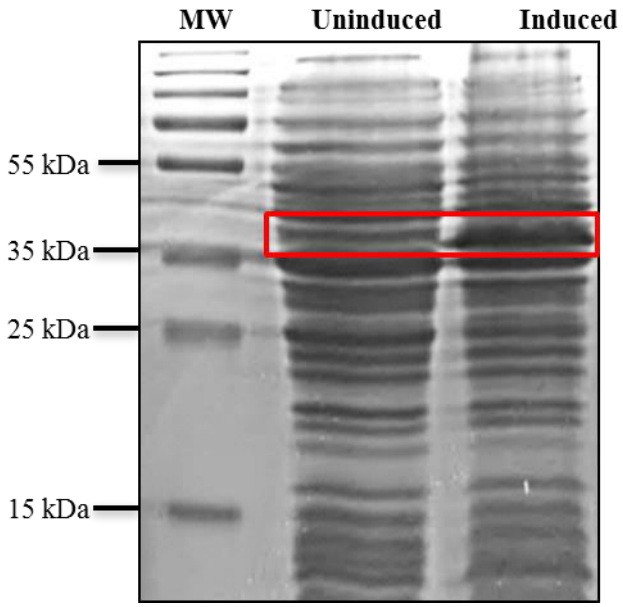


Figure 28: Induction of the 6xHis-Flag-CXXC5 synthesis. The expression of the recombinant 6xHis-Flag-CXXC5 was induced in bacterial cells, when OD_{600} reached to 0.6, by 1mM IPTG for 6h at 30°C. Extracts of un-induced and induced bacterial samples were run on 10% SDS-PAGE and the gel was stained with Commassie Brilliant Blue. Induction in contrast to uninduced sample was shown in red rectangular. MW depicts molecular weight markers.

Once we verified that the recombinant protein can be synthesized upon induction, we then proceeded to examine whether the protein is in the soluble fraction. To accomplish this, bacterial pellets were lysed in a lysis buffer after sonication and centrifuged. Pellet and supernatant fractions were collected and subjected to 10% SDS-PAGE followed by Commassie Brilliant Blue staining of the gel (Figure 29). We found that although the majority of 6xHis-Flag-CXXC5 was in the pellet, a significant portion of the synthesized protein was also observable in the supernatant.

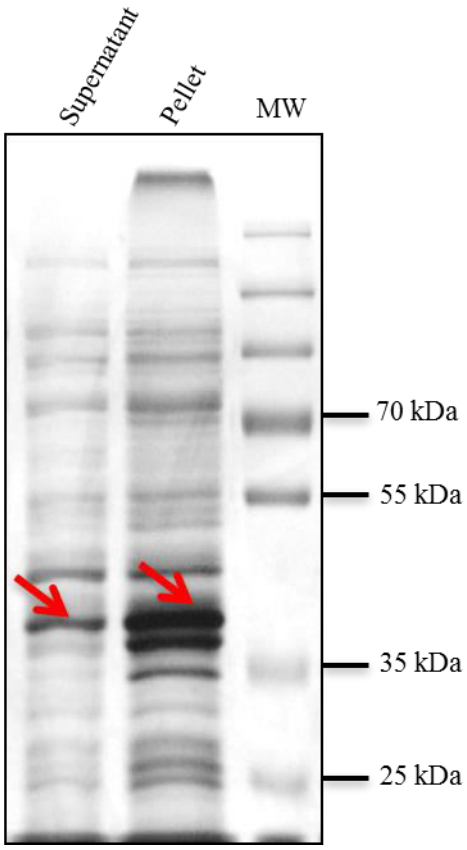


Figure 29: 6xHis-Flag-CXXC5 in soluble fraction. Bacterial pellet was lysed with 630 µl lysis buffer supplemented with 70 µl lysozyme (10 mg/ml) in PBS followed by sonication. Lysed cells were centrifuged and the supernatant was collected. A small amount (20 µl) of the pellet or the supernatant fraction was subjected to 10% SDS-PAGE. Shown is the Commassie Brilliant Blue staining of the gel. MW depicts molecular weight markers.

3.4.3.2. Purification of the recombinant 6xHis-Flag-CXXC5 Protein

After the verification that 6xHis-Flag-CXXC5 is detectable in the supernatant fraction, the supernatant of an IPTG-induced bacterial sample was subjected to Ni-NTA Spin columns (Qiagen, Germany). Proteins bound to the column were eluted three successive times with an elution buffer after washing the column four times with a wash buffer (Table 10 in Section 2.5.4.2). Protein contents at various stages of column purification were analyzed on 10% SDS-PAGE followed by Commassie Brilliant Blue staining of the gel (Figure 30). Eluates revealed single protein band that shows an electrophoretic migration approximating 37 kDa.

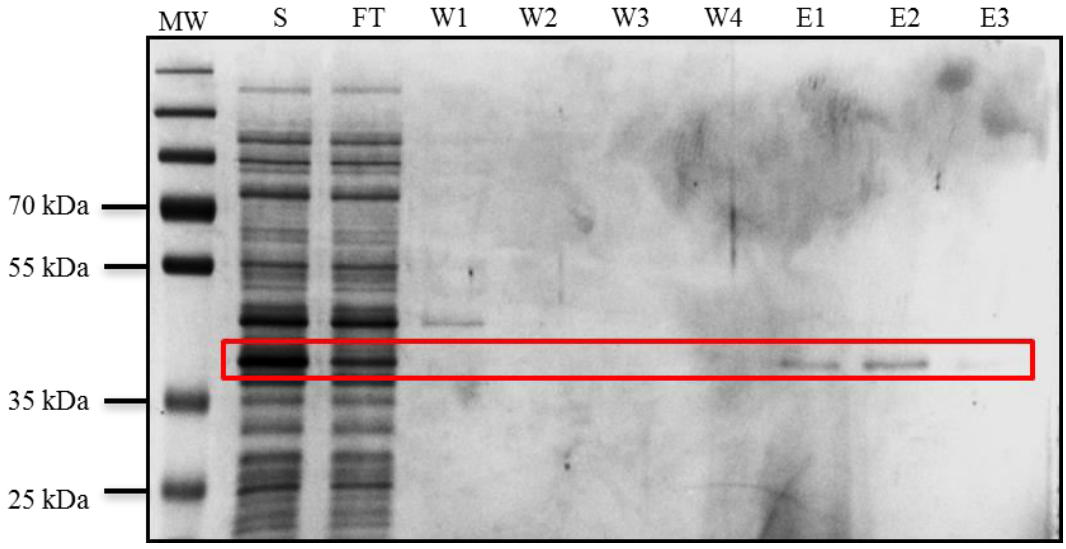


Figure 30: Column purification of the 6xHis-Flag-CXXC5 protein. Five ml of bacterial culture was used for the purification of CXXC5 protein under native conditions using the Ni-NTA column system. Red rectangle indicates the recombinant 6xHis-Flag-CXXC5. S: Supernatant; FT: Flow through (after the column binding); W1, W2, W3 and W4 indicate washes; E1, E2 and E3 denote elutions. MW: Molecular weight marker.

To ensure that the column purified recombinant protein is CXXC5, fractions of column purification steps were subjected to WB analysis using the ab106533 antibody (Figure 31). Results indicate that eluates contain the recombinant 6xHis-Flag-CXXC5 protein.

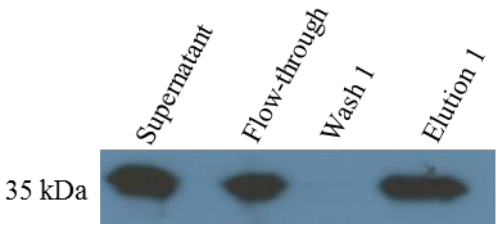


Figure 31: WB for the column purified 6xHis-Flag-CXXC5. Western blot analysis of the Ni-NTA column purified recombinant CXXC5 protein. 2.5 µg protein fractions from the column purification scheme were subjected 10% SDS-PAGE followed by WB analysis using the ab106533 antibody (1:10000).

3.4.4. EMSA (Electrophoretic Mobility Shift Assay) Studies

3.4.4.1. Controls for EMSA

To examine the ability of the 6xHis-fusion Flag-tagged CXXC5 to interact with non-methylated CpG containing DNA, we used EMSA. Biotin-labeled DNAs consist of none, CpG or methylated CpG dinucleotide embedded within a larger oligomer with no CpG features [12] depicted in Figure 32.

CpG DNA	5'- CTGTCAGTAT CCGG TACAGGTCTA -3'
	3'- GACAGTCATA GGCC ATGTCCAGAT -5'
Methyl-CpG DNA	5'- CTGTCAGTAT CmCGG TACAGGTCTA -3'
	3'- GACAGTCATA GmGCC ATGTCCAGAT -5'
Non-CpG DNA	5'- CTGTCAGTAT CAGT TACAGGTCTA -3'
	3'- GACAGTCATA GTCA ATGTCCAGAT -5'

Figure 32: DNA sequences of CpG, methylated CpG and non-CpG oligonucleotides that are labeled with biotin at their 5' ends.

We used the LightShift® Chemiluminescent EMSA System (Thermo, Scientific, CA, USA) for EMSA experiments. To ensure that the system is functional, we initially assessed the DNA binding of the recombinant EBNA (Epstein–Barr virus nuclear antigen) protein provided with the system. The biotin-labeled double-stranded DNA fragment containing sequences for EBNA (EBNA-DNA, 60 bp) were co-incubated with EBNA in EMSA binding buffer for 30-minute. Reactions were loaded onto a native 5% PAGE for electrophoresis. The unlabeled EBNA-DNA (60 bp) was used as the competitor. The biotin-labeled EBNA-DNA showed a retardation in electrophoretic migration when EBNA was present and the retarded complex was lost in the presence of the competitor unlabeled EBNA-DNA, as expected (Figure 33).

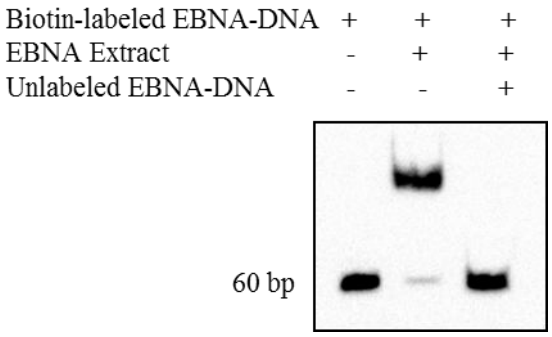


Figure 33: EMSA for EBNA-DNA interactions. The binding of EBNA to the biotin-labeled EBNA-DNA was used to assess the functionality of the LightShift® Chemiluminescent EMSA System. Reactions were set up according to the manufacturer's instructions. The chemiluminescent Nucleic Acid Detection Module (Thermo Scientific, CA, USA) was used for the detection of the biotin-labeled DNA.

We also used ERα to ensure that intracellular proteins under the laboratory conditions could also be used for EMSA. To accomplish this, we transiently transfected ERα-negative MDA-MB-231 cells with pcDNA3.1(-) bearing none or the ERα cDNA. 36h after transfections, whole cell extracts (WCEs, in Section 2.5.4.3) were obtained. WCEs at various concentrations and biotin-labeled double stranded DNA fragment bearing the consensus ERE (40 bp) (in Section 2.5.5.1 in Table 14) were incubated in EMSA binding buffer with or without the Flag-M2 antibody to confirm the specificity of protein-DNA interactions. Samples were then subjected to electrophoresis. Results revealed that ERα at increasing amounts binds to ERE as reflected in the retardation of the electrophoretic mobility of the biotin-labeled DNA bearing ERE (Figure 34). That the inclusion of the Flag-M2 antibody (lane 5) further retards the electrophoretic mobility of the protein-DNA complex verifies that the ERE bound protein is Flag-ERα.

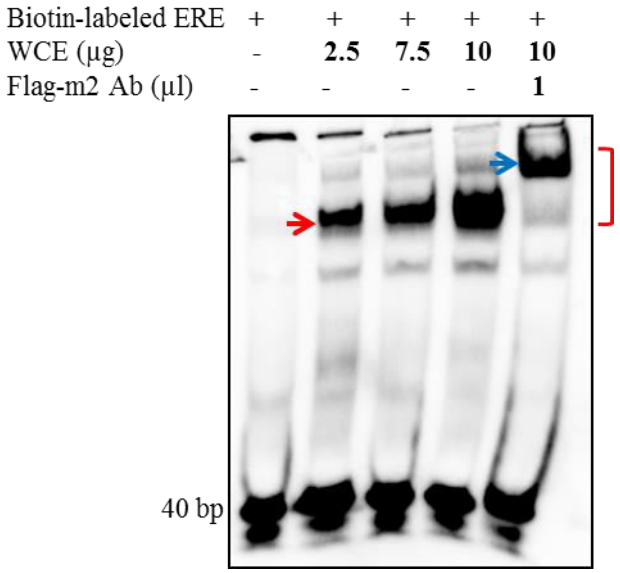


Figure 34: EMSA for ERα-ERE interactions. WCEs of Flag-ERα synthesizing MDA-MB-231 cells and double stranded biotin-labeled DNA fragment bearing the consensus ERE were subjected to EMSA. The red arrow indicates the Flag ERα-ERE complex and the blue arrow denotes the supershifted Flag ERα-ERE complex due to the presence of the Flag-M2 antibody.

3.4.4.2. EMSA for CXXC5

To assess the DNA binding ability of CXXC5, we used EMSA with increasing concentrations of WCEs of MDA-MB-231 cells transfected with pcDNA3.1(-)-Flag-*CXXC5* cDNA together with the 5'-biotinylated and non-methylated CpG dinucleotide containing DNA fragment (Figure 35). Results revealed various electrophoretically retarded DNA bands suggesting the presence of many intracellular proteins capable of binding to DNA. There was also a further retardation of the electrophoretic migration of the protein-DNA bands when the Flag-M2 antibody was present. Although this implies the presence of Flag-CXXC5 in reactions, the similar pattern of EMSA results with WCEs of MDA-MB-231 cells transfected with the pcDNA3.1(-) vector bearing no cDNA together with the 5'-biotinylated and non-methylated CpG dinucleotide containing DNA (Figure 36) points to the difficulty of assessing/deciphering CXXC5-DNA interactions. This was also the case when we used the unlabeled non-methylated CpG (as competitor

DNA), methylated CpG or non-CpG containing DNA in the absence or presence of the ab106533 antibody (Figure 37).

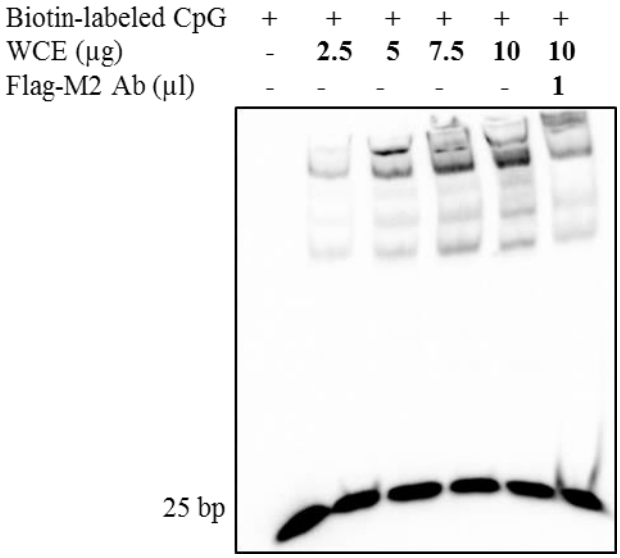


Figure 35: EMSA for the overexpressed CXXC5 protein. Increasing amount of WCEs obtained from cells transfected with pcDNA3.1(-) bearing Flag-*CXXC5* cDNA and the double stranded biotin-labeled DNA fragment bearing the non-methylated CpG were subjected to the EMSA. 1µl Flag-M2 antibody was used for specifying the shift.

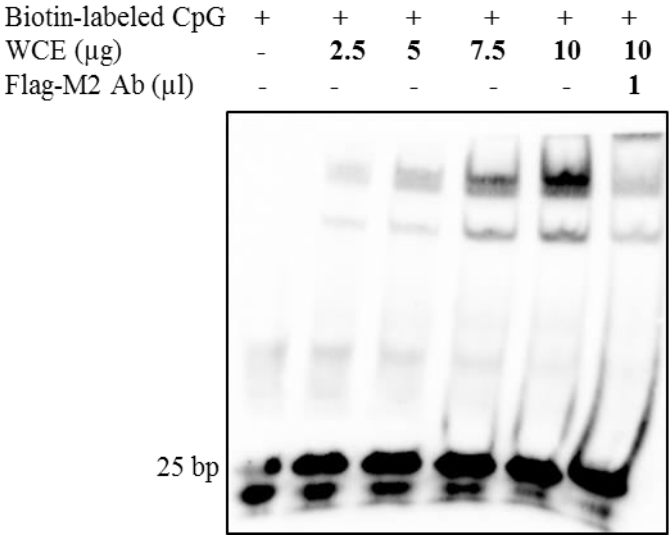


Figure 36: EMSA for the endogenous CXXC5 protein. Increasing amount of WCEs obtained from MDA-MB-231 cells transfected with pcDNA3.1(-) bearing no cDNA and double stranded biotin-labeled DNA fragment bearing the non-methylated CpG were subjected to the EMSA. 1µl Flag-M2 antibody was used for supershift.

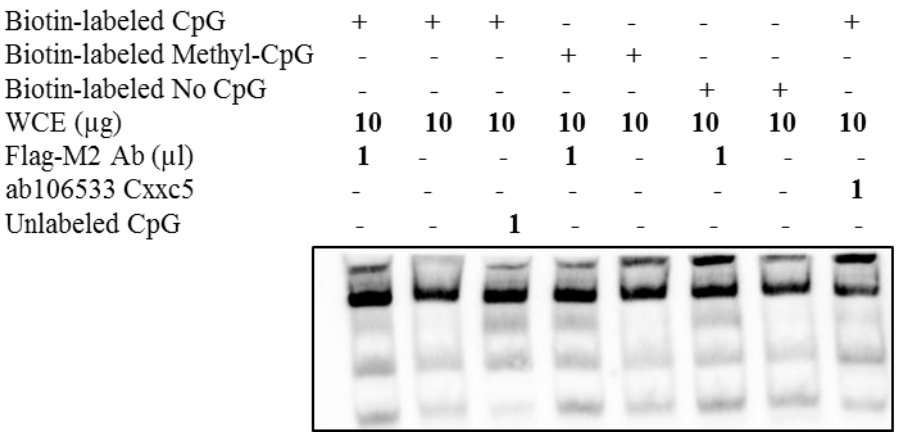


Figure 37: EMSA of the overexpressed CXXC5 together with the controls. 10 µg WCE of MDA-MB-231 cells transfected with pcDNA(-)Flag-*CXXC5* cDNA and the double stranded biotin-labeled DNA fragment bearing the non-methylated CpG, methylated CpG or non-CpG were subjected to EMSA. For each DNA set, 1 µl Flag-M2 antibody was used for supershift. For the non-methylated CpG DNA, unlaballed (cold competitor) DNA and ab106533 antibody were also used.

To assess the existence of CXXC5 complexed with DNA and/or decipher the CXXC5-DNA complex from other protein-DNA bands, we performed WB analysis using the Flag-M2 antibody. However, we could not detect Flag-CXXC5. The difficulty of using WCEs in assessing the ability of CXXC5 to interact with DNA necessitated the use of the 6xHis column purified recombinant full length CXXC5 for EMSA. We again failed to observe a DNA binding at any condition that we have tested (data not shown). Although the underlying reason(s) is unclear, our inability to detect the binding of CXXC5 to DNA could stem from the bacterial expression system we used which may not be suitable for the folding, or the correct folding, of the protein.

A recent study showed that the bacterially produced recombinant CXXC domain of IDAX, CXXC4, which share a high amino-acid identity (~ 87%) with that of CXXC5, fused to the carboxyl-terminus of GST (Glutathione S-transferase) interacts with a non-methylated CpG dinucleotide containing DNA *in vitro* [22]. It is therefore possible that the isolated CXXC domain alone or as a fusion with GST rather than the full-length protein produced in bacteria binds to DNA. To test this possibility, we are in the process of producing recombinant CXXC domain of both

CXXC5 and IDAX without or with GST fusion. Even if we were to observe a DNA binding with CXXC domains, this only would imply that the isolated CXXC domain, as predicted, has the ability to interact with CpG dinucleotides but not the full length protein which may require *in situ* constitutive or signal-mediated conformational changes, for example protein processing, acetylation and/or phosphorylation, to expose and allow CXXC domain to interact with DNA. To address this issue, we are currently testing yeast and insect cell expression approaches that have eukaryotic protein processing capabilities associated with mammalian systems to produce a recombinant CXXC5. We are also planning to examine interactions of CXXC5 *in situ* with DNA using Fluorescence recovery after photobleaching (FRAP) approach described previously [7].

Alternatively, CXXC5 in contrast to our expectation is not a DNA binding protein. The scarcity of information about the structural-functional features of CXXC5 has been limiting our ability to test whether the recombinant protein has a function. Yet, we have in the process of exploring the putative interacting protein partners of CXXC5 using yeast two-hybrid and mammalian proximity biotinylation approaches. Findings could allow us to make better predictions and to design appropriate experimental systems towards the delineation of the structural-functional features of CXXC5.

CHAPTER 4

CONCLUSION AND FUTURE DIRECTIONS

Our predictions using *in silico* approaches suggest that:

1. Although the amino-terminus of CXXC5 shows no significant homology with any known mammalian proteins, the carboxyl-terminus contains a highly conserved CXXC domain present in the ZF-CXXC domain protein family, through which the family proteins bind to non-methylated CpG dinucleotides, and a nuclear localization signal. Our tertiary structure and homology studies therefore predict that CXXC5 localized in the nucleus is a non-methylated CpG dinucleotide binding protein.

2. Our observation is that *CXXC5* is an E2 responsive gene regulated by ERα through the ERE-dependent signaling pathway predicts that the *CXXC5* gene locus contains ERE-like sequences. Using Web-based sequence prediction tools, we located seven putative ERE sequences at the upstream and downstream of the translation initiation and termination sequences of the gene, respectively.

Our experimental results reveal and we suggest that:

1. The commercially available ab106533 CXXC5 antibody recognizes the overexpressed wild-type and Flag-tagged CXXC5 under denaturing conditions (WB) in primarily nuclear extracts of cell lines derived from breast carcinomas. In addition, the ab106533 CXXC5 antibody recognizes both the endogenously synthesized CXXC5 and exogenously introduced Flag-CXXC5 under native conditions (ICC) in the nuclei of MCF7 cells.

    a) Although the ab106533 CXXC5 antibody recognizes the overexpressed CXXC5 in WBs of transfected MCF7 cells, the

antibody detects several non-specific proteins with varying molecular masses as well. By using *CXXC5* specific siRNAs, we were able to conclusively verify and decipher the presence of the CXXC5 protein in WB. We also detected CXXC5 in the nuclei of untransfected MCF7 cells by ICC similar to those of transfected cells. Since the ab106533 CXXC5 antibody recognizes several proteins with different molecular masses in WB, it is imperative to prove that ICC detects the endogenous CXXC5 with the use of our siRNA approach as well.

b) It appears that CXXC5 is synthesized at low levels in MCF7 cells. Extending the screening of cell lines, as we performed with some triple negative and ER-positive cell lines, will allow the selection of cell models with varying levels (high or low) of CXXC5 that could be important for the functional analysis of CXXC5 through knock-down or knocking-in studies.

2. We observed that the overexpressed wild-type CXXC5 shows an electrophoretic migration slower than that of the endogenous CXXC5 in WB analyses. Pointing to a higher molecular mass of the overexpressed wild-type CXXC5, this could be due to a protein processing of the endogenous CXXC5 that are circumvented/prevented with the overexpressed protein. Alternatively, or in addition to, aberrant post-translational modifications (PTM) of the overexpressed CXXC5 which is not observed with the endogenous protein could contribute to the higher molecular mass of the overexpressed CXXC5. Protein processing of the endogenous CXXC5 could be comparatively assessed by protein sequencing following immunoprecipitation. Based on the PTM predictions, CXXC5 has 17 putative phosphorylation and 2 N-terminally acetylation sites. The phosphorylation and acetylation status of the overexpressed CXXC5 can be examined by preventing phosphorylation and/or acetylation to reveal whether differences in molecular masses are due to aberrant PTMs of the overexpressed protein.

3. Although the CXXC5 staining in the nuclei of MCF7 cells was observed to overlap with DNA at interphase, the CXXC5 showed a distinct subnuclear localization at metaphase, late anaphase/telophase and telophase that did not overlap with DNA assessed by confocal microscopy. Moreover, it appeared that nuclear levels of CXXC5 are augmented, reflected as more intense staining, in dividing cells compared to levels observed in interphase cells. These findings suggest that the synthesis, degradation and/or subnuclear localization of CXXC5 is cell-cycle dependent.

a) To examine the possible variations in protein levels of CXXC5 at various cell cycle stages, ICC and WB experiments can be performed using synchronized cells.

b) To analyze the changes in the subcellular localization of CXXC5 according to cell-cycle phases, CXXC5 can be traced by adding fluorescent tag, for example green fluorescent protein (GFP), by genome editing and using time lapse microscopy.

4. Homology modeling studies suggest that CXXC5 is a non-methylated CpG binding protein due to the presence of the CXXC domain. The use of the overexpressed CXXC5 protein in WCEs of transfected cells failed to decipher CXXC5-DNA interactions in EMSA. We also failed to observe the binding of the recombinant protein obtained from a bacterial expression system to various DNA. These could be due to:

a) WCEs of the CXXC5 overexpressing cells contain many other CpG binding proteins. This renders the interpretation of EMSA results difficult.

b) To circumvent this possible problem, we used the bacterially produced recombinant CXXC5. However we observed no CpG dinucleotide binding. It is possible that the recombinant CXXC5 protein may not be folded or folded correctly. To ensure that this is or is not the case, a protein, for example IDAX (CXXC4) which

shows a high amino acid identity to CXXC5 and was previously shown to bind to CpG dinucleotides *in vitro*, could be used as positive control. Since bacterial systems do largely not have folding machinery mimicking eukaryotic systems, yeast or insect cell protein expression approaches with eukaryotic protein processing capabilities similar to mammalian systems could be used to produce a recombinant CXXC5. This could be followed with Atomic Force Microscopy (ATM) and/or crystallography studies that allow us to generate various models for structural and functional analysis of CXXC5.

c) It is also possible that CXXC5 requires a specific PTM that alters conformation of the protein to expose or allow the CXXC domain to bind to CpG dinucleotides in genome. Examination of DNA-CXXC5 interactions *in situ* using fluorescence recovery after photobleacing (FRAP) approach could be carried out.

d) On the other hand, CXXC5 may not be a DNA binding protein in contrast to our predictions. CXXC5 could act as protein partners for transcription factors and/or signaling cascade proteins, as some studies suggest. To address these issues, we are in the process of exploring the putative CXXC5 partner proteins by using yeast two-hybrid and also mammalian proximity biotinylation systems.

5. Since the *CXXC5* locus encompasses several ERE like sequences as an indication of an E2-ERα responsive gene regulated through the ERE-dependent signaling pathway, the identification and verification of the regulatory ERE sequence(s) at the gene locus should be carried out using Chromatin ImmunoPrecipitation (ChIP) assay followed by EMSA and luciferase reporter assays.

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] K. Polyak, "Science in medicine Breast cancer : origins and evolution," *J. Clin Invest*, vol. 117, no. 11, pp. 3155–3163, 2007.

[3] 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, no. 11, pp. 2696–712, Nov. 2005.

[4] J. S. Lewis-Wambi and V. C. Jordan, "Estrogen regulation of apoptosis: how can one hormone stimulate and inhibit?," *Breast Cancer Res.*, vol. 11, no. 3, p. 206, Jan. 2009.

[5] 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.

[6] C. Klinge, "Estrogen receptor interaction with estrogen response elements," *Nucleic Acids Res.*, vol. 29, no. 14, pp. 2905–2919, 2001.

[7] 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.," *J. Mol. Endocrinol.*, vol. 49, no. 3, pp. 249–66, Dec. 2012.

[8] J. M. Hall, J. F. Couse, and K. S. Korach, "The multifaceted mechanisms of estradiol and estrogen receptor signaling.," *J. Biol. Chem.*, vol. 276, no. 40, pp. 36869–72, Oct. 2001.

[9] Q. Zhang, M. Ye, X. Wu, S. Ren, M. Zhao, C. Zhao, G. Fu, Y. Shen, H. Fan, G. Lu, M. Zhong, X. Xu, Z. Han, J. Zhang, J. Tao, and Q. Huang, "Cloning and Functional Analysis of cDNAs with Open Reading Frames for 300 Previously Undefined Genes Expressed in CD34 + Cloning and Functional Analysis of cDNAs with Open Reading Frames for 300 Previously Undefined Genes Expressed in CD34 + Hematopoietic," *Genome Res.*, vol. 10, pp. 1546–1560, 2000.

[10] a Renneville, C. Roumier, V. Biggio, O. Nibourel, N. Boissel, P. Fenaux, and C. Preudhomme, "Cooperating gene mutations in acute myeloid leukemia: a review of the literature.," *Leukemia*, vol. 22, no. 5, pp. 915–31, May 2008.

[11] C. Meyer, E. Kowarz, J. Hofmann, a Renneville, J. Zuna, J. Trka, R. Ben Abdelali, E. Macintyre, E. De Braekeleer, M. De Braekeleer, E. Delabesse, M. P. de Oliveira, H. Cavé, E. Clappier, J. J. M. van Dongen, B. V Balgobind, M. M. van den Heuvel-Eibrink, H. B. Beverloo, R. Panzer-Grümayer, a Teigler-Schlegel, J. Harbott, E. Kjeldsen, S. Schnittger, U. Koehl, B. Gruhn, O. Heidenreich, L. C. Chan, S. F. Yip, M. Krzywinski, C. Eckert, a Möricke, M. Schrappe, C. N. Alonso, B. W. Schäfer, J. Krauter, D. a Lee, U. Zur Stadt, G. Te Kronnie, R. Sutton, S. Izraeli, L. Trakhtenbrot, L. Lo Nigro, G. Tsaur, L. Fechina, T. Szczepanski, S. Strehl, D. Ilencikova, M. Molkentin, T. Burmeister, T. Dingermann, T. Klingebiel, and R. Marschalek, "New insights to the MLL recombinome of acute leukemias.," *Leukemia*, vol. 23, no. 8, pp. 1490–9, Aug. 2009.

[12] M. D. Allen, C. G. Grummitt, C. Hilcenko, S. Y. Min, L. M. Tonkin, C. M. Johnson, S. M. Freund, M. Bycroft, and A. J. Warren, "Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase.," *EMBO J.*, vol. 25, no. 19, pp. 4503–12, Oct. 2006.

[13] C. Xu, C. Bian, R. Lam, A. Dong, and J. Min, "The structural basis for selective binding of non-methylated CpG islands by the CFP1 CXXC domain.," *Nat. Commun.*, vol. 2, p. 227, Jan. 2011.

[14] T. Cierpicki, L. E. Risner, J. Grembecka, S. M. Lukasik, R. Popovic, M. Omonkowska, D. S. Shultis, N. J. Zeleznik-le, and H. Bushweller, "NIH Public Access," vol. 17, no. 1, pp. 62–68, 2010.

[15] H. K. Long, N. P. Blackledge, and R. J. Klose, "ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection.," *Biochem. Soc. Trans.*, vol. 41, no. 3, pp. 727–40, Jun. 2013.

[16] T.-P. Gu, F. Guo, H. Yang, H.-P. Wu, G.-F. Xu, W. Liu, Z.-G. Xie, L. Shi, X. He, S. Jin, K. Iqbal, Y. G. Shi, Z. Deng, P. E. Szabó, G. P. Pfeifer, J. Li, and G.-L. Xu, "The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes.," *Nature*, vol. 477, no. 7366, pp. 606–10, Sep. 2011.

[17] Y. Xu, C. Xu, A. Kato, W. Tempel, J. G. Abreu, C. Bian, Y. Hu, D. Hu, B. Zhao, T. Cerovina, J. Diao, F. Wu, H. H. He, Q. Cui, E. Clark, C. Ma, A. Barbara, G. J. C. Veenstra, G. Xu, U. B. Kaiser, X. S. Liu, S. P. Sugrue, X. He, J. Min, Y. Kato, and Y. G. Shi, "Tet3 CXXC domain and dioxygenase

activity cooperatively regulate key genes for Xenopus eye and neural development.," *Cell*, vol. 151, no. 6, pp. 1200–13, Dec. 2012.

[18] L. Core, J. Waterfall, and J. Lis, "Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters," *Science*, vol. 322, pp. 1845–1848, 2008.

[19] S. Cross and A. Bird, "CpG islands and genes," *Curr. Opin. Genet. Dev.*, vol. 5, pp. 309–314, 1995.

[20] S. Knappskog, L. M. Myklebust, C. Busch, T. Aloysius, J. E. Varhaug, P. E. Lønning, J. R. Lillehaug, and F. Pendino, "RINF (CXXC5) is overexpressed in solid tumors and is an unfavorable prognostic factor in breast cancer.," *Ann. Oncol.*, vol. 22, no. 10, pp. 2208–15, Oct. 2011.

[21] F. Pendino, E. Nguyen, I. Jonassen, B. Dysvik, A. Azouz, M. Lanotte, E. Ségal-Bendirdjian, and J. R. Lillehaug, "Functional involvement of RINF, retinoid-inducible nuclear factor (CXXC5), in normal and tumoral human myelopoiesis.," *Blood*, vol. 113, no. 14, pp. 3172–81, Apr. 2009.

[22] M. Ko, J. An, H. S. Bandukwala, L. Chavez, T. Aijö, W. a Pastor, M. F. Segal, H. Li, K. P. Koh, H. Lähdesmäki, P. G. Hogan, L. Aravind, and A. Rao, "Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX.," *Nature*, vol. 497, no. 7447, pp. 122–6, May 2013.

[23] L. M. Iyer, M. Tahiliani, A. Rao, and L. Aravind, "Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids," *Cell Cycle*, vol. 8, no. 11, pp. 1698–1710, Oct. 2009.

[24] O. Bernatik, R. S. Ganji, J. P. Dijksterhuis, P. Konik, I. Cervenka, T. Polonio, P. Krejci, G. Schulte, and V. Bryja, "Sequential activation and inactivation of Dishevelled in the Wnt/beta-catenin pathway by casein kinases.," *J. Biol. Chem.*, vol. 286, no. 12, pp. 10396–410, Mar. 2011.

[25] M. S. Kim, S. K. Yoon, F. Bollig, J. Kitagaki, W. Hur, N. J. Whye, Y.-P. Wu, M. N. Rivera, J. Y. Park, H.-S. Kim, K. Malik, D. W. Bell, C. Englert, A. O. Perantoni, and S. B. Lee, "A novel Wilms tumor 1 (WT1) target gene negatively regulates the WNT signaling pathway.," *J. Biol. Chem.*, vol. 285, no. 19, pp. 14585–93, May 2010.

[26] A. Astori, H. Fredly, T. Aloysius, and L. Bullinger, "CXXC5 (Retinoid-Inducible Nuclear Factor, RINF) is a Potential Therapeutic Target in High-Risk Human Acute Myeloid Leukemia.," *Oncotarget*, vol. 4, no. 9, pp. 1438–1448, 2013.

[27] H.-Y. Kim, D.-H. Yang, S.-W. Shin, M.-Y. Kim, J.-H. Yoon, S. Kim, H.-C. Park, D. W. Kang, D. Min, M.-W. Hur, and K.-Y. Choi, "CXXC5 is a transcriptional activator of Flk-1 and mediates bone morphogenic protein-induced endothelial cell differentiation and vessel formation.," *FASEB J.*, pp. fj.13–236216–, Oct. 2013.

[28] Z. Min, W. Ruipeng, W. Yanyi, D. Feici, L. U. Fei, G. A. O. Dong, C. Danying, Z. Zhonghe, and S. H. U. Hongbing, "The CXXC finger 5 protein is required for DNA damage-induced p53 activation," vol. 52, no. 2006, pp. 528–538, 2009.

[29] Y. Shiloh and Y. Ziv, "The ATM protein: the importance of being active.," *J. Cell Biol.*, vol. 198, no. 3, pp. 273–5, Aug. 2012.

[30] S. Banin, L. Moyal, S. Shieh, Y. Taya, C. Anderson, L. Chessa, N. Smorodinsky, C. Prives, Y. Reiss, Y. Shiloh, and Y. Ziv, "Enhanced phosphorylation of p53 by ATM in response to DNA damage," *Science (80-.).*, vol. 281, no. September, pp. 1674–1678, 1998.

[31] C. E. Canman, D. Lim, K. A. Cimprich, Y. Taya, K. Tamai, and K. Sakaguchi, "Activation of the ATM Kinase by Ionizing Radiation and Phosphorylation of p53," *Science.*, vol. 281, pp. 1996–1999, 1998.

[32] X. Wang, P. Liao, X. Fan, Y. Wan, Y. Wang, Y. Li, Z. Jiang, X. Ye, and X. Mo, "CXXC5 Associates with Smads to Mediate TNF- Induced Apoptosis," *Current Molecular Medicine*., pp. 1–12, 2013.

[33] S. Fulda and K.-M. Debatin, "Extrinsic versus intrinsic apoptosis pathways in anticancer chemotherapy.," *Oncogene*, vol. 25, no. 34, pp. 4798–811, Aug. 2006.

[34] A. Ashkenazi, "Targeting the extrinsic apoptosis pathway in cancer.," *Cytokine Growth Factor Rev.*, vol. 19, no. 3–4, pp. 325–31, 2008.

[35] J. Massagué, "TGF-β signal transduction," *Annu. Rev. Biochem.*, vol. 67, pp. 753–791, 1998.

[36] J. Massagué, J. Seoane, and D. Wotton, "Smad transcription factors," *Genes Dev.*, vol. 19, pp. 2783–2810, 2005.

[37] B. Schmierer and C. Hill, "Kinetic analysis of Smad nucleocytoplasmic shuttling reveals a mechanism for transforming growth factor β-dependent nuclear accumulation of Smads," *Mol. Cell. Biol.*, vol. 25, no. 22, pp. 9845–9858, 2005.

[38] S. Aras, O. Pak, N. Sommer, R. Finley, M. Hüttemann, N. Weissmann, and L. I. Grossman, "Oxygen-dependent expression of cytochrome c oxidase

subunit 4-2 gene expression is mediated by transcription factors RBPJ, CXXC5 and CHCHD2.," *Nucleic Acids Res.*, vol. 41, no. 4, pp. 2255–66, Feb. 2013.

[39] R. Fukuda, H. Zhang, J. Kim, L. Shimoda, C. V Dang, and G. L. Semenza, "HIF-1 regulates cytochrome oxidase subunits to optimize efficiency of respiration in hypoxic cells.," *Cell*, vol. 129, no. 1, pp. 111–22, Apr. 2007.

[40] M. B. Treppendahl, L. Möllgård, E. Hellström-Lindberg, P. Cloos, and K. Grønbaek, "Downregulation but lack of promoter hypermethylation or somatic mutations of the potential tumor suppressor CXXC5 in MDS and AML with deletion 5q.," *Eur. J. Haematol.*, vol. 90, no. 3, pp. 259–60, Mar. 2013.

[41] S. K. Horrigan, Z. H. Arbieva, H. Y. Xie, J. Kravarusic, N. C. Fulton, H. Naik, T. T. Le, and C. A. Westbrook, "Delineation of a minimal interval and identification of 9 candidates for a tumor suppressor gene in malignant myeloid disorders on 5q31," *Blood*.,vol. 95, no. 7, pp. 2372–2378, 2000.

[42] T. X. Liu, M. W. Becker, J. Jelinek, W.-S. Wu, M. Deng, N. Mikhalkevich, K. Hsu, C. D. Bloomfield, R. M. Stone, D. J. DeAngelo, I. a Galinsky, J.-P. Issa, M. F. Clarke, and a T. Look, "Chromosome 5q deletion and epigenetic suppression of the gene encoding α-catenin (CTNNA1) in myeloid cell transformation," *Nat. Med.*, vol. 13, no. 1, pp. 78–83, Dec. 2006.

[43] P. Mhawech and A. Saleem, "Myelodysplastic syndrome: review of the cytogenetic and molecular data," *Crit. Rev. Oncol. Hematol.*, vol. 40, pp. 229–238, 2001.

[44] M. S. Tallman and J. K. Altman, "Curative Strategies in Acute Promyelocytic Leukemia," pp. 391–399, 2008.

[45] E. Gasteiger and C. Hoogland, "Protein identification and analysis tools on the ExPASy server," *The Proteomics Protocols Handbook*, pp. 571–608, 2005.

[46] R. D. Finn, A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E. L. L. Sonnhammer, J. Tate, and M. Punta, "Pfam: the protein families database.," *Nucleic Acids Res.*, vol. 42, no. Database issue, pp. D222–30, Jan. 2014.

[47] 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.

[48] N. Blom, S. Gammeltoft, and S. Brunak, “Sequence and structure-based prediction of eukaryotic protein phosphorylation sites,” *J. Mol. Biol.*, vol. 294, pp. 1351–1362, 1999.

[49] L. Kiemer, J. D. Bendtsen, and N. Blom, “NetAcet: prediction of N-terminal acetylation sites.,” *Bioinformatics*, vol. 21, no. 7, pp. 1269–70, Apr. 2005.

[50] V. B. Bajic, S. L. Tan, A. Chong, S. Tang, A. Ström, J. Gustafsson, C. Lin, and E. Liu, “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.

[51] 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.

[52] B. Kuntal, P. Aparoy, and P. Reddanna, “EasyModeller: A graphical interface to MODELLER,” *BMC Res. Notes*, pp. 1–5, 2010.

[53] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, “UCSF Chimera--a visualization system for exploratory research and analysis.,” *J. Comput. Chem.*, vol. 25, no. 13, pp. 1605–12, Oct. 2004.

[54] 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.

[55] A. Bateman, L. Coin, R. Durbin, R. D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E. L. L. Sonnhammer, D. J. Studholme, C. Yeats, and S. R. Eddy, “The Pfam protein families database.,” *Nucleic Acids Res.*, vol. 32, no. Database issue, pp. D138–41, Jan. 2004.

[56] E. L. L. Sonnhammer, S. R. Eddy, E. Birney, A. Bateman, and R. Durbin, “Pfam : multiple sequence alignments and HMM-profiles of protein domains,” *Nuc. Acids. Res.,* vol. 26, no. 1, pp. 320–322, 1998.

[57] C. M. Gould, F. Diella, A. Via, P. Puntervoll, C. Gemünd, S. Chabanis-Davidson, S. Michael, A. Sayadi, J. C. Bryne, C. Chica, M. Seiler, N. E. Davey, N. Haslam, R. J. Weatheritt, A. Budd, T. Hughes, J. Pas, L. Rychlewski, G. Travé, R. Aasland, M. Helmer-Citterich, R. Linding, and T. J. Gibson, “ELM: the status of the 2010 eukaryotic linear motif resource.,” *Nucleic Acids Res.*, vol. 38, pp. D167–80, Jan. 2010.

[58] T. Hunter, “Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling,” *Cell*, vol. 80, pp. 225-236, 1995.

[59] B. Polevoda and F. Sherman, “N-terminal Acetyltransferases and Sequence Requirements for N-terminal Acetylation of Eukaryotic Proteins,” *J. Mol. Biol.*, vol. 325, no. 4, pp. 595–622, Jan. 2003.

[60] 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- reveals binding to diverse ERE sequences and widespread binding within repetitive DNA elements,” *Nucleic Acids Res.*, vol. 38, no. 7, pp. 2355–2368, Jan. 2010.

[61] A. Schafferhans and G. Klebe, “Docking ligands onto binding site representations derived from proteins built by homology modelling,” *J. Mol. Biol.*, vol. 307, pp. 407-427, 2001.

[62] M. Behjati, I. Torktaz, M. Mohammadpour, G. Ahmadian, and A. J. Easton, “Comparative modeling of CCRL1, a key protein in masked immune diseases and virtual screening for finding inhibitor of this protein,” *Biomed. informatics*, vol. 8, no. 6, pp. 336–340, 2012.

[63] 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.

[64] S. Movérare-Skrtic, A. E. Börjesson, H. H. Farman, K. Sjögren, S. H. Windahl, M. K. Lagerquist, A. Andersson, A. Stubelius, H. Carlsten, J.-Å. Gustafsson, and C. Ohlsson, “The estrogen receptor antagonist ICI 182,780 can act both as an agonist and an inverse agonist when estrogen receptor α AF-2 is modified.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 3, pp. 1180–5, Jan. 2014.

[65] P. Yi, S. Bhagat, R. Hilf, R. a Bambara, and M. Muyan, “Differences in the abilities of estrogen receptors to integrate activation functions are critical for subtype-specific transcriptional responses.,” *Mol. Endocrinol.*, vol. 16, no. 8, pp. 1810–27, Aug. 2002.

APPENDIX A

PRIMERS

Table A 1. Primer List

Primer Name	Sequence (5'-3')
CXXC5_qPCR_FP (Origene)	CGGTGGACAAAAGCAACCCTAC
CXXC5_qPCR_REP (Origene)	CGCTTCAGCATCTCTGTGGACT
XNmetN_CXXC5_FP	CGCATATACTCGAGATTACCATGGAAGCTAGCTCGAGCCTCGGCGGTGGCTC
CXXC5_EpolyaB_REP	CGCATGGGATCCTTTATTAGAATTCCTGAAACCACCGGAAGGCGG
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
CXXC5_Ala_Mutant_FP	AATCGAAAGGCTGCCGCAGCGATTTGCAAATTCAGAAAATGTGAG
CXXC5_Ala_Mutant_REP	TTTGCAAATCGCTGCGGCAGCCTTTCGATTCCTACAACTGCTGCA
CXXC5_N_Term_REP	CGCATGGGATCCTTTATTAGAATTCCATGCCGCAGCGTTTCCGCTT
CXXC5_C_Term_FP	CGCATATACTCGAGATTACCATGGAAGCTAGCAGCTCCGGCAAGAAGAAGCGG
FP_BamHI_NcoI_Flag_NheI	CGCATGGATCCACCATGGAAGATTACAAGGATGACGACGATAAGGCTAGC
REP_CXXC5_EcoRI_polyA_XhoI	CGCATGCTCGAGTTTATTAGAATTCCTGAAACCACCGGAAGGCGG

APPENDIX B

siRNA SEQUENCES

Table B.1. siRNA Sequences

siRNA Name	Sequence (5'-3')
Hs_CXXC5_2	CAGCAGTTGTAGGAATCGAAA
Hs_CXXC5_7	CAAGAGCGGTATCATCAGTGA
Hs_CXXC5_9	AGGGATT CGGGCGAAGACAAA
Hs_CXXC5_10	TCAGATTTGCAAATTCAGAAA

APPENDIX C

OLIGONUCLEOTIDES

Table C.1. Oligonucleotides

Oligonucleotide Name	Sequence (5'-3')
Non-methylated CpG_Up	CTGTCAGTATCCGGTACAGGTCTA
Non-methylated CpG_Down	TAGACCTGTACCGGATACTGACAG
Methylated CpG_Up	CTGTCAGTATCC^{M}GGTACAGGTCTA
Methylated CpG_Down	TAGACCTGTACC^{M}GGATACTGACAG
Non-CpG_Up	CTGTCAGTATCAGTTACAGGTCTA
Non-CpG_Down	TAGACCTGTAACTGATACTGACAG
Up_ERE	GCGCCGAGTTATCAGGTCAGAGTGACCTGGAGATTCCTTA
Down_ERE	ATATAAGGAATCTCCAGGTCACTCTGACCTGATAACTCGG

APPENDIX D

CHARCOAL DEXTRAN TREATMENT OF FETAL BOVINE SERUM

10 g Charcoal, dextran coated (C6241, Sigma Aldrich, Germany) is added to 500 ml fetal bovine serum (FBS) in its own bottle. An autoclaved magnetic fish is used for stirring of the FBS in the medium speed not to break the charcoal into fine particles and incubation is performed at 4^{o}C for overnight. After the incubation, the mixed FBS with charcoal is divided as equally weighted into two sterile Nalgene bottles and centrifuged at 10800xg (8000 rpm for Sorvall SLA-3000 rotor) for 30 minutes at 4^{o}C. Both of the supernatants are transferred into a 0.45 µM sterile filter unit (Sarstedt, 83.1823, Germany) and filtered. For the second round of the treatment, again 10 g Charcoal, dextran coated is added and stirred with a sterile magnetic fish for 4-6 hours at 4^{o}C. After the incubation, the mixture is divided equally weighted into two sterile Nalgene bottles and centrifuged at 10800xg for 30 minutes at 4^{o}C. The supernatants are transferred into a 0.45 µM sterile filter unit (Sarstedt, 83.1823, Germany) and filtered in the biological safety cabinet. Finally, charcoal dextran treated fetal bovine serum (CD-FBS) is aliquoted as 40 ml into 50 ml falcons and stored at -20^{o}C.

APPENDIX E

GENOMIC DNA CONTAMINATION CONTROL

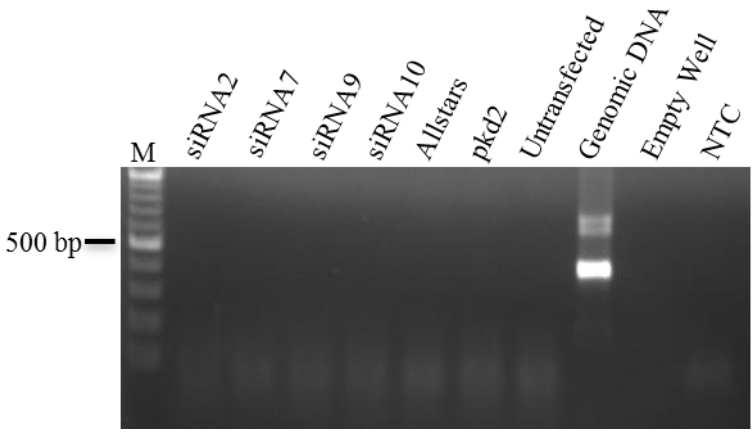


Figure E. 1. Genomic DNA contamination control for RNA samples. PCR was performed by using 500 ng of RNA samples as template with the *GAPDH* specific primers: GAPDH_F: 5'-GGGAGCCAAAAGGGTCATCA-3' and GAPDH_R: 5'-TTTCTAGACGGCAGGTCA GGT-3'. PCR conditions were as follows: incubation at 95^0C for 10 minutes, 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 (GeneRuler™ DNA Ladder Mix 100-10000 bp, Thermo Scientific, CA, USA). *PKD2* is the positive control siRNA. siRNA treated RNA samples were used for the representative image and the all RNAs that were used in this study were controlled likewise.

APPENDIX F

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**
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**

APPENDIX G

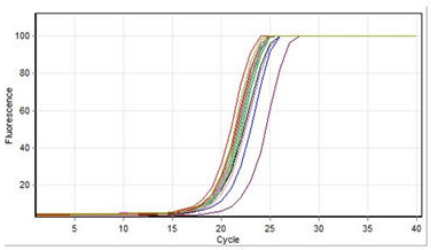
PERFORMANCE OF RT-QPCR REACTIONS

A representative example for the RT-qPCR results of *CXXC5*, similar conditions were used for *pS2/TFF1*, *RPLP0* and *PUM1*.

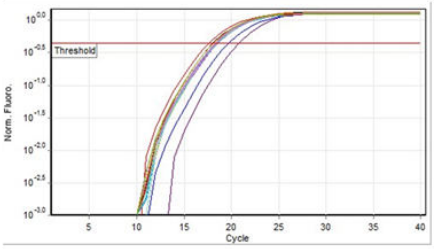
Quantitation information

Threshold	0,1494
Left Threshold	1,000
Standard Curve Imported	No
Standard Curve (1)	conc= 10^(-0,288*CT + 11,108)
Standard Curve (2)	CT = -3,468*log(conc) + 38,521
Reaction efficiency (*)	0,94256 (* = 10^(-1/m) - 1)
M	-3,46773
B	38,52061
R Value	0,99992
R^2 Value	0,99985
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	

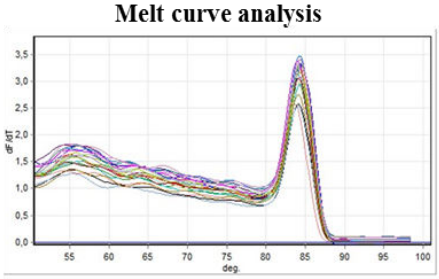
Raw data for Cycling A Green information

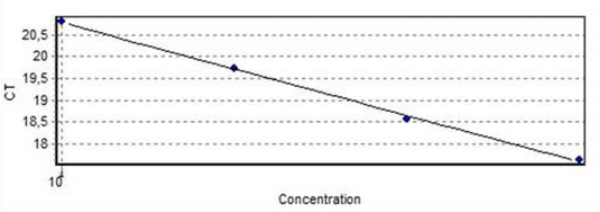


Quantitation data for Cycling A Green information



Melt curve analysis





No.	Colour	Name	Type	Ct	Given Conc (copies/ul)	Calc Conc (copies/ul)	% Var
1		1:10 (24h E2)	Standard	17,64	80.000	77.575	3,0%
2		1:20 (24h E2)	Standard	18,58	40.000	41.886	4,7%
3		1:40 (24h E2)	Standard	19,72	20.000	20.004	0,0%
4		1:80 (24h E2)	Standard	20,80	10.000	9.846	1,5%
23		NTC	NTC				

APPENDIX H

BUFFERS

TBE Buffer (5X) (1L)

Tris-base	54 g
Boric acid	27.5 g
EDTA	20 ml (0.5M, pH 8.0.)

6X Laemmli Buffer (10 ml)

Tris-base	3,75 ml (1M Tris, pH 6.8)
SDS	1.2 g
Glycerol	6 ml (100%)
Bromophenol Blue	1.2 mg

⇨ β-mercaptoethanol is added freshly (Final concentration is 30%)