

EPIGENETIC CHARACTERIZATION OF CXXC5 GENE LOCUS

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



BY
GİZEM KARS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
MOLECULAR BIOLOGY AND GENETICS

AUGUST 2018

Approval of the thesis:

EPIGENETIC CHARACTERIZATION OF THE CXXC5 GENE LOCUS

submitted by **GİZEM KARS** in partial fulfillment of the requirements for the degree of **Master of Science in Molecular Biology and Genetics Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, **Graduate School of Natural and Applied Sciences**

Prof. Dr. Orhan Adalı
Head of Department, **Biological Sciences**

Prof. Dr. Mesut Muyan
Supervisor, **Biological Sciences Dept., METU**

Examining Committee Members:

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

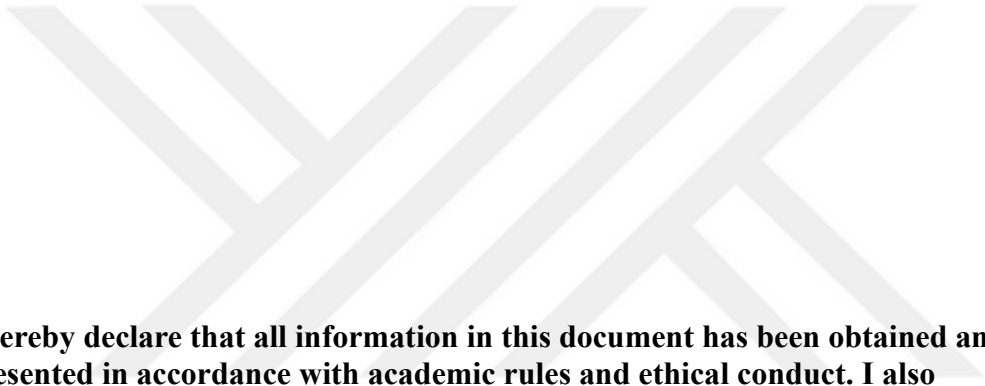
Prof. Dr. Mesut Muyan
Biological Sciences Dept., METU

Assoc. Prof. Dr. Çağdaş Devrim Son
Biological Sciences Dept., METU

Asst. Prof. Dr. Erkan Kiriş
Biological Sciences Dept., METU

Asst. Prof. Dr. Murat Cevher
Molecular Biology and Genetics Dept., Bilkent University

Date: 29.08.2018



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

Name, Last name : Gizem Kars

Signature :

ABSTRACT

EPIGENETIC CHARACTERIZATION OF *CXXC5* GENE LOCUS

Kars, Gizem

M.S., Molecular Biology and Genetics Department

Supervisor: Prof. Dr. Mesut Muyan

August 2018, 77 pages

17 β -estradiol (E2), the main circulating form of estrogen, induces diverse cellular responses through Estrogen Receptor α and β (ER α and ER β). E2 signaling mediates physiology and pathophysiology of several tissues by regulating E2 responsive genes.

Previous studies of our laboratory described a novel E2 responsive gene, *CXXC5* that is regulated through ER α . *CXXC5* is a member of Zinc Finger CXXC domain protein family. Like other members of family, *CXXC5* appears to function as epigenetic factor, co-regulator or transcription factor. Also several studies indicate that *CXXC5* is a poor prognostic factor in tissue malignancies including breast cancer which shows high dependency to E2 signaling. Thus, understanding the transcription of *CXXC5* is important. Despite the importance, transcriptional regulation of *CXXC5* has not been understood and mechanism of E2 mediated expression of *CXXC5* has not been explored.

It is known that E2 regulated gene expression involves dynamic interaction of transcription factors and epigenetic modifiers. Consequently, several epigenetic alterations of DNA including methylation, chromatin structure or histone modifications are critical components of gene expressions. Here, we aimed to test epigenetic features of *CXXC5* locus and effects of E2, at possible regulatory sites.

Initially we carried out *in silico* analysis to define putative promoter of *CXXC5*. Then we conducted DNA methylation and nucleosome profiling for the promoter and for previously identified enhancer region of *CXXC5* in MCF7 cells in the absence or presence of E2.

Our studies suggest that *CXXC5* has a CpG Island (CGI) promoter which is dramatically hypomethylated independently from E2. While the enhancer region and exon 2 as gene body is hypermethylated. Moreover, we detected a nucleosome free subregion at 5' of the *CXXC5* exon 1, in the untreated cells. The Enhancer region was identified as occupied independently from E2. In summary, although general description of epigenetic features of *CXXC5* regulatory regions has been identified in our study, detailed mechanism of E2 regulation could be explained by further studies.

Keywords: *CXXC5*, Estrogen, Transcription, DNA Methylation, Nucleosome

ÖZ

CXXC5 GEN LOKUSUNUN EPİGENETİK KARAKTERİZASYONU

Kars, Gizem

Yüksek Lisans, Moleküler Biyoloji ve Genetik Bölümü

Tez Yöneticisi : Prof. Dr. Mesut Muyan

Ağustos 2018, 77 Sayfa

Dolaşım sistemindeki başlıca östrojen çeşidi olan 17β -estradiol (E2), östrojen reseptörü-alfa ve beta ($ER\alpha$ ve $ER\beta$) aracılığıyla çeşitli hücresel tepkilere sebep olur. Östrojen sinyali çeşitli dokuların fizyolojisini düzenler ve patofizyolojik durumlarda rol oynar.

Laboratuvarımızda daha önce yapılan çalışmalar *CXXC5* genini $ER\alpha$ aracılığıyla düzenlenen yeni bir E2 yanıt geni olarak tanımlamıştır. *CXXC5*, çinko-pramak CXXC domain protein ailesinin bir üyesidir. Bu protein ailesinin diğer üyeleri gibi, *CXXC5* proteini de epigenetik faktör, ko-düzenleyici ve transkripsiyon faktörü olarak işlev görmektedir. Çalışmalar, çeşitli dokuların kanserlerinde *CXXC5* proteininin hastalık seyrini kötüleştirdiğini göstermiştir. Meme kanseri de bunlardan biridir. *CXXC5* meme kanseri tedavisinde bir hedef aday olabilir. Ayrıca meme kanseri gelişimi ve ilerleyişi E2 sinyaliyle bağlantılıdır. Dolayısıyla, E2 sinyalinin *CXXC5* geni üzerindeki etkilerini anlamak önemlidir. *CXXC5* gen ifadesinin nasıl düzenlendiği henüz tam olarak bilinmemektedir ve E2 sinyaline bağlı *CXXC5* gen anlatımındaki değişikliklerin mekanizması da henüz açıklanmamıştır.

E2 sinyali, transkripsiyon faktörleri ve epigenetik düzenleyicilerin dinamik etkileşimleri sayesinde gen ifadesini düzenler. Sonuç olarak DNA metilasyonu,

kromatin yapısı ve/veya histon modifikasyonlarındaki deęişiklikler gen ifadesini deęiştirir.

Bu alıřmada, *CXXC5* gen lokusunda bulunan olası dzenleyici blgelerin E2'den nasıl etkilendiklerini anlamaya alıřtık. ncelikle, bilgisayar ortamında *CXXC5*' in promotrn tanımlamaya ynelik analizler gerekleřtirdik. Sonrasında promotr blgesinin ve daha nceki alıřmalarımızda belirlenmiř olan hızlandırıcı (enhancer) blgesinin, E2 varıęında/yokluęunda, DNA metilasyon ve nkleozom profilini anlamaya ynelik alıřmalar yaptık.

alıřmalarımız *CXXC5* geninin promotrnn ilk ekzonun bařında bulunduęunu ve CpG adası olduęunu nerdi. Bu blgenin hipometile olduęu tespit edilirken, hızlandırıcı blgenin hipermetile olduęu gzlenmiřtir. Metilasyon profilinin E2 sinyalinden etkilenmedięi anlařılmıřtır. Ayrıca, aynı blgeler iin gerekleřtirilen nkleozom profili alıřmalarına gre, promotr blgesinin E2 yokluęunda nkleozomlar ve/veya dięer DNA'ya baęlanan proteinler tarafından iřgal edilmemiř olduęu gsterilmiřtir. Hızlandırıcı blge ise E2 sinyalinden baęımsız olarak, nkleozomlar veya dięer proteinler tarafından iřgal edilmiř grnmektedir. Genel hatlarıyla *CXXC5* promotr ve hızlandırıcı blgelerinin epigenetik zellikleri tespit edilmiř olsa da, E2 baęlantılı gen reglasyonunun mekanizması daha detaylı alıřmalarla aıklanabilir. Bunun iin alıřmalarımız devam etmektedir.

Anahtar Kelimeler: *CXXC5*, strojen, Transkripsiyon, DNA Metilasyonu, Nkleozom

ACKNOWLEDGMENTS

I would like to thank to my supervisor Prof. Dr. Mesut Muyan for his support and advice throughout my graduate study. I have learned a lot about scientific approach, critical thinking and planning future steps. I will always be grateful for his counseling that encourages me to be a well-rounded scientist.

I want to thank to all my thesis committee members, Prof. Dr. Elif Erson Bengan, Prof. Dr. Mesut Muyan, Assoc. Prof. Dr. Çağdaş Devrim Son, Assoc. Prof. Dr. Murat Cevher and Asst. Prof. Dr. Erkan Kiriş for their kind interest in being at my thesis defense and their valuable feedbacks.

I know it would not be easy to study in a miserable, unsteady and unfriendly environment. Luckily, our lab has been always friendly and easily adaptable thanks to Gamze Ayaz, Pelin Yaşar, Çağla Ece Olgun, Burcu Karakaya, Negin Rezizadeh, Kerim Yavuz and Gizem Turan. I progressed with their endless help and informative discussions. They mean a lot not just as lab mates but also as friends. I want to thank to our former lab member, Sırma Damla User, for her patience in teaching me several experiments when I was a very new master student.

I would like to express my thankfulness to my parents. They have supported me throughout my life. I will always cherish their endless love.

I also want to thank to Muhammet Durmaz. He has always encouraged me. No one else could cheer me up that much when I feel exhausted.

I am so glad to have my dear friends Cansu Dinçer, Didem Toker, Mertcan Özturan, Öner Murat Akbaba and Yağmur Kılıç. More than 10 years, they have been very supportive, fun, real source of motivation to be socialized despite all the

busyness. I also want to express my thanks to my home-mates; Gamze Ađaayak, Ekinsu Özmen, Aybüke Gülkaya and Süheyla the fluffy cat.

Finally, I would like to thank TUBİTAK for supporting this study through 114Z243 and 117Z213 projects.



TABLE OF CONTENTS

ABSTRACT.....	v
ÖZ.....	vii
ACKNOWLEDGMENTS.....	ix
TABLE OF CONTENTS.....	xi
LIST OF FIGURES.....	xiv
LIST OF TABLES.....	xvii
CHAPTERS	
1. INTRODUCTION.....	1
1.1 Estrogen Signaling.....	1
1.1.2. Estrogen Signaling and Breast Cancer.....	3
1.1.3. E2-ER α Regulated Transcription.....	3
1.1.3.1. Epigenetic Modifications as a Consequence of E2 Signaling.....	4
1.2. <i>CXXC5</i> as an E2 Responsive Gene.....	7
1.2.1 <i>CXXC5</i> Protein Product.....	8
1.2.2. Studies on <i>CXXC5</i>	9
1.3. Aim of Study.....	11
2. MATERIALS AND METHODS.....	13
2.1. Cell Culture and Cell Lines.....	13
2.2. Treatments.....	13

2.3. Treatment Control by Expression Analysis.....	14
2.3.1. Total RNA Isolation.....	14
2.3.2 Genomic DNA Contamination Control.....	14
2.3.3 cDNA Synthesis.....	15
2.3.4. Polymerase Chain Reaction to Control cDNA Synthesis.....	15
2.3.5. Real Time-Quantitative Polymerase Chain Reaction (RT-qPCR).	15
2.4. Endogenous Methylation and Nucleosome Occupation Analysis.....	16
2.4.1. Chromatin Shearing and GpC Methylase Reaction.....	17
2.4.2. Phenol:Chloroform:Isoamylalcohol Purification of DNA.....	18
2.4.3. Genomic DNA isolation and Bisulfite Conversion for Endogenous Methylation Profile.....	18
2.4.4 Bisulfite Polymerase Chain Reaction and TA Cloning.....	19
2.4.5. Sequencing and Analysis.....	20
3. RESULTS AND DISCUSSION.....	21
3.1. <i>In Silico</i> Analysis of the <i>CXXC5</i> Locus.....	21
3.1.1. <i>CXXC5</i> Might Be Regulated by a CpG Island Promoter.....	21
3.1.1. Nucleosome Positioning of <i>CXXC5</i>	26
3.2. Control of E2 Treatment on <i>CXXC5</i> Expression.....	27
3.3 Analysis of Endogenous Methylation Pattern of the <i>CXXC5</i> Promoter and the Enhancer Region.....	28
3.3.1. Methylation Profile of the Putative Promoter of <i>CXXC5</i>	30
3.3.2. Methylation Profile of the ERE Containing Enhancer Region and the Exon 2 of <i>CXXC5</i>	32
3.4. Investigation of Nucleosome Positions.....	33
3.4.1. Nucleosome Positioning of the Putative Promoter of <i>CXXC5</i>	34
3.4.2. Nucleosome Positioning Around the Putative Enhancer.....	36

4. CONCLUSION AND FUTURE DIRECTIONS.....	39
REFERENCES.....	43
APPENDICES	
A. PRIMER SEQUENCES.....	55
B. PCR CONTROLS.....	57
C. MIQE CHECKLIST.....	59
D. RT-qPCR REACTIONS.....	65
E. GpC METHYLASE ACTIVITY CONTROL.....	67
F. LOLLIPOP REPRESENTATIONS OF METHYLATION.....	69

LIST OF FIGURES

FIGURES

Figure 1. Schematic Representation of common structure of Estrogen Receptors..	1
Figure 2. Representation of ERE-dependent (A) and ERE-independent (B) pathways of E2 signaling.....	2
Figure 3. Cytosine methylation and demethylation [20].....	5
Figure 4: Genomic Representation of <i>CXXC5</i>	8
Figure 5: Protein structure of <i>CXXC5</i>	8
Figure 6: <i>CXXC5</i> in Wnt signaling.....	10
Figure 7: Putative Promoter of <i>CXXC5</i>	22
Figure 8. GC Content Distribution and GC-Skew for 5' of <i>CXXC5</i>	24
Figure 9: CpG Island Calculation for the Putative Promoter and Enhancer Regions of <i>CXXC5</i>	25
Figure 10: DNase I Hypersensitivity Data from UCSC-Genome Browser for Promoter and ERE region of <i>CXXC5</i>	27
Figure 11: Upregulation of <i>CXXC5</i> After E2 treatment.....	28
Figure 12: Bisulfite Conversion of Cytosine to Uracil.....	29
Figure 13: CpG Methylation of the 5' loci of <i>CXXC5</i>	31
Figure 14: Methylation Profile of the Region Encompassing Enhancer.....	33
Figure 15. Nucleosome Analysis of the 5' region of <i>CXXC5</i>	36

Figure 16. GpC Methylation Profile of the Region Encompassing the Enhancer and Exon2 Of <i>CXXC5</i>	37
Figure 17. Genomic DNA Contamination Control After RNA Isolation.....	57
Figure 18. GAPDH PCR For cDNA Synthesis Control.....	58
Figure 19. Amplification Curve, Melt Curve and Standard Curve of <i>CXXC5</i> and RPLP0 qPCR Reaction.....	65
Figure 20. GpC Methylase Enzymatic Activity Control with HaeIII Restriction..	67
Figure 21. Lollipop Graph Representation of Endogenous Methylation of the 5' end of the putative <i>CXXC5</i> Promoter from MCF7 cells.....	69
Figure 22. Lolipop Graph Representation of Endogenous Methylation of the Putative Enhancer of <i>CXXC5</i> from MCF7 cells.....	70
Figure 23. Lollipop Graph Representation of Endogenous Methylation of Exon2 from MCF7 cells.....	71
Figure 24. Lollipop Graph Representation of NOME sequences of <i>CXXC5</i> 5' Putative Promoter from MCF7 cells.....	72
Figure 25. Lollipop Graph Representation of NOME sequences of <i>CXXC5</i> Putative Enhancer from MCF7 cells.....	73
Figure 26. Lollipop Graph Representation of Endogenous Methylation of <i>CXXC5</i> 5' Putative Promoter from HL60 cells.....	74
Figure 27. Lollipop Graph Representation of Endogenous Methylation of <i>CXXC5</i> Putative Enhancer.....	75

Figure 28. Lollipop Graph Representation of NOME sequences of CXXC5
Putative Promoter.....76

Figure 29. Lollipop Graph Representation of NOME sequences of CXXC5
Putative Enhancer.....77



LIST OF TABLES

TABLES

Table 2: Reaction Conditions and Product Size Information.....	16
Table 3: GpC Methyltransferase Reaction Set Up.....	17
Table 4: Bisulfite PCR Conditions.....	19
Table 6. <i>GAPDH</i> Primers.....	55
Table 7. qPCR Primers.....	55
Table 8. Bisulfite Primers for <i>CXXC5</i>	56
Table 9. MIQE CHECK LIST.....	59



CHAPTER 1

INTRODUCTION

1.1 Estrogen Signaling

Estrogen is a steroid hormone synthesized from androgen by aromatase enzyme activity in several tissues and organs including the ovary, adipose, bone and brain. 17 β -estradiol (E2) is main circulating form of estrogen hormone. E2 is involved in the physiology and pathophysiology of human body. E2 signaling is one of the important event for the regulation of reproduction, cardiovascular development, immune system and nervous system. E2 also contributes to the initiation and development of target tissue malignancies [1], [2]. E2 exerts its effect on cells primarily through estrogen receptor alpha and beta (ER α and ER β), which are belong to nuclear hormone receptor (NHR) family. Like other NHR family members, ER α and ER β function as transcription factors activated by E2 binding [2]. ER α and ER β share common structural characteristics reflected in similar functional features that involve transactivation, ligand binding, dimerization and nuclear localization (Figure 1) [3], [4]. ER α and ER β are expressed in various levels at different cell types and elicit diverse cellular responses [5], [6].



Figure 1. Schematic Representation of common structure of Estrogen Receptors.

A/B region is the ligand independent transactivation domain responsible for protein-protein interaction. C region is DNA binding domain. D region is hinge domain and contains a nuclear localization signal. The E/F region contains a ligand binding domain, dimerization domain and ligand dependent transactivation domain [7]

Upon E2 binding, ERs undergo conformational changes, which generate interaction surfaces for co-regulatory proteins and for basal transcription factors. The E2-ER complex regulates gene expressions via two distinct ways; Estrogen Response Element (ERE)-independent and ERE-dependent pathways (Figure 2). The ERE-independent pathway is the route that ER interacts with other transcription factors bound to their cognate response elements. While, the ERE dependent pathway features the direct interaction of ER with estrogen response element, ERE. The consensus ERE is a palindromic DNA sequence, 5'-GGTCAnnnTGACC -3' [2]. Ligand activated ER binds to consensus ERE, as well as several derivatives of the consensus sequence. In addition, it is known that ER α can bind ERE in the absence of E2 [8]. ERE sequences can be located at proximal promoters or distal regions from transcription start sites [9]. ERE location and sequence is important in target gene responses to E2 [10], [11].

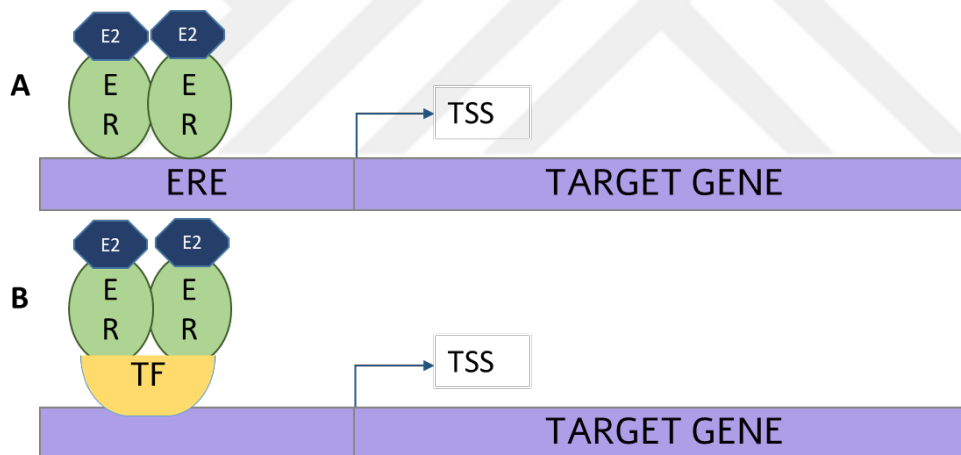


Figure 2. Representation of ERE-dependent (A) and ERE-independent (B) pathways of E2 signaling.

1.1.2. Estrogen Signaling and Breast Cancer

Breast cancer was recognized as a hormone dependent carcinoma since the removal of ovary leads to cancer regression. Consequently, hormone receptors (estrogen receptor (ER), progesterone receptor (PR) in conjunction with human epidermal growth factor receptor (HER2)) have been important for the classification of breast cancer. Hormone receptor-positive breast cancers make up 70-80% of all cases [12], [13]. Ovarian estrogen is one of the crucial steroid hormone involved in mammary gland development and tumorigenic response of breast tissue. ER α as primary receptor for E2 in mammary gland, mediates initiation and progression of breast cancer. Studies with hormone-receptor positive cell models like MCF7, indicates that ER α promotes S-phase entry by regulating many genes involved in proliferation [14]. Therefore, E2-ER α regulated transcription of responsive genes is important to understand breast cancer initiation and progression.

1.1.3. E2-ER α Regulated Transcription

Conformational changes of ER α after interacting with E2, increase DNA and protein binding affinity of receptor. Also ER α -ERE interactions promote several conformations to recruit co-regulatory proteins. Studies showed that the amino-terminally located AF-1 and the carboxyl-terminally located AF-2 regions of ER α provide surfaces for co-regulators that control interactions between receptor molecules and transcriptional apparatus as well as chromatin remodeling factors. Moreover, in the absence of E2 ligand, RNA Polymerase II is not recruited to promoter and transcription complex is disassembled [2],[15]. In brief, ER-dependent regulation of transcription is composed of several combinations of regulatory complexes that interact with DNA and transcription factors and also remodel chromatin structure dynamically as a consequence of epigenetic alterations [16].

1.1.3.1. Epigenetic Modifications as a Consequence of E2 Signaling

Epigenetic modifications are heritable and reversible chemical alterations on DNA and DNA binding histone proteins, which provide genetic control beyond nucleotide sequence. Epigenetic factors have an important role throughout development as they modulate tissue-specific gene expression, genomic imprinting and regulation of cell cycle [17]. Genome wide studies define ER α binding sites (ERBSs) as DNase I Hypersensitive regions and several histone modifications accepted as enhancer markers accumulated around ERBSs. Promoters of well-characterized E2 responsive genes showed assembly of co-activator complexes together with chromatin remodelers to provide chromatin environment permissive to transcription. Overall, these co-activators function in cellular events like nucleosome stabilization, destabilization, translocation and chromatin decompaction as well as the establishment of active state of chromatin via histone modifications due to the recruitment of DNA and Histone modifying enzymes [16], [18]. To sum up, studies have indicated that ER mediated events are tightly coupled to epigenetic changes. DNA base modifications and regulation of chromatin structure are important mechanisms of ER mediated regulation of transcription.

Cytosine Methylation

DNA methylation is one of the important phenomena among epigenetic modifications. In higher eukaryotes including humans, the 5th position of cytosine base can have a methyl group (5-methylcytosine, 5mC) within a Cytosine-phosphate-Guanine dinucleotide context (CpG). DNA Methyltransferase (DNMT) is the responsible enzyme that methylates cytosine. DNMT1 methylates CpG of hemimethylated DNA, which is important in epigenetic inheritance. DNMT3a/b enzymes perform de-novo methylation. In addition, Ten-Eleven Translocation (TET) enzyme catalyzes the conversion of the modified cytosine into

5-hydroxymethylcytosine (5hmC) and plays a key role in the active DNA demethylation event (Figure 3) [19].

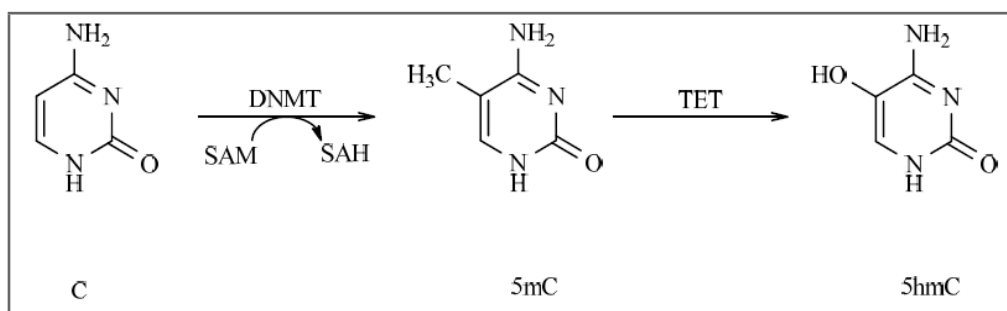


Figure 3. Cytosine methylation and demethylation [20].

CpG Island (CGI) contains high frequency of CpG dinucleotide clusters. Methylation status of CGI has regulatory effects on gene expression especially when they are located at the 5' of a gene, even some gene promoter, which are called as CGI promoters [20]. When a CGI has a few number of methylated cytosine, it is referred to as hypomethylated state. Hypomethylated CGI promotes gene expression by allowing transcription factors to bind to cognate responsive elements on DNA. On the other hand, highly methylated CGI is referred to as hypermethylated. Hypermethylation is a repressive marker. Transcription factor-DNA interaction is blocked by Methyl Binding Proteins (MBP), which bind to methylated CpG and recruit co-repressor proteins acting as histone modifiers [20], [21].

E2-ER signaling affects methylation profile of E2 responsive genes. For example, the initial demethylation at promoter of the well-known E2 responsive *PS2* (*TFF1*) gene, enhances transcription [22]. Cyclical recruitment of ER α , DNMTs and several chromatin remodelers is observed on PS2 promoter after E2 stimulation. This suggests that methylation-demethylation cycle is critical for ER α -mediated transcriptional regulation [22], [23]. Also it is observed that local hypermethylation or hypomethylation of DNA induced by E2, plays crucial roles

in regulating expression of several genes, both in physiology and pathophysiology [24], [25].

Chromatin Structure and Nucleosomes

Nucleosome is another important epigenetic regulatory factor for gene expression. Nucleosome is the main structural unit of chromatin, which packs DNA of eukaryotes [26]. DNA packaging is required to organize long genetic material within cell nucleus as well as transcription regulation.

Histones are positively charged protein components of nucleosomes. Thus, they bind to negatively charged DNA with ionic interactions. Two copies of Histone 2A, 2B, 3 and 4 generate an octomeric core structure to fold 147 bp DNA onto nucleosome and further organizational units with the help of linker Histone 1. Histones can bind variety of DNA sequences and can show binding preferential to some of those that contains high degree of A-T residue in contrast to C-G content [27].

Amino-terminus of histone called histone tail is not folded into globular histone structure, rather it is exposed to environment. This allows the histone tail to be post-translationally modified by histone modifying proteins.

Combinatorial effect of these different chemical modifications can alter chromatin structure [28]. Primary chemical modifications of histones are phosphorylation, acetylation and methylation. Mostly, acetylation and methylation neutralize positive charge of histone core and cause the loosening of DNA-nucleosome interaction or the ejection of nucleosome. Methylation does not affect charge of histones, but like other modifications, is critical for recruitment of chromatin remodelers [28], [29].

Studies showed that the presence of chemically modified histones and mediation of nucleosome occupancy regulate transcription initiation and elongation. Nucleosomes can prevent transcription initiation by rendering DNA inaccessible

to transcription machinery or enhance by placing promoter and enhancer regions in close proximity [30], [31].

E2 induction and ER α dependent regulation of transcription also involves chromatin remodeling [2], [32]. Particularly, induction of transcription is known to be highly correlated with nucleosome depletion or remodeling at promoter and enhancer regions. For instance, E2 responsive *PS2* (trefoil factor 1, *TFF1*) gene has two nucleosomes that are located at the TATA promoter and the ERE site. E2-ER α signaling is shown to cause alterations in the location of these nucleosomes [33], [34]. Moreover, the recruitment of several histone modifying enzymes like Histone Acetyl Transferases (HATs) and Histone Methyl Transferases (HMTs) as well as the chromatin remodeler SWI/SNF is shown to be responsible for this regulation. These proteins initially stabilize nucleosome positions and then cause chromatin disruption locally in a dynamic way [16]. Another study suggests that ER α binding to ERE, promotes H3K4 methyl transferase recruitment and activates FOXA1 dependent chromatin decomposition at promoter [35]. Citrullination of Histone 3 Arginine 26 has also effect on chromatin remodeling in response to E2-ER signaling [36].

In summary, cross talk between DNA methylation, histone modification and chromatin remodeling is one of the main mechanism behind E2-ER signaling in mediating gene expression.

1.2. *CXXC5* as an E2 Responsive Gene

Our earlier studies suggest that *CXXC5* is one of the E2 responsive genes regulated by ER α [37], [38]. These studies also showed the augmentation of *CXXC5* expression occurs through a putative promoter region containing a functional, non-consensus ERE sequence located at 240 bp upstream of Translation Start Site (TrSS) [38]. *CXXC5* also known as Retinoid Inducible Factor (*RINF*), is found on chromosome 5 and spans around 35,5 kb on human genome according to GRCh38 assembly (Figure 4). *CXXC5* has 11 exons that

result in 14 transcript variants that encode the same protein. The main transcript (NM016463.8) is 2661 bp mRNA including 969 bp coding region which encodes ~33 kDa protein product [39].

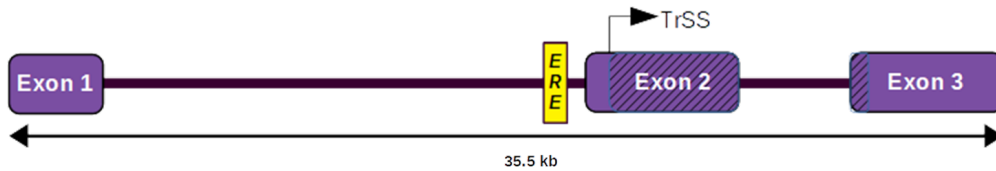


Figure 4: Genomic Representation of CXXC5.

CXXC5 spans ~ 35.5 kb on the genome and main transcript has three exons. The first and the second exon is separated with a long (30 kb) first intron. 240 bp upstream of translation start site (TrSS) contains a non-consensus functional ERE sequence. Striped regions of Exon 2 and 3 indicates open reading frame of the gene.

1.2.1 CXXC5 Protein Product

CXXC5 protein belongs to the Zinc Finger-CXXC domain protein family (ZF-CXXC) due to a highly conserved zinc finger domain. The tertiary structure of the protein is not fully resolved; however, it is known that CXXC domain is positioned between 256th and 297th amino acids. This region also contains Nuclear Localization Signal (NLS) between 257th and 262nd amino acids (Figure 5) [40]. CXXC5 is localized at the nucleus and can be phosphorylated at threonine residue at the 53rd position [41], [42].

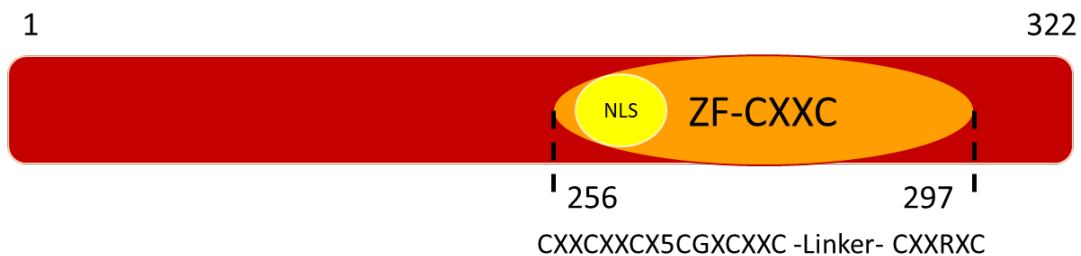


Figure 5: Protein structure of CXXC5. Orange region depicts Zinc Finger CXXC domain and yellow region indicates NLS. Cysteine rich conserved amino acid sequence also is indicated.

The CXXC domain is crucial for the molecular and biological functions of ZF-CXXC domain protein family, providing structural basis for DNA binding. The CXXC domain contains two cysteine rich clusters with two Zn²⁺ ions and functions as a non-methylated CpG binding module. Most members of ZF-CXXC family contains intrinsic chromatin remodeling functions and recruit other remodelers to regulate gene expression [43].

1.2.2. Studies on CXXC5

In 2009, CXXC5 was identified as a retinoid inducible factor (RINF) in NB4 myeloid cell line. Studies showed that retinoic acid induces CXXC5 expression and thereby, terminal maturation of leukemic myeloblasts [41]. CXXC5 is important in the differentiation of granulocytes during normal development; and dysregulation of expression of or mutations in CXXC5 can cause development of leukemia [41], [44], [45]. Several further studies on CXXC5 in Acute Myeloid Leukemia (AML) and Myelodysplastic Syndrome (MSD) patients found a strong correlation with high degree of CXXC5 expression and the overall survival and relapse rate [45], [46].

Although limited, studies suggest that CXXC5 is an important transcription factor in development and differentiation of other tissues and cell types. As a negative feedback of Wnt/ β catenin pathway (Figure 6), CXXC5 is negatively regulates osteoblast differentiation and wound healing by interacting with dishevelled protein which is important in Wnt signal transduction in cells [47], [48].

Several clinic researches target CXXC5-Dishevelled interaction to re-establish bone formation [49], [50].

Also, Bone Morphogenic Factor (BMP) signaling upregulates Vascular Endothelial Growth Factor Receptors (VEGFRs) via CXXC5 leading to endothelial cell differentiation and vessel formation [51]. Neuronal differentiation and brain development are also regulated by CXXC5. Through Wnt/ β catenin

pathway, the upregulation of CXXC5 expression as an activator for myelin genes is shown to be critical for Neuronal Stem Cells and promotes oligodendrocyte differentiation [52]. Myogenin genes important for skeletal myogenesis are also shown to be regulated by CXXC5 [53].

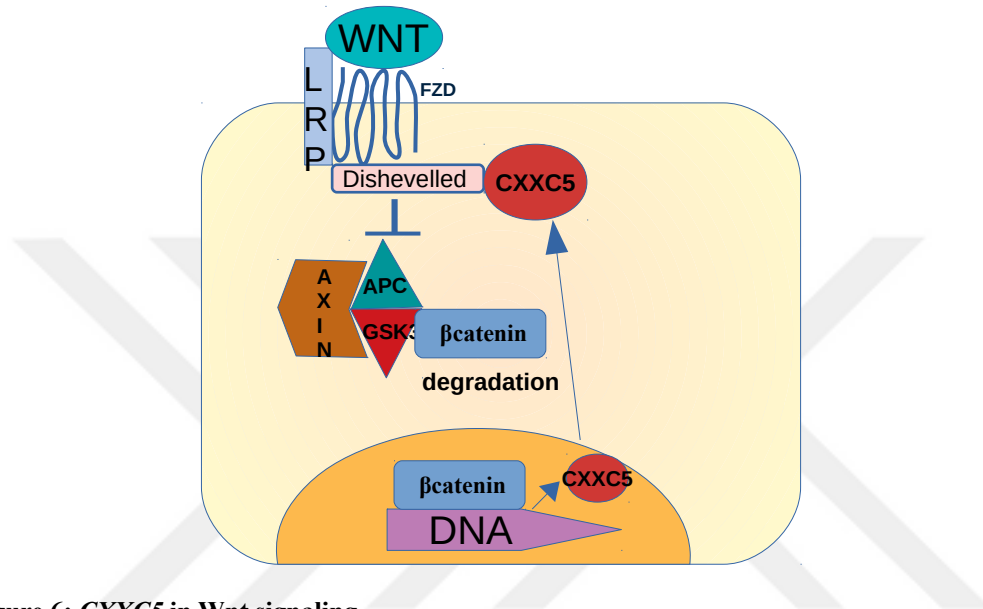


Figure 6: CXXC5 in Wnt signaling. β-catenin activated by Wnt signaling upregulates CXXC5 expression, which generates negative feed back loop. [47].

CXXC5 involves in the regulation of metabolism by modulating cytochrome c oxidase gene expression that has an oxygen responsive element (ORE). CXXC5 is suggested to bind ORE and negatively regulates transcription [54]. Studies on plasmacytoid dendritic cells concluded that CXXC5 together with TET2 maintain hypomethylated state of Interferon Regulatory Factor 7 (*IRF7*) promoter, which is regulated in response to pathogenic nucleic acid sensed by Toll-like receptor 7-9 (TLR7-9) [55]. Another identified function of CXXC5 in immune cells is the repression of CD40 ligand encoding gene, which is a marker of CD4⁺ helper T cells. In CD8⁺ cytotoxic T cells, CXXC5 represses CD40 by interacting with chromatin remodeler SUV39H1 [56].

Apoptosis is another cellular event that *CXXC5* is shown to be involved. *CXXC5* induces TNF α activated apoptosis in HEK293 cells [57]. In addition to TNF α , *CXXC5* is shown to be a target gene and positive feedback factor of TGF β signaling that activates apoptotic events [58]. Hence, *CXXC5* is identified as a tumor suppressor in hepatocellular carcinoma and peripheral nerve sheath tumors [58]–[60]. In ER+ breast cancers, *CXXC5* is found to be as a poor prognostic factor with an overexpression status as well as, in malignant melanoma, prostate cancer tissue and thyroid tumor, within which *CXXC5* is overexpressed when compared with normal/benign tissues [61]–[63].

1.3. Aim of Study

CXXC5 as a novel member of CXXC domain containing family member, impacts physiology and pathophysiology of several tissues as transcription factor, co-regulator and/or epigenetic factor. However, a very little is known about *CXXC5* protein structure, function as well as the transcriptional regulation of *CXXC5*. Previous studies of our laboratory showed that *CXXC5* expression is upregulated by E2-ER α through ERE dependent pathway in MCF7 breast cancer cell line [38]. Several studies observed that *CXXC5* is overexpressed in ER+ breast cancer tissues and affect prognosis of cancer [62]. These findings give rise to a prediction that *CXXC5* could be an important target for cancer treatment. Thus, a better understanding of *CXXC5* expression and its relationship to E2-ER signaling in MCF7 cells as a model cell line for ER+ breast cancer could provide a necessary prelude to explore the importance of this protein in physiology and pathophysiology.

Due to the fact that E2-ER signaling mediates transcription through complex and interdependent events that include epigenetic alterations, we aimed here to explore whether or not E2-ER α signaling mediated *CXXC5* transcription also involves alterations at regulatory regions of the *CXXC5* gene.



CHAPTER 2

MATERIALS AND METHODS

2.1. Cell Culture and Cell Lines

MCF7 cells were gift of Dr. Rengül Çetin Atalay from CanSyL of the Middle East Technical University, Ankara, Turkey. HL60 cells were gift of Dr. Can Özen, Middle East Technical University, Ankara, Turkey. Both cell lines were maintained in 8% fetal bovine serum (FBS, Biochrom, Germany, S0115), 1% penicillin-streptomycin (Lonza, Belgium, 17-602E) and 2.4 mM L-glutamine (Lonza, Belgium, 17-605E) added Dulbecco's Modified Eagle Medium (DMEM) with high glucose (4.5 g/L) without phenol red (Lonza, Belgium, 12-917F). Cells were grown at 37°C with 5% CO₂ in 95% humidified incubator. MCF7 cells maintained as monolayer and HL60 as suspension. The duration of maintenance did not exceed six weeks for both cell lines.

2.2. Treatments

For hormone depletion, MCF7 cells were grown in 8% charcoal dextran-fetal bovine serum (CD-FBS), 1% penicillin-streptomycin and 2.4 mM L-glutamine supplemented DMEM with high glucose without phenol red for 48 hours before treatment. Cells were treated with 10⁻⁸ M steroid hormone 17β-estradiol (E2, Sigma Aldrich, Germany, E2257) or 0,001% molecular biology grade ethanol as vehicle control. MCF7 cells cells were treated without or with E2 for 3h. Cells were then collected, counted and the counts were recorded. Collected cells were

divided in three part to use as treatment control, chromatin accessibility and endogenous methylation profiling analyses.

2.3. Treatment Control by Expression Analysis

2.3.1. Total RNA Isolation

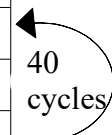
After treatment, MCF7 cells were detached by trypsin-EDTA (Lonza, Belgium, 17-161E) collected and washed with 1X phosphate buffer saline (PBS, Lonza, Belgium, BE17-516F). For total RNA isolation, Quick-RNA MiniPrep kit from Zymo Research, USA (R1055) was used according to manufacturer instructions. Concentration of total RNA is measured by BioDrop μ LITE (BioDrop, UK, 80-3006-51)

2.3.2 Genomic DNA Contamination Control

To avoid genomic DNA contamination in RNA samples, on column DnaseI treatment was carried out. Also using isolated RNA as template, genomic DNA contamination was tested by conventional PCR method. 0.5 μ M Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) forward and reverse primers were used in PCR with 1X Taq Buffer with KCl, 2mM MgCl₂, 200 μ M dNTP mix and 1U Taq Polymerase (Thermo Scientific, USA, EP0402). 3-step reaction conditions were adjusted as indicated in Table 1.

Table 1. PCR conditions

Initial denaturation	95°C for 3 minutes
Denaturation	95°C for 30 seconds
Annealing	65°C for 30 seconds
Extension	72°C for 30 seconds
Final extension	72°C for 10 minutes



Human genomic DNA was used as positive control at the same set of reaction. PCR was evaluated on agarose gel electrophoresis. If genomic DNA contamination was observed in RNA samples, Dnase I digestion was repeated according to Zymo Research Quick RNA MiniPrep kit protocol.

2.3.3 cDNA Synthesis

From 500 ng total RNA, cDNA conversion was carried out using RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific, USA, K1621). Firstly, Oligo(dT)₁₈ primers and RNA template were mixed and volume was completed to 12 µl with Rnase-Dnase free water. The mixture was incubated at 65°C for 5 minutes and samples were kept on ice for a couple of minutes. To complete reverse transcription, reaction mixture with a final 1X concentration of reaction buffer, 1mM dNTP mix, 1 U/µl Riblock Rnase Inhibitor and 10 U/µl RevertAid M-MuLV reverse transcriptase was prepared as master mix and added on previous reaction as 8 µl to end up with 20 µl total reaction which is incubated at 42°C for one hour for cDNA synthesis and at 70°C for 5 minutes to terminate reaction.

2.3.4. Polymerase Chain Reaction to Control cDNA Synthesis

To ensure that cDNA synthesis was successful, conventional PCR was carried out with GAPDH primers. Reaction conditions were same that is indicated in 2.3.2 Genomic DNA Contamination Control.

2.3.5. Real Time-Quantitative Polymerase Chain Reaction (RT-qPCR)

In order to control the effect of E2 treatment, RT-qPCR was conducted using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad, USA, 172-5271). Reaction setup is composed of 0.3 µM final concentration of forward and reverse primer in 20 µl total reaction. As template, 1:20 cDNA dilutions were used for E2 treated and vehicle control. Reaction conditions given in Table 2, were adjusted on CFX Connect Real Time System (Bio-Rad, USA) according to each primer set.

Ribosomal Protein Lateral Stalk Subunit P0 (*RPLP0*) gene expression is used for normalization and calculated as fold change. Primer sequences are available in Appendix A.

Table 2: Reaction Conditions and Product Size Information

Steps / Genes	<i>RPLP0</i>	<i>CXXC5</i>
Polymerase activation and Initial DNA denaturation	95°C for 10 minutes	
Denaturation	94°C for 30 sec	
Annealing	60°C	
Extension and plate read	72°C for 30 sec	
Rapid heating	95°C for 10 sec	
Melt Curve Generation	55°C to 99°C, increment 1.0°C, 5.0 sec	
Product size (base pairs)	191	145

Results are indicated by Fold change calculation ($\Delta\Delta C_t$), whose formula is shown below. MIQE Check List is available in Appendix B

$\Delta\Delta C_t$ Calculation for Gene Expression Analysis

$$\frac{(1 + RE_{CXXC5})^{Ct_{EtOH-E2}}}{(1 + RE_{RPLP0})^{Ct_{EtOH-E2}}}$$

2.4. Endogenous Methylation and Nucleosome Occupation Analysis

After 3 hours E2 treatment, cells to be used in DNA isolation for endogenous methylation analysis were stored in -80°C as pellet until processing. For nucleosome analysis, 3×10^6 number of cells were fixed with 1% final concentration of Formaldehyde for 10 minutes at room temperature. To stop

fixation, 0.125 M Glycine was added and incubated for additional 10 minutes at room temperature. Then cells were centrifuged at 600 g for seven minutes and washed with 1X PBS twice. Until used, cell pellets were stored at -80°C with addition of PMSF and PIC to prevent protein degradation during freeze-thaw.

2.4.1. Chromatin Shearing and GpC Methylase Reaction

Fixed and frozen cells were thawed on ice and re-suspended with ice-cold lysis buffer (Active motif, USA, 54000) supplemented with PIC and PMSF. Cell suspensions were incubated on ice for 30 minutes and pelleted at 2500g for 10 minutes in 4°C microcentrifuge. Pellets were washed with 1X nuclear wash buffer (Active Motif, USA, 5400) twice. While second wash, 7.5×10^5 cells were aliquot according to counts before fixation and centrifuged again at 2500g for 10 minutes in 4°C microcentrifuge. Pellets were re-suspended with 100 μ l 1X GpC Methyltransferase buffer (NEB, USA, M0227S) supplemented with PIC and PMSF. After, samples were sonicated with 30 second on-and-off intervals at 100% amplitude. 40 minute of active sonication (80 minutes in total) was applied. Sonicated samples were spun down at 14000g for 10 minutes in 4°C microcentrifuge and supernatant was collected into a new tube.

For methyl transferase reaction, sonicated chromatin samples from 2.5×10^5 cells (33 μ l) was used. Sheared chromatin mixed with GpC Methyltransferase and S-adenosylmethionine (SAM) (NEB, USA, M0227S) was incubated at 37°C for 4 hours. Reaction set up is indicated at Table 3.

Table 3: GpC Methyltransferase Reaction Set Up

Reagent	Volume (μ l)	Notes
Sheared chromatin (250k cells)	33	
10 X Reaction Buffer	4	
GpC Methylase (4U/ μ l)	5	
20X SAM (3.2mM)	10	Add 2 μ l after each hour of reaction
Water	Up to 50	

At the end of the 4-hour incubation, NaCl at the final concentration of 300 mM was added onto chromatin samples; chromatin samples were then reverse cross-linked at 95°C for 15 minutes. After cool down to room temperature, final concentration of 0.2 mg/ml RNase A (Thermo Scientific, USA, EN0531) was added into chromatin samples to degrade RNAs at 37°C for 30 minutes. Additionally final concentration of 0.25 mg/ml Proteinase K (Thermo Scientific, USA, AM2542) was added into samples to degrade proteins and samples were incubated at 65°C for overnight.

2.4.2. Phenol:Chloroform:Isoamylalcohol Purification of DNA

To purify DNA, samples were mixed with 1 volume of buffered phenol:chloroform:isoamylalcohol mixture and centrifuged to generate phase separation. The top aqueous phase containing DNA was transferred to a new tube. 1/10 volume of sodium acetate and 2 volumes of 100% ice cold EtOH was added to the sample and mixed well. Samples were incubated at -80°C for 1 hour. DNA precipitate was washed with 70% EtOH twice, air dried and re-suspended with ultrapure water.

2.4.3. Genomic DNA isolation and Bisulfite Conversion for Endogenous Methylation Profile

Pelleted cells were stored at – 80°C until genomic DNA isolation. For isolation, Quick-gDNA MiniPrep kit from Zymo Research (USA, D3024) was used according to manufacturer's protocol and the sample concentration was measured. Isolated gDNA samples were stored at – 20°C.

500 ng of gDNA was subjected to bisulfite conversion reaction using EZ-DNA Methylation Lightning kit (Zymo Research, USA, D5030) according to the company's instructions. The concentration of bisulfite converted DNA was recorded.

2.4.4 Bisulfite Polymerase Chain Reaction and TA Cloning

Bisulfite DNA was used as template for PCR reaction. Bisulfite primers were designed using Methyl Viewer software developed by Ian Carr from the University of Leeds, United Kingdom. Primers are listed on Appendix. LongAmp Taq Polymerase (NEB, USA, M0323) was used as 2.5 units in 50 μ l total reaction with final concentration of 1X LongAmp Taq Reaction Buffer, 0.5 μ M forward and reverse primers. Template DNA was added to reaction at 90°C to prevent non-specific primer annealing and first 2 cycles of reaction was driven by just reverse primer that is complement to sense strand in order to avoid primer dimer. After that, forward primer was added at 90°C. Reaction conditions are described in Table 4.

Table 4: Bisulfite PCR Conditions

Steps / Regions	5' region 1	5' region 2	ERE region 1	ERE region 2
Initial Denaturation	94°C for 40 seconds			
Denaturation	94°C for 30 seconds			
Annealing	55°C			
Extension	65°C for 40 seconds			
	10 cycles			
Denaturation	94°C for 30 seconds			
Annealing	45°C			
Extension	65°C for 40 seconds			
	40 cycles			
Final Extension	65°C for 10 minutes			

PCR products were loaded onto a 2 % agarose gel; and appropriate size fragments were gel extracted using Zymoclean Gel DNA Recovery Kit (Zymo Research, USA, D4007) according to manufacturer's instructions.

Recovered PCR products were ligated into PGEM-T Vector (Promega Corporation, USA, A3600) according to the PGEM-T and PGEM-T Easy Vector System Manual. Ligation products were transformed into XL1-Blue (Agilent Technologies, Santa Clara, CA, USA) competent cells. Transformed bacterial cells were grown on LB agar plate containing 100 µg/ml ampicillin. For blue-white selection, 40 µg/ml IPTG (Zymo Research, USA, I-1001-5) and 1mM X-Gal (Zymo Research, USA, X-1001-5) were spread on plates before bacterial inoculation. Plates were incubated at 37°C for overnight and colony screening was carried out with white colonies. Positive colonies were grown in 4ml LB culture containing 100 µg/ml Ampicillin at 37°C for overnight. Plasmids were isolated using QIAprep Spin Plasmid MiniPrep kit(Qiagen, Germany, 27106).

2.4.5. Sequencing and Analysis

PRZ Biotechnology (Ankara, Turkey) sequenced isolated plasmids. Sanger sequencing method with T7 or SP6 PGEMT vector primers were used. Sequences were analyzed by Methyl Viewer Software [64].

CHAPTER 3

RESULTS AND DISCUSSION

3.1. *In Silico* Analysis of the *CXXC5* Locus

3.1.1. *CXXC5* Might Be Regulated by a CpG Island Promoter

Our ongoing studies conducted in the laboratory by Pelin Yaşar, to find the promoter of *CXXC5* in MCF7 cells, suggest that *CXXC5* has a putative promoter located at the 5' end of the Exon 1. Result of genome wide Pol II-ChIP-Seq studies analyzed by Eukaryotic Promoter Database (EPD) also correlates with this findings (Figure 7A). This putative *CXXC5* promoter does not have identified conserved promoter elements like TATA or CAAT box. However, when we investigated transcription factor ChIP-Seq studies deposited and analyzed by Gene Transcription Regulation Database (GTRD), we identified that there are a number of sites for transcription factors (TFs), for example SP1 and CTCF that are GC rich DNA sequence binding proteins and are critical for Pol II recruitment [65]. We concluded that this putative promoter of *CXXC5* might be a TATA-box lacking, GC-rich promoter. Examples of GC-box binding TFs that are enriched on the 5' region of *CXXC5* are represented in Figure 7B.

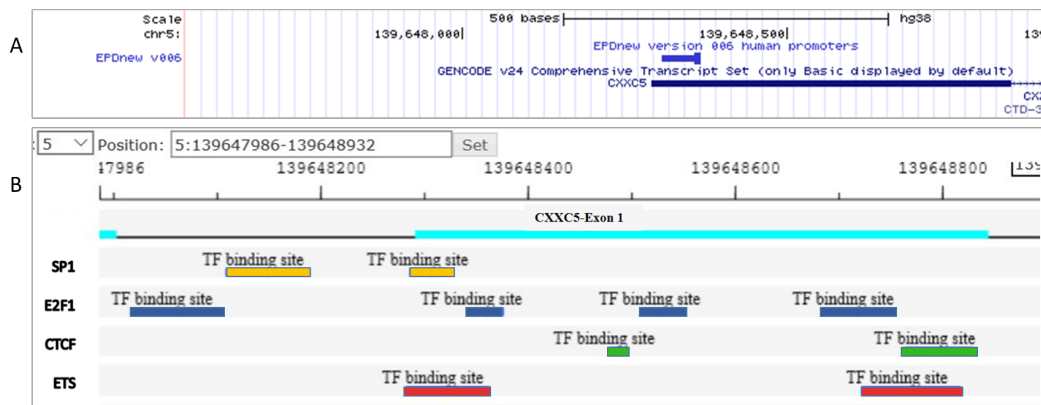


Figure 7: Putative Promoter of *CXXC5*.

EPDnew track indicates that *CXXC5* has a promoter region located in the exon 1 (A). Gene Transcription Regulation Database demonstrates the presence of several transcription factor binding sites on the putative promoter which seems to be enriched in GC-rich sequence binding transcription factors, including SP1, E2F1, CTCF, ETS and which are widely observed on GC rich promoters (B).

In summary, preliminary studies to determine promoter of *CXXC5* raise the possibility that the *CXXC5* expression is regulated by a CpG Island (CGI) promoter. CGIs colocalize with the majority (more than 70%) of promoters in the human genome. CGI promoters are characterized with DNA sequences that are highly rich in G and C nucleotide content and in CpG dinucleotides and show significant strand asymmetry in the distribution of G and C, or GC skew. Importantly, CGI promoters are predominantly nonmethylated [66]. CGIs are also organized in a characteristic chromatin structure that predisposes them toward a promoter activity. Compared to non-CGI promoters, the transcription start sites (TSSs) contain few nucleosomes and the nucleosomes at surrounding regions of TSSs are ornamented with histone markers, such as H3K4me3, that signify the poised or active transcriptional state. Furthermore, CGI promoters often lack core promoter elements such as the TATA box. Although the mechanism(s) is unclear, transcription initiation at CGI promoters appears to be mediated by the recruitment of Pol II by transcription factors bound to cognate response elements. Studies indicate that CGI promoters tend to be enriched for transcription factor-

binding motifs including SP1, ETS, CTCF that are capable of recruiting Pol II [67], [68]. This diversity in *cis*-element motifs consequently leads to a plasticity for the selection of TSS, dependent upon *trans*-acting factors specific to a given CGI promoter. As indicated, *CXXC5* has these properties, so we wanted to evaluate in detail our prediction that the expression of *CXXC5* is regulated by a CGI promoter.

We initially carried out an *in silico* analysis to examine 5' of *CXXC5* in terms of CGI promoter features. Firstly, we calculated GC content as $[\text{Count}(\text{G} + \text{C}) / \text{Count}(\text{A} + \text{T} + \text{G} + \text{C}) * 100\%]$ from -1000 to +1000 nucleotide of Exon 1 (Biologics International Corp, USA). As seen in CGI promoters, *CXXC5* has high GC content, which is 66%. Moreover, GC level increases nearby exon 1 (Figure 8A). Then by using GenSkew Tool (University of Vienna, Austria), we evaluated GC skew of the same region by the formula : $[\text{count}(\text{G} - \text{C}) / \text{count}(\text{G} + \text{C})]$. According to the cumulative distribution of the GC skew of *CXXC5*, the beginning of exon 1 has an increasing GC skew, which is another important feature of CGI promoters (Figure 8B).

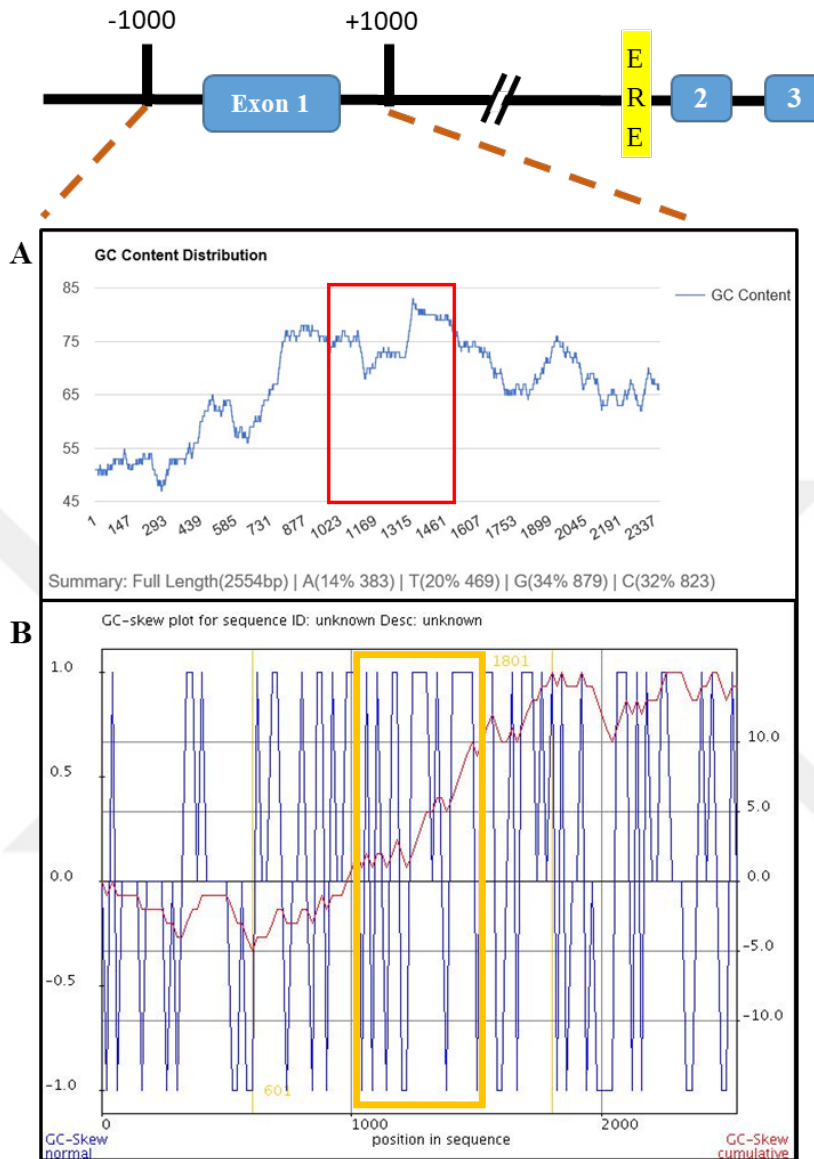


Figure 8. GC Content Distribution and GC-Skew for 5' of CXXC5.

GC content of indicated region was calculated as $[\text{Count}(G + C)/\text{Count}(A + T + G + C) * 100\%]$ with a window size 200. Average GC content was 66 % and also there is an increase in GC content distribution nearby exon 1 (A). GC Skew was calculated according to $[\text{count}(G - C) / \text{count}(G + C)]$ formula with a window size 200. Cumulative GC Skew was indicated as red line, which increases towards exon 1. Blue line is normal GC skew. From exon 1 beginning to the end, there is an increase in GC skew which is a common feature of CGI promoters (B). The elocation of exon 1 is indicated by red and yellow square respectively.

In prediction of CGI, there are several criteria that need to be taken into account: the percentage of G-C content should be higher than 60, the length of CGI should be longer than 300 bp and the observed/expected ratio of GC should be more than 0.6, and the promoter should be located on 5' end of CGI. Our *in silico* analyses using Methyl Primer Express software (Figure 9A) suggest that the putative promoter of *CXXC5* contained in a CGI is a CGI promoter. CGI evaluation was also conducted for the putative enhancer region. We found that in the exon 2 there is a CGI (Figure 9B), which might be regulated through E2-ER α , through binding to an upstream ERE. This in turn raises possibility that E2-ER α signaling could alter epigenetic features, including DNA methylation and nucleosome positioning, of both the promoter and enhancer regions.

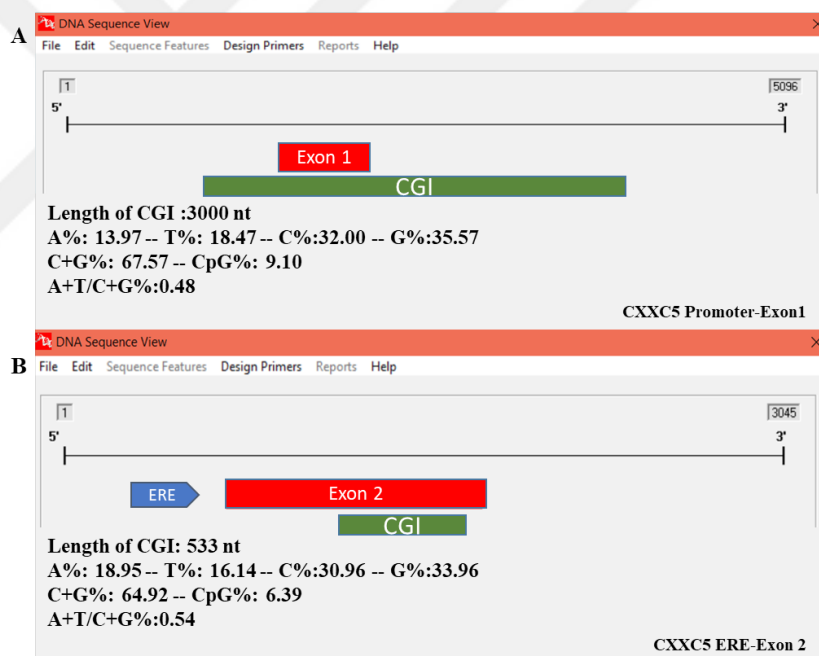


Figure 9: CpG Island Calculation for the Putative Promoter and Enhancer Regions of *CXXC5*.

CGI evaluation for promoter (A) and enhancer (B) showed that promoter is within a CGI and enhancer has a CGI nearby itself. Length of CGI, percentages of each nucleotide, total percentage of guanine and cytosine, percentage of CpG dinucleotide among GC contents and percent ratio of AT content to GC content are presented on figure.

3.1.2. Nucleosome Positioning of *CXXC5*

As mentioned above, CGI promoters display a transcriptionally paused or active state characterized with dispersed nucleosome containing histone modifications that signify paused/active transcription state. CGI promoters and enhancers containing no or sparse nucleosomes indicate open chromatin structures available for the binding of transcription factors. Several positions of a given locus might be dissociated from nucleosomes upon signal transduction that activates gene expression. It is known that E2 signaling affects nucleosome positioning thereby altering chromatin structure [33] Before we address this possibility, we also carried out an initial *in silico* analysis for the epigenetic state of the putative promoter and enhancer regions of *CXXC5*.

DNase I is a restriction enzyme that randomly digests accessible, free of proteins, DNA and is used for the identification of nucleosome free regions called DNase I Hypersensitive sites. Genome-wide DNase I Hypersensitive sites, indicating possible nucleosome free regions, have been identified by [John Stamatoyannopoulos lab](#) from the University of Washington and deposited to UCSC, Genome Browser for visualization. Using this database, we found that in MCF7 cells that the putative promoter region of *CXXC5* displays DNase I peaks, while the encoding Exon 2, including the enhancer ERE region, located about 30kb downstream of the Exon 1 does not (Figure 10).

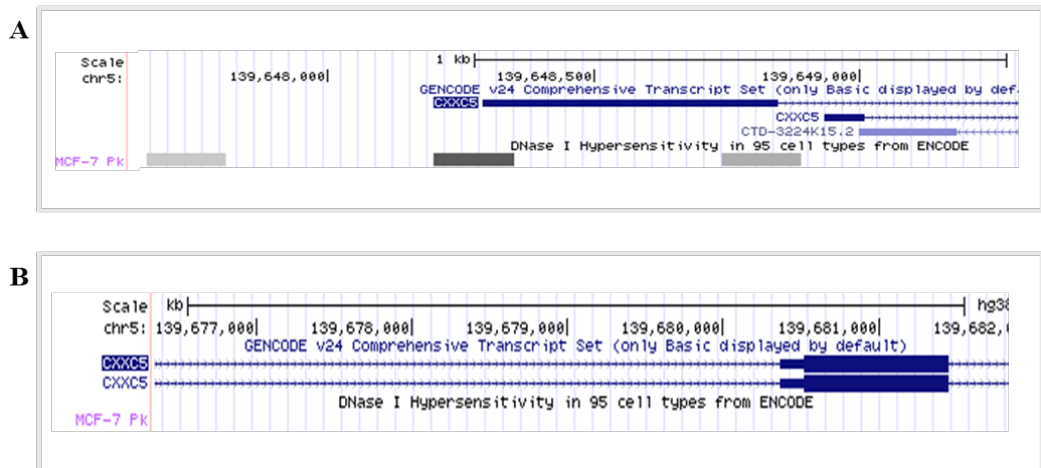


Figure 10: DNase I Hypersensitivity Data from UCSC-Genome Browser for Promoter and ERE region of *CXXC5*.

The track named as MCF-7 Pk shows DNase I Peaks from ENCODE database. 5' promoter region of *CXXC5* has a DNase I hypersensitive region in MCF7 cells (A) while the enhancer ERE region juxtaposed to Exon2 does not (B). This information suggests that the promoter of *CXXC5* may have nucleosome free regions but not the enhancer region.

3.2. Control of E2 Treatment on *CXXC5* Expression

Previous studies from our laboratory showed that *CXXC5* is upregulated in the presence of E2, which is detectable at 3 and 24-hour treatments [38]. To reiterate this finding, we carried studies in MCF7 cells, grown in CD-FBS containing medium, DMEM, for 48 h to remove steroid hormones, and treated without (EtOH) or with 10^{-8} M E2 for 3h. Cells were then subjected to RNA isolation and RT-qPCR. qPCR was carried out with 3 biological replicates in triplicates. For the calculation of fold change, the expression of *CXXC5* was normalized to that of *RPLP0*. EtOH control was set to 1 and fold change of E2 treatments are statistically analyzed with one-tailed, paired t-test of mean with 95% confidence interval (Figure 11) (Microsoft Excel version of 2016).

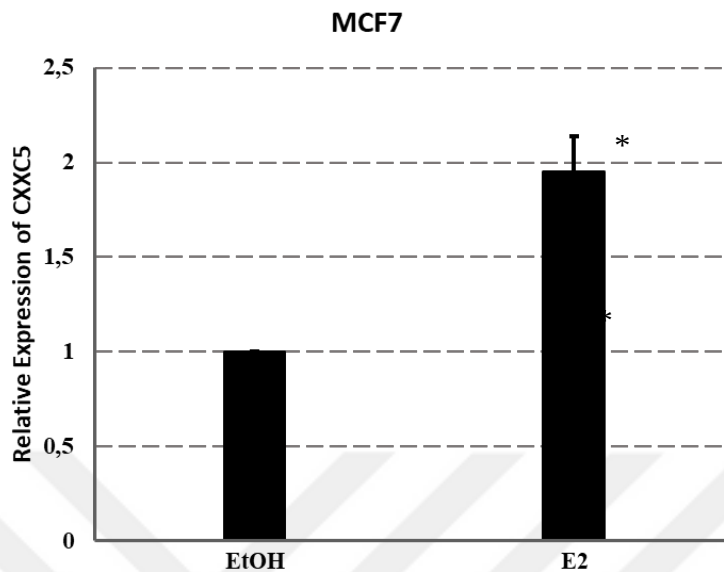


Figure 11: Upregulation of CXXC5 After E2 treatment.

MCF7 cells were grown in CD-FBS containing DMEM medium for 48 hours. Cells were then treated with E2 (10^{-8} M) or vehicle control EtOH for 3 hours. After total RNA isolation and cDNA conversion, qPCR was performed with samples from three biological replicates, each biological replicate contained three technical repeats. Expression is normalized to *RPLP0* ($p < 0.05$; one tailed, paired t-test with 95% confidence interval).

3.3 Analysis of Endogenous Methylation Pattern of the CXXC5 Promoter and the Enhancer Region

Since E2 augments *CXXC5* expression at 3h, we wanted to initially assess whether or not the change in the gene expression is reflected as changes in the methylation pattern of the regulatory regions of *CXXC5*. For the detection of endogenous methylation, we used MCF7 cells treated without (EtOH) or with E2 for 3h. In addition to MCF7 cell line, we used HL60 cell line derived from human leukemia in which the expression of *CXXC5* gene is first studied [41] HL60 cells do not synthesize ER α and consequently are unresponsive to E2 (unpublished observation). This allow us to comparatively assess whether methylation profiling of the selected *CXXC5* loci show ligand and/or cell-type dependent differences.

In order to explore the endogenous methylation profile of DNA, we used bisulfite-sequencing method, which involves successive chemical reactions on isolated gDNA.

Firstly, cytosine of DNA is sulfonated during bisulfite reaction. Then, de-amination and de-sulfonation reactions occur and cytosine is converted to uracil (Figure 12) [69].

The methylated cytosine is protected from the bisulfite reaction. Hence, the methylated cytosine cannot be converted to uracil. Bisulfite conversion disrupts double stranded structure of DNA. Due to the base substitution, DNA strands are no longer complementary. For this reason, purified DNA after bisulfite reaction is assessed with RNA parameters.

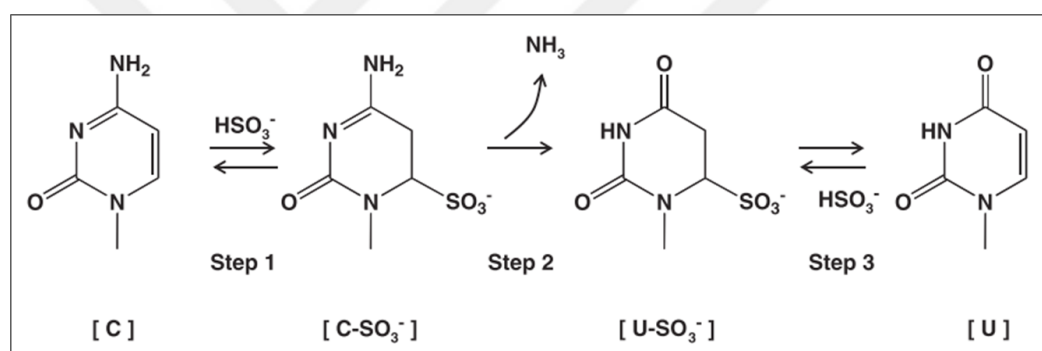


Figure 12: Bisulfite Conversion of Cytosine to Uracil.

First step is represented sulfonation reaction of cytosine. Following reactions are de-amination and de-sulfonation respectively [69].

The purified bisulfite DNA is then used as the template in PCR reactions for a targeted analysis of DNA methylation.

Using PCR, we generated overlapping 400-600 nucleotide long PCR products of the putative promoter and enhancer regions of *CXXC5*. PCR products were then cloned into PGEM-T Vector (Promega Corporation, A3600) and positive colonies were sequenced with T7 or SP6 universal primers. In order to increase statistical

relevance of the results, we sequenced 10 positive clones from 2 biological replicates of EtOH/E2 treatments for both 5' region of the putative promoter and enhancer of *CXXC5*. In addition, we also analyzed the whole exon 2, as control of the methylated gene body region (Data shown in Appendix). Bisulfite sequences were aligned to the genomic sequence of the corresponding *CXXC5* regions using MethylViewer Software [64]. CpG dinucleotides were analyzed to show whether or not cytosine is converted to thymine. We considered that cytosine residues converted to thymine are unmethylated while unconverted ones are methylated.

3.3.1. Methylation Profile of the Putative Promoter of *CXXC5*

We analyzed CpG sites between -873rd and 180th nucleotides, according to the first nucleotide of the Exon 1. As a part of the CpG island, this region contains also the predicted promoter of *CXXC5* as mentioned in section 3.1. *In Silico* Locus Analysis.

While we present the detailed analysis of the methylation profile, as lollipop graph, of the examined loci in Appendix F, we here summarize the results here as a line graph (Excel 2016, Microsoft Corporation, USA) that depicts methylation percentage of CpG positions calculated as $[(Methylated\ CpG\ \# / total\ sequenced\ clone) * 100]$ (Figure 13). According to our results, the 5' region of the *CXXC5* Exon 1 shows high degree of CpG methylation which declines precipitously towards the beginning of Exon 1. There was no significant change between samples treated without (EtOH) or with E2.

Methylation of 5' Region of CXXC5 Encompassing Promoter

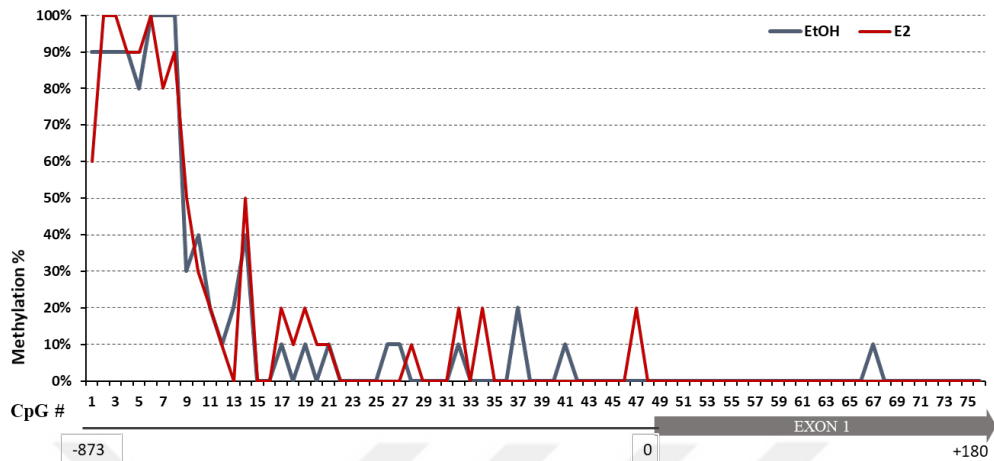


Figure 13: CpG Methylation of the 5' loci of *CXXC5*.

Each CpG dinucleotide is represented as numbered positions. Blue line corresponds to EtOH while red line corresponds to E2 treated samples. From 10 different clones, methylation percentages are calculated individually for each dinucleotide position. It is observed that towards the 1st exon, the 5' region of *CXXC5* progressively becomes unmethylated independently from E2.

These results suggest that E2 does not have an effect on endogenous methylation profile of the 5' region of the *CXXC5* promoter. It should be noted that there can be dynamic changes at some of the CpG positions in response to E2 that we could not detect at 3h. A time course approaches could be useful to detect these type of changes.

The same protocol was also used for HL60 leukemia cell line maintained at steady-state growth condition. As mentioned, the putative promoter of *CXXC5* was deduced from series of *in silico* analysis from genome wide data and preliminary studies conducted on MCF7 cells in laboratory. We wanted to examine whether the endogenous methylation profile of the putative CGI promoter of *CXXC5* in HL60 cells resembles to that of MCF7 cells or there are cell type-dependent differences. According to our results, HL60 also shows the same pattern of methylation observed in MCF7 cells (Appendix F).

3.3.2. Methylation Profile of the ERE Containing Enhancer Region and the Exon 2 of CXXC5

A functional ERE located at 240 bp upstream of translation start site. E2 treatment could induce alterations in DNA methylation of this region. To investigate this possibility, CpG sites at the 5' region of the exon 2 (from -513th nucleotide to +240th) was analyzed. 10 clones from two biological replicates were sequenced. Methylation percentage of each CpG site corresponding to analyzed region is presented in line graph below (Figure 14); a detailed lollipop graph of each sequenced clone is also presented in Appendix F.

Results show that the analyzed region is highly methylated and E2 treatment does not have any significant effect on DNA methylation. Except for the first exons, exon methylation is a common phenomenon and it is thought to be involved in regulation of co-transcriptional splicing events and protect genes from intragenic transcriptional initiation [69]. Similarly, the same regulatory region was highly methylated in HL60 cells (Appendix F).

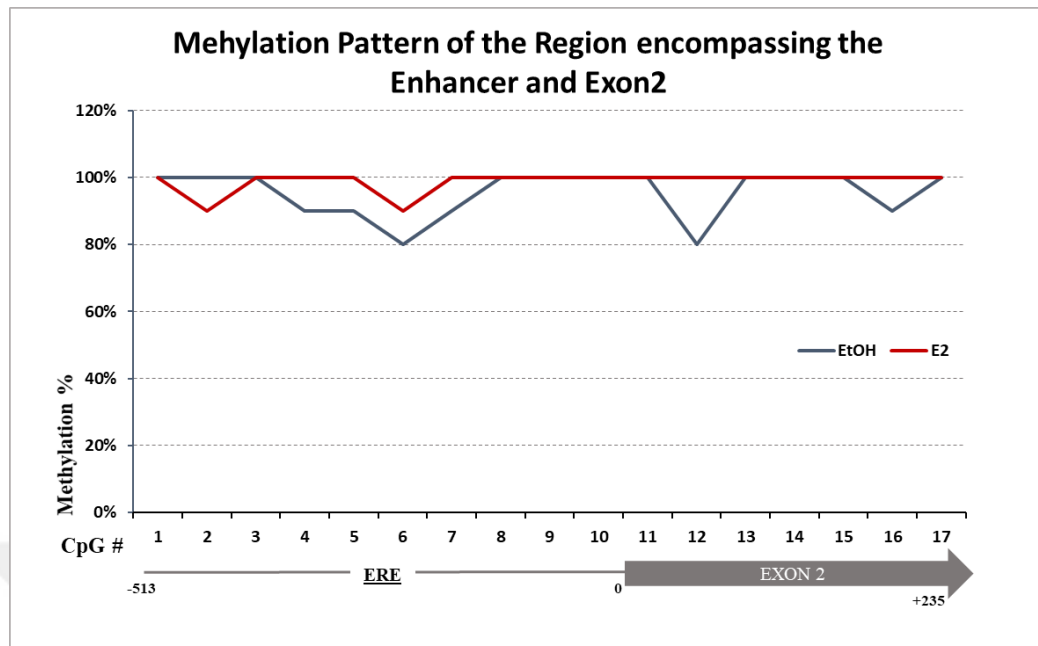


Figure 14: Methylation Profile of the Region Encompassing Enhancer.

Each CpG dinucleotide from 10 different clones is represented as numbered positions and methylation percentages are calculated individually for each. Blue line corresponds to EtOH while red line corresponds to E2. Methylation percentage of each CpG site in MCF7 cells is between 80-100 independently from E2.

3.4. Investigation of Nucleosome Positions

Nucleosomes as basic units of chromatin structure are important in gene regulation. Thereby, understanding of nucleosome positioning during gene expression is crucial. To investigate nucleosome positioning of the putative promoter and the enhancer region of *CXXC5* in the absence/presence of E2, we carried out Nucleosome Occupancy and Methylome (NOME) Sequencing.

NOME uses a GpC Methylase enzyme of bacterial origin, which is absent in vertebrates, to exogenously methylate GpC dinucleotides of DNA, differently from endogenous methylation that occurs on CpG dinucleotides. After protein-chromatin interaction is fixed, nuclei are isolated and incubated with GpC Methylase. The enzyme can methylate only available GpC dinucleotides. Nucleosomes or any other proteins fixed on DNA prevent accessing of the

methylase to dinucleotides. Bisulfite reaction is used again to discriminate between methylated and unmethylated cytosine. After bisulfite reaction, purified DNA is used as the template to amplify target regions by PCR.

Because there can be GpCpG triplets in DNA, the possible endogenous methylation profile of triplets should also be considered during analysis. Some GpC sites might be methylated because its cytosine actually belongs to endogenously methylated CpG dinucleotides. In this case, methylation does not indicate an open chromatin structure.

HaeIII is a methyl sensitive enzyme which cleaves GG[^]CC sequence. It cannot cleave DNA when GpC is methylated. To ensure the NOME-processed GpC sites are methylated with the experimental approach we are using, we carried out a GpC Methylase reaction using linearized pBS-KS vector. Since the GpC sites of the purified vector DNA is not occupied with any protein, the methylase should add methyl groups to any GpC that it encounters. After methylase reaction, an enzyme digestion with HaeIII (NEB, USA, R0108S) was conducted on samples. Digestion efficiency was analyzed on agarose gel. We detected a very efficient GpC Methylase activity: While the vector DNA not subjected to methylation reaction was efficiently cleaved with the enzyme, the enzyme had no effect on the methylated vector (Result are showed in Appendix).

For NOME analysis, 3h EtOH/E2 treated cells were fixed and experiment was conducted as indicated. After bisulfite PCR, we cloned PCR products into PGEM-T vector for sanger sequencing.

3.4.1. Nucleosome Positioning of the Putative Promoter of *CXXC5*

Nucleosome profiles at promoter regions of genes can provide information about the regulatory state of gene expression. There might be nucleosome free regions that are available to transcription machinery or several regions might become nucleosome free because of signaling cascades to regulate transcription.

The putative *CXXC5* promoter region was investigated to understand its nucleosome profile and the effect of E2 on this profile. The same regions were amplified as described in endogenous methylation analysis. Nine clones for EtOH and 10 for E2 were sequenced. Methylation of GpC dinucleotides were calculated as percentage by $[(Methylated\ GpC\ \# / total\ sequenced\ clone) * 100]$. A line graph was generated with Excel (Version 2016, Microsoft Corporation, USA). Detailed profile of each sample is represented as a lollipop graph at Appendix F.

According to our results, exogenous methylation was rarely observable at the 5' region of *CXXC5* in the absence or presence of E2. Since GpC sites are exogenously methylated when available, methylated sites are presumed to be nucleosome free. Consequently, our results suggest that the 5' region of *CXXC5* is occupied. Nevertheless, we could not determine precise nucleosome positions at these regions, which show a very high GC dinucleotide content. This region could also be occupied with proteins bound to DNA and/or nucleosomes that prevent accessibility to exogenous methylation. There are methylated sites observed in all sequenced samples (GpC positions 10-23) due to the endogenously methylated CpG dinucleotides that cannot be thought as nucleosome free. Only a short length of DNA containing repetitive GC sequence (between 75th and 91st GpC positions) shows 50% exogenous methylation level in EtOH samples indicating that the region may be unoccupied. Among E2 samples, exogenous methylation at these sites was not observed (Figure 15, also see lollipop graph at Appendix). 75th -91st GpC sites are located at the beginning of the exon 1 and our data is correlated with genome wide DNase I Hypersensitivity data described in section 3.1. *Locus Analysis*.

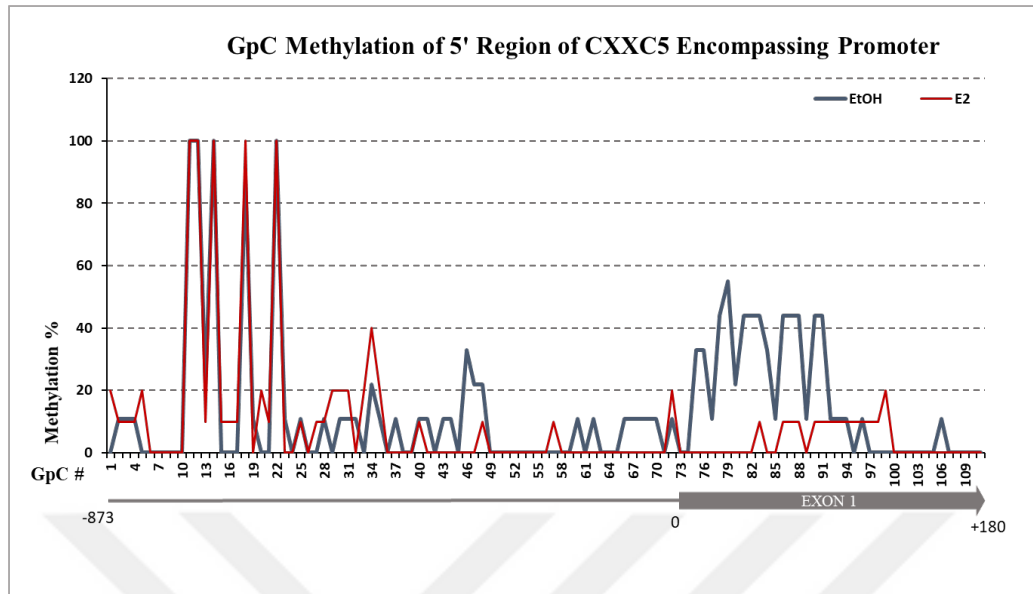


Figure 15. Nucleosome Analysis of the 5' region of CXXC5.

The percent methylation profiles of GpC dinucleotides obtained with NOME are represented. Blue line indicates EtOH and red line indicates E2 treated samples. Since GpC sites are exogenously methylated when DNA is available, in NOME assay, methylated sites are presumed to be nucleosome free. GpC dinucleotides positioned between 10 and 23 corresponds to GCG triplets methylated due to endogenously methylated CpG, thus cannot be accepted as free DNA while, GpC positioned between 75-91 are exogenously methylated in EtOH treated samples, indicating possible unoccupied DNA.

NOME results for the indicated exogenously methylated region might have biological significance. This indeed may be case, because the same nucleosome positioning assay in E2-unresponsive HL60 cells display the same NOME profile to that of the MCF7 cells in the absence of E2. When we carried out nucleosome positioning assay on HL60 cells, we identified the same region as possible nucleosome free structure observed in EtOH, but not E2, treated MCF7 cells (Appendix F).

3.4.2. Nucleosome Positioning Around the Putative Enhancer

The second exon of *CXXC5* is GC rich, contains translation start site and has an upstream functional ERE. In order to assess whether or not there are changes in

nucleosome positioning at this region in response to E2, NOME analysis was carried out. Region from -513th nucleotide upstream to +235th nucleotide of the exon 2 was analyzed. The percent methylation of each GpC dinucleotides was calculated according to $[(Methylated\ GpC\ \# / total\ sequenced\ clone) * 100]$ formula and a line graph was generated with Excel (Version 2016, Microsoft Corporation). We found that there was no significant exogenous methylation in samples treated without (EtOH) or with E2. The GpC positions between 11-13, 36-40, 42-46, 57-60 and 67-69 are indicated as 100% methylated; however, these sites have GpCpG triplets that seem to be methylated because of endogenously methylated CpG (Figure 15). We also present a lollipop graph combining endogenous methylation together with exogenous GpC methylation in Appendix F.

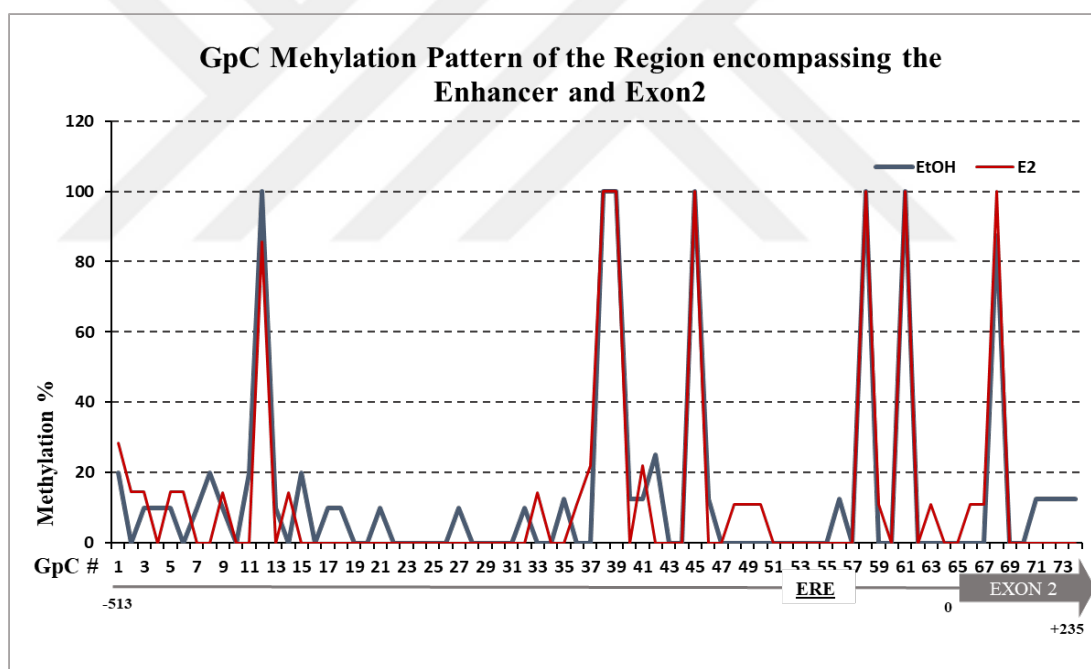


Figure 16. GpC Methylation Profile of the Region Encompassing the Enhancer and Exon2 of CXXC5.

GpC positions at enhancer and 5' of exon 2 was analyzed in terms of exogenous methylation as an indication of open chromatin structure. GpC positions that have 100% methylation are GCG triplets and methylation appears due to CpG sites that were shown to be methylated endogenously. Other positions do not have a significant level of GpC methylation. Thus, result indicates that the analyzed region is occupied.

This result suggests that the region of interest might be occupied with nucleosomes or other proteins, correlating with DNase I Hypersensitive data referred in section 3.1. *In Silico Analysis*.



CHAPTER 4

CONCLUSION AND FUTURE DIRECTIONS

Previous studies from this and other laboratories indicate that the expression of *CXXC5* is mediated in response to various signaling pathway. Importantly, the de-regulated expression of *CXXC5* appears to be a critical prognostic factor for various disease state and malignancies including breast cancer. Since *CXXC5* is also an estrogen responsive gene, deciphering mechanisms underlying the expression of *CXXC5* could contribute to a better understanding of the role of the *CXXC5* protein in physiology and pathophysiology of E2-ER α signaling. By using MCF7 cell line as a cell model for estrogen receptor positive breast cancer in comparison with ER α -negative, hence E2 unresponsive, HL60 cells derived from an acute promyelocytic leukemia, we here assessed whether or not a short duration (3h) E2-ER α signaling alters DNA methylation and nucleosome occupancy of the putative promoter and enhancer regions of *CXXC5*.

Our *in silico* and experimental results can be summarized as follows:

1. The putative *CXXC5* promoter is located within a CGI characterized by high density G and C nucleotides, CpG dinucleotides and GC skew. Based on our bisulfite sequencing results, we conclude that the *CXXC5* promoter displays a hypomethylated state, in contrast to the hypermethylated gene body including the enhancer regions and exon 2 independent of cell-type.

2. E2 does not change the DNA methylation pattern of the promoter nor does the enhancer region of *CXXC5* in MCF7 cells, which have a methylation pattern similar to that observed in HL60 cells.
3. While very effective in assessing the nucleosome occupancy in non-CGI promoters, Nucleosome occupancy assay (NOME) in our hands provided limited information at the *CXXC5* locus due to remarkably high GC content. However, it appears that E2 treatment could alter the nucleosome occupancy of the promoter region of *CXXC5* without affecting that of the enhancer region in MCF7 cells, which have nucleosome occupancy profile in the absence of E2 similar to that of HL60 cells grown under a steady-state condition. Our findings suggest that the augmentation of the *CXXC5* expression by E2-ER α through a distal enhancer alters nucleosome occupancy by a mechanism involving nucleosome sliding or stabilization at the *CXXC5* promoter or DNA becomes occupied with transcription-related proteins.

In general, we could define DNA methylation and nucleosome properties of *CXXC5* putative regulatory sites. However, a mechanistic explanation for the enhanced expression in response to E2 is yet to be further investigated.

There could be some dynamic changes of specific CpG dinucleotides that we might have missed with our experimental approach that assess single treatment point (3h). Consequently, approaches that dynamically assess DNA methylation and nucleosome occupancy at both the promoter and enhancer would be useful.

Currently, we are conducting Micrococcal Nuclease Assay to further define and refine nucleosome occupied/free positions as an additional method to NOME. Other approaches could also be used to assess the nucleosome occupancy in detailed at the *CXXC5* locus. These could include Histone core chromatin immunoprecipitation (ChIP) assays that might provide quantitative results by qPCR analysis.

Beside occupancy, E2 signaling induces histone tail modifications of nucleosomes as well. *CXXC5* might be regulated by several histone modifications in the presence of E2 without affecting DNA methylation and/or nucleosome occupancy. To identify these histone modifications in the absence/presence of E2 might contribute to the mechanism by which E2 mediates the expression of *CXXC5*.

Another question currently investigated by our laboratory is that how could a distant ERE located at the end of the very long intron (app. 30kb) regulates transcription of *CXXC5*. One of the hypothesis is that there could be a chromatin loop formation that brings the enhancer and promoter to a close proximity. Kerim Yavuz, one of the members of our laboratory, is currently using a chromatin capture assay (3C) to test this possibility.



REFERENCES

- [1] M. Marino, P. Galluzzo, and P. Ascenzi, “Estrogen signaling multiple pathways to impact gene transcription,” *Curr Genomics*, vol. 7, no. 8, pp. 497–508, 2006.
- [2] P. Yaşar, G. Ayaz, S. D. User, G. Güpür, and M. Muyan, “Molecular mechanism of estrogen–estrogen receptor signaling,” *Reprod. Med. Biol.*, vol. 16, no. 1, pp. 4–20, 2017.
- [3] N. Heldring, A. Pike, S. Andersson, J. Matthews, G. Cheng, E. Treuter, M. Warner, J. Hartman, M. Tujague, and A. Stro, “Estrogen Receptors : How Do They Signal and What Are Their Targets,” *Physiol. Rev.*, vol. 87, pp. 905–931, 2007.
- [4] L. R. Nelson and S. E. Bulun, “Estrogen production and action,” *J. Am. Acad. Dermatol.*, vol. 45, pp. S116–S124, 2001.
- [5] Y. Huang, X. Li, and M. Muyan, “Estrogen receptors similarly mediate the effects of 17 β -estradiol on cellular responses but differ in their potencies,” *Endocrine*, vol. 39, no. 1, pp. 48–61, 2011.
- [6] 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, 2002.
- [7] J. Huang, X. Li, T. Qiao, R. a Bambara, R. Hilf, and M. Muyan, “A tale of two estrogen receptors (ERs): how differential ER-estrogen responsive element interactions contribute to subtype-specific transcriptional responses.,” *Nucl. Recept. Signal.*, vol. 4, p. e015, 2006.

- [8] Mmuyan, “The Ligand-Mediated Nuclear Mobility and Interaction with Estrogen Responsive Elements of Estrogen Receptors are Subtype-Specific,” vol. 6, no. 3, pp. 247–253, 2009.
- [9] C. Y. Lin, V. B. Vega, J. S. Thomsen, T. Zhang, S. L. Kong, M. Xie, K. P. Chiu, L. Lipovich, D. H. Barnett, F. Stossi, A. Yeo, J. George, V. A. Kuznetsov, Y. K. Lee, T. H. Charn, N. Palanisamy, L. D. Miller, E. Cheung, B. S. Katzenellenbogen, Y. Ruan, G. Bourque, C. L. Wei, and E. T. Liu, “Whole-genome cartography of estrogen receptor alpha binding sites.,” *PLoS Genet.*, vol. 3, no. 6, p. e87, 2007.
- [10] C. M. Klinge, “Estrogen receptor interaction with estrogen response elements,” *Nucleic Acids Res.*, vol. 29, no. 14, pp. 2905–2919, 2001.
- [11] P. Yi, M. D. Driscoll, J. Huang, S. Bhagat, R. Hilf, R. a Bambara, and M. Muyan, “The effects of estrogen-responsive element- and ligand-induced structural changes on the recruitment of cofactors and transcriptional responses by ER alpha and ER beta.,” *Mol. Endocrinol.*, vol. 16, no. 4, pp. 674–693, 2002.
- [12] I. H. Russo and J. Russo, “Role of hormones in mammary cancer initiation and progression.,” *J. Mammary Gland Biol. Neoplasia*, vol. 3, no. 1, pp. 49–61, 1998.
- [13] S. J. Dawson, O. M. Rueda, S. Aparicio, and C. Caldas, “A new genome-driven integrated classification of breast cancer and its implications,” *EMBO J.*, vol. 32, no. 5, pp. 617–628, 2013.
- [14] X.-H. Liao, D.-L. Lu, N. Wang, L.-Y. Liu, Y. Wang, Y.-Q. Li, T.-B. Yan, X.-G. Sun, P. Hu, and T.-C. Zhang, “Estrogen receptor α mediates proliferation of breast cancer MCF-7 cells via a p21/PCNA/E2F1-dependent pathway,” *FEBS J.*, vol. 281, no. 3, pp. 927–942, 2014.

- [15] G. Sathya, P. Yi, S. Bhagat, R. A. Bambara, and R. Hilf, “Structural regions of ER α critical for synergistic transcriptional responses contain co-factor interacting surfaces,” *Mol. Cell. Endocrinol.*, vol. 192, no. 1–2, pp. 171–185, 2002.
- [16] R. Métivier, G. Penot, M. R. Hübner, G. Reid, H. Brand, M. Koš, and F. Gannon, “Estrogen receptor- α directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter,” *Cell*, vol. 115, no. 6, pp. 751–763, 2003.
- [17] J. Turek-Plewa and P. P. Jagodzinski, “The role of mammalian DNA methyltransferases in the regulation of gene expression,” *Cell. Mol. Biol. Lett.*, vol. 10, no. 4, pp. 631–647, 2005.
- [18] S. Murakami, A. Nagari, and W. L. Kraus, “Dynamic assembly and activation of estrogen receptor α enhancers through coregulator switching,” *Genes Dev.*, vol. 31, no. 15, pp. 1535–1548, 2017.
- [19] T. O. Tollefsbol, *Epigenetics: The new science of genetics*, First Edit. Elsevier Inc., 2011.
- [20] R. J. Klose and A. P. Bird, “Genomic DNA methylation: The mark and its mediators,” *Trends Biochem. Sci.*, vol. 31, no. 2, pp. 89–97, 2006.
- [21] J. T. Attwood, R. L. Yung, and B. C. Richardson, “DNA methylation and the regulation of gene transcription,” *Cell. Mol. Life Sci. C.*, vol. 59, no. 2, pp. 241–257, 2014.
- [22] R. Métivier, R. Gallais, C. Tiffoche, C. Le Péron, R. Z. Jurkowska, R. P. Carmouche, D. Ibberson, P. Barath, F. Demay, G. Reid, V. Benes, A. Jeltsch, F. Gannon, and G. Salbert, “Cyclical DNA methylation of a transcriptionally active promoter,” *Nature*, vol. 452, no. 7183, pp. 45–50, 2008.

- [23] S. Kangaspeska, B. Stride, R. Métivier, M. Polycarpou-Schwarz, D. Ibberson, R. P. Carmouche, V. Benes, F. Gannon, and G. Reid, “Transient cyclical methylation of promoter DNA.,” *Nature*, vol. 452, no. 7183, pp. 112–5, 2008.
- [24] E. Hervouet, P.-F. Cartron, M. Jouvenot, and R. Delage-Mourroux, “Epigenetic regulation of estrogen signaling in breast cancer.,” *Epigenetics*, vol. 8, no. 3, pp. 237–45, 2013.
- [25] K. P. Singh, J. Treas, T. Tyagi, and W. Gao, “DNA demethylation by 5-aza-2-deoxycytidine treatment abrogates 17 beta-estradiol-induced cell growth and restores expression of DNA repair genes in human breast cancer cells,” *Cancer Lett.*, vol. 316, no. 1, pp. 62–69, 2012.
- [26] R. D. Kornberg, “Chromatin Structure : A Repeating Unit of Histones and DNA,” *Science (80-.)*, vol. 184, pp. 868–871, 1974.
- [27] V. Ramakrishnan, “HISTONE STRUCTURE AND THE ORGANIZATION OF THE NUCLEOSOME,” *Annu. Rev. Biophys. Biomol. Struct.*, vol. 26, pp. 83–112, 1997.
- [28] A. J. Bannister and T. Kouzarides, “Regulation of chromatin by histone modifications,” *Nat. Publ. Gr.*, vol. 21, no. 3, pp. 381–395, 2011.
- [29] C. D. Allis and T. Jenuwein, “The molecular hallmarks of epigenetic control,” *Nat. Rev. Genet.*, vol. 17, no. 8, pp. 487–500, 2016.
- [30] J. A. Knezetic and D. S. Luse, “The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro,” *Cell*, vol. 45, no. 1, pp. 95–104, 1986.
- [31] Q. Lu, L. L. Wallrath, and S. C. R. Elgin, “The role of a positioned nucleosome at the *Drosophila melanogaster* hsp26 promoter,” *EMBO*, vol. 14, no. 1, pp. 4738–4746, 1995.

- [32] V. Berno, L. Amazit, C. Hinojos, J. Zhong, M. G. Mancini, Z. Dave, and M. A. Mancini, “Activation of Estrogen Receptor- α by E2 or EGF Induces Temporally Distinct Patterns of Large-Scale Chromatin Modification and mRNA Transcription,” *PLoS One*, vol. 3, no. 5, 2008.
- [33] G. F. Sewack and U. Hansen, “Nucleosome positioning and transcription-associated chromatin alterations on the human estrogen-responsive pS2 promoter,” *J. Biol. Chem.*, vol. 272, no. 49, pp. 31118–31129, 1997.
- [34] C. Giamarchi, M. Solanas, C. Chailleux, P. Augereau, F. Vignon, H. Rochefort, and H. Richard-Foy, “Chromatin structure of the regulatory regions of pS2 and cathepsin D genes in hormone-dependent and -independent breast cancer cell lines,” *Oncogene*, vol. 18, no. 2, pp. 533–541, 1999.
- [35] K. W. Jeong, C. Andreu-Vieyra, J. S. You, P. A. Jones, and M. R. Stallcup, “Establishment of active chromatin structure at enhancer elements by mixed-lineage leukemia 1 to initiate estrogen-dependent gene expression,” *Nucleic Acids Res.*, vol. 42, no. 4, pp. 2245–2256, 2014.
- [36] X. Zhang, M. Bolt, M. J. Guertin, W. Chen, S. Zhang, B. D. Cherrington, and D. J. Slade, “Peptidylarginine 26 citrullination facilitates estrogen receptor α target gene activation,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 33, pp. 13331–13336, 2012.
- [37] 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 α are required to elicit cellular alterations,” *J. Biol. Chem.*, vol. 284, no. 22, pp. 15277–88, May 2009.
- [38] P. Yaşar, G. Ayaz, and M. Muyan, “Estradiol-Estrogen Receptor α Mediates the Expression of the CXXC5 Gene through the Estrogen Response

- Element-Dependent Signaling Pathway,” *Sci. Rep.*, vol. 6, no. July, pp. 1–12, 2016.
- [39] W. J. Kent, C. W. Sugnet, T. S. Furey, and K. M. Roskin, “The Human Genome Browser at UCSC,” *Cold Spring Harb. Lab. Press*, vol. 12, pp. 996–1006, 2002.
- [40] A. Bateman, M. J. Martin, C. O’Donovan, M. Magrane, E. Alpi, R. Antunes, B. Bely, M. Bingley, C. Bonilla, R. Britto, B. Bursteinas, H. Bye-AJee, A. Cowley, A. Da Silva, M. De Giorgi, T. Dogan, F. Fazzini, L. G. Castro, L. Figueira, P. Garmiri, G. Georghiou, D. Gonzalez, E. Hatton-Ellis, W. Li, W. Liu, R. Lopez, J. Luo, Y. Lussi, A. MacDougall, A. Nightingale, B. Palka, K. Pichler, D. Poggioli, S. Pundir, L. Pureza, G. Qi, S. Rosanoff, R. Saidi, T. Sawford, A. Shypitsyna, E. Speretta, E. Turner, N. Tyagi, V. Volynkin, T. Wardell, K. Warner, X. Watkins, R. Zaru, H. Zellner, I. Xenarios, L. Bougueleret, A. Bridge, S. Poux, N. Redaschi, L. Aimo, G. ArgoudPuy, A. Auchincloss, K. Axelsen, P. Bansal, D. Baratin, M. C. Blatter, B. Boeckmann, J. Bolleman, E. Boutet, L. Breuza, C. Casal-Casas, E. De Castro, E. Coudert, B. CuChe, M. Doche, D. Dornevil, S. Duvaud, A. Estreicher, L. Famiglietti, M. Feuermann, E. Gasteiger, S. Gehant, V. Gerritsen, A. Gos, N. Gruaz-Gumowski, U. Hinz, C. Hulo, F. Jungo, G. Keller, V. Lara, P. Lemercier, D. Lieberherr, T. Lombardot, X. Martin, P. Masson, A. Morgat, T. Neto, N. Nospikel, S. Paesano, I. Pedruzzi, S. Pilbout, M. Pozzato, M. Pruess, C. Rivoire, B. Roechert, M. Schneider, C. Sigrist, K. Sonesson, S. Staehli, A. Stutz, S. Sundaram, M. Tognolli, L. Verbregue, A. L. Veuthey, C. H. Wu, C. N. Arighi, L. Arminski, C. Chen, Y. Chen, J. S. Garavelli, H. Huang, K. Laiho, P. McGarvey, D. A. Natale, K. Ross, C. R. Vinayaka, Q. Wang, Y. Wang, L. S. Yeh, and J. Zhang, “UniProt: The universal protein knowledgebase,” *Nucleic Acids Res.*, vol. 45, no. D1, pp. D158–D169, 2017.

- [41] 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. 3172–3181, 2009.
- [42] H. Zhou, S. Di Palma, C. Preisinger, M. Peng, A. N. Polat, A. J. R. Heck, and S. Mohammed, “Toward a comprehensive characterization of a human cancer cell phosphoproteome,” *J. Proteome Res.*, vol. 12, no. 1, pp. 260–271, 2013.
- [43] 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, 2013.
- [44] M. B. Treppendahl, L. Möllgård, E. Hellström-Lindberg, P. Cloos, and K. Grønbæk, “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. 259–260, 2013.
- [45] A. Astori, H. Fredly, T. Aloysius, L. Bullinger, V. M.-D. Mas, P. de la Grange, F. Delhommeau, K. M. Hagen, C. Récher, I. Dusanter-Fourt, S. Knappskog, J. R. Lillehaug, F. Pendino, and Ø. Bruserud, “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.
- [46] A. Kühnl, P. J. M. Valk, M. A. Sanders, A. Ivey, R. K. Hills, K. I. Mills, R. E. Gale, M. F. Kaiser, R. Dillon, M. Joannides, A. Gilkes, T. Haferlach, S. Schnittger, E. Duprez, D. C. Linch, R. Delwel, B. Löwenberg, C. D. Baldus, E. Solomon, A. K. Burnett, and D. Grimwade, “Downregulation of the Wnt inhibitor CXXC5 predicts a better prognosis in acute myeloid leukemia,” *Blood*, vol. 125, no. 19, pp. 2985–2994, 2015.

- [47] H. Kim, J. Yoon, J. Yun, K. Cho, S. Lee, Y. Rhee, H. Jung, H. J. Lim, H. Lee, J. Choi, J. Heo, W. Lee, K. T. No, D. Min, and K. Choi, “CXXC5 is a negative-feedback regulator of the Wnt/ β -catenin pathway involved in osteoblast differentiation,” *Cell Death Differ.*, vol. 22, pp. 912–920, 2015.
- [48] S.-H. Lee, M.-Y. Kim, H.-Y. Kim, Y.-M. Lee, H. Kim, K. A. Nam, M. R. Roh, D. S. Min, K. Y. Chung, and K.-Y. Choi, “The Dishevelled-binding protein CXXC5 negatively regulates cutaneous wound healing,” *J. Exp. Med.*, vol. 212, no. 7, pp. 1061–1080, 2015.
- [49] I. Lee, S. Choi, J. H. Yun, S. hwa Seo, S. Choi, K. Y. Choi, and W. Lee, “Crystal structure of the PDZ domain of mouse Dishevelled 1 and its interaction with CXXC5,” *Biochem. Biophys. Res. Commun.*, vol. 485, no. 3, pp. 584–590, 2017.
- [50] H.-Y. Kim, S. Choi, J.-H. Yoon, H. J. Lim, H. Lee, J. Choi, E. J. Ro, J.-N. Heo, W. Lee, K. T. No, and K.-Y. Choi, “Small molecule inhibitors of the Dishevelled-CXXC5 interaction are new drug candidates for bone anabolic osteoporosis therapy,” *EMBO Mol. Med.*, vol. 8, no. 4, pp. 375–87, 2016.
- [51] 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.*, vol. 28, no. 2, pp. 615–626, 2014.
- [52] M. Y. Kim, H. Y. Kim, J. Hong, D. Kim, H. Lee, E. Cheong, Y. Lee, J. Roth, D. G. Kim, D. S. Min, and K. Y. Choi, “CXXC5 plays a role as a transcription activator for myelin genes on oligodendrocyte differentiation,” *Glia*, vol. 64, no. 3, pp. 350–362, 2016.
- [53] G. Li, X. Ye, X. Peng, Y. Deng, W. Yuan, Y. Li, X. Mo, X. Wang, Y. Wan, X. Liu, T. Chen, Z. Jiang, X. Fan, X. Wu, and Y. Wang, “CXXC5 regulates

- differentiation of C2C12 myoblasts into myocytes,” *J. Muscle Res. Cell Motil.*, vol. 35, no. 5–6, pp. 259–265, 2014.
- [54] S. Aras, O. Pak, N. Sommer, R. Finley, M. Hü Ttemann, 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. 2255–2266, 2013.
- [55] S. Ma, X. Wan, Z. Deng, L. Shi, C. Hao, Z. Zhou, C. Zhou, Y. Fang, J. Liu, J. Yang, X. Chen, T. Li, A. Zang, S. Yin, B. Li, J. Plumas, L. Chaperot, X. Zhang, G. Xu, L. Jiang, N. Shen, S. Xiong, X. Gao, Y. Zhang, and H. Xiao, “Epigenetic regulator CXXC5 recruits DNA demethylase Tet2 to regulate TLR7/9-elicited IFN response in pDCs,” *J. Exp. Med.*, vol. 214, no. 5, pp. 1471–1491, 2017.
- [56] Y. Tsuchiya, T. Naito, M. Tenno, M. Maruyama, H. Koseki, I. Taniuchi, and Y. Naoe, “ThPOK represses CXXC5, which induces methylation of histone H3 lysine 9 in *Cd40lg* promoter by association with SUV39H1: implications in repression of CD40L expression in CD8⁺ cytotoxic T cells,” *J. Leukoc. Biol.*, vol. 100, no. 2, pp. 327–338, 2016.
- [57] X. Wang, P. Liao, X. Fan, Y. Wan, Y. Wang, Y. Li, Z. Jiang, X. Ye, X. Mo, K. Ocorr, Y. Deng, X. Wu, and W. Yuan, “CXXC5 Associates with Smads to Mediate TNF-alpha Induced Apoptosis,” *Curr. Mol. Med.*, vol. 13, no. 8, pp. 1385–1396, 2013.
- [58] X. Yan, J. Wu, Q. Jiang, H. Cheng, J. J. Han, and Y. Chen, “CXXC5 suppresses hepatocellular carcinoma by promoting TGF- β -induced cell cycle arrest and apoptosis,” *J. Mol. Cell Biol.*, vol. 1, no. March, pp. 1–12, 2017.

- [59] Z. Cui, Y. Shen, K. H. Chen, S. K. Mittal, J. Y. Yang, and G. J. Zhang, “KANK1 inhibits cell growth by inducing apoptosis through regulating CXXC5 in human malignant peripheral nerve sheath tumors,” *Sci. Rep.*, vol. 7, pp. 1–2, 2017.
- [60] S. Tan, H. Li, W. Zhang, Y. Shao, Y. Liu, H. Guan, J. Wu, Y. Kang, J. Zhao, Q. Yu, Y. Gu, K. Ding, M. Zhang, W. Qian, Y. Zhu, H. Cai, C. Chen, P. E. Lobie, X. Zhao, J. Sun, and T. Zhu, “NUDT21 negatively regulates PSMB2 and CXXC5 by alternative polyadenylation and contributes to hepatocellular carcinoma suppression,” *Oncogene*, pp. 1–14, 2018.
- [61] 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. 2208–2215, 2011.
- [62] L. Fang, Y. Wang, Y. Gao, and X. Chen, “Overexpression of CXXC5 is a strong poor prognostic factor in ER+ breast cancer,” *Oncol. Lett.*, vol. 0, pp. 395–401, 2018.
- [63] I. Benedetti, A. M. De Marzo, J. Geliebter, and N. Reyes, “CXXC5 expression in prostate cancer: implications for cancer progression,” *Int. J. Exp. Pathol.*, vol. 98, no. 4, pp. 234–243, 2017.
- [64] C. E. Pardo, I. M. Carr, C. J. Hoffman, R. P. Darst, A. F. Markham, D. T. Bonthron, and M. P. Klade, “MethylViewer: Computational analysis and editing for bisulfite sequencing and methyltransferase accessibility protocol for individual templates (MAPit) projects,” *Nucleic Acids Res.*, vol. 39, no. 1, pp. 1–18, 2011.
- [65] I. Yevshin, R. Sharipov, T. Valeev, A. Kel, and F. Kolpakov, “GTRD: A database of transcription factor binding sites identified by ChIP-seq experiments,” *Nucleic Acids Res.*, vol. 45, no. D1, pp. D61–D67, 2017.

- [66] S. R. Hartono, I. F. Korf, and F. Chédin, “GC skew is a conserved property of unmethylated CpG island promoters across vertebrates,” *Nucleic Acids Res.*, vol. 43, no. 20, pp. 9729–9741, 2015.
- [67] J. M. Landolin, D. S. Johnson, N. D. Trinklein, J. M. Landolin, D. S. Johnson, N. D. Trinklein, S. F. Aldred, C. Medina, H. Shulha, Z. Weng, and R. M. Myers, “Sequence features that drive human promoter function and tissue specificity Sequence features that drive human promoter function and tissue specificity,” pp. 890–898, 2010.
- [68] A. Deaton and A. Bird, “CpG islands and the regulation of transcription,” *Genes Dev.*, vol. 25, no. 10, pp. 1010–1022, 2011.
- [69] M. Monti, “Epigenetics protocols,” *Eur. J. Histochem.*, vol. 56, no. 2, p. 8, 2012.



APPENDIX A

PRIMER SEQUENCES

Table 6. *GAPDH* Primers

Forward	GGGAGCCAAAAGGGTCATCA
Reverse	TTTCTAGACGGCAGGTCAGGT

Table 7. qPCR Primers

GENE NAME		PRIMER SEQUENCE
RPLP0	Forward	GGAGAAACTGCTGCCTCATA
	Reverse	GGAAAAAGGAGGTCTTCTCG
CXXC5	Forward	CGGTGGACAAAAGCAACCCTAC
	Reverse	CGCTTCAGCATCTCTGTGGACT

Table 8. Bisulfite Primers for *CXXC5*

CXXC5 REGION		PRIMER SEQUENCE
5' Promoter Region 1	Forward	ATTTTAAAGAATTGGAGAAAATTTTATT
	Reverse	AATTCCTAATCAAAAATCCTAAAAAATAAA
5' Promoter Region 2	Forward	TTATTTTTTAGGATTTTGTATTAGGAATTT
	Reverse	CTTCAAAAAATAAACTCTACC
ERE Enhancer Region 1	Forward	TGATAGGATTTTATTTTGTTATTTAGG
	Reverse	CACCAAATCCATCCTAATAAAAAATT
ERE Enhancer Region 2	Forward	GTGTGTTTATAGATTGGTAGTTTTTTTA
	Reverse	CTCGAAAAATAATATATCATCT

APPENDIX B

PCR CONTROLS

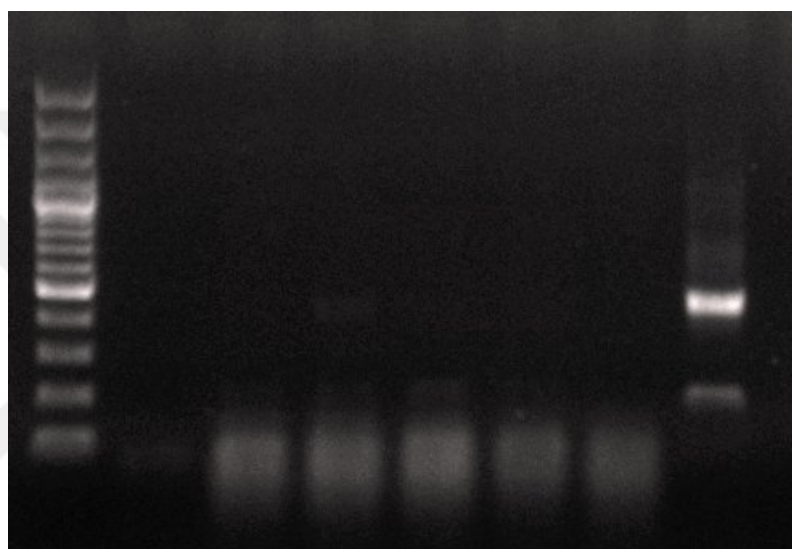


Figure 17. Genomic DNA Contamination Control After RNA Isolation

After total RNA was isolated and on-column DNase treatment carried out, PCR was carried out from 500 ng RNA template by using GAPDH primers. RNA samples does not have DNA contamination. Only last sample which is gDNA template as positive control of PCR has amplification of GAPDH

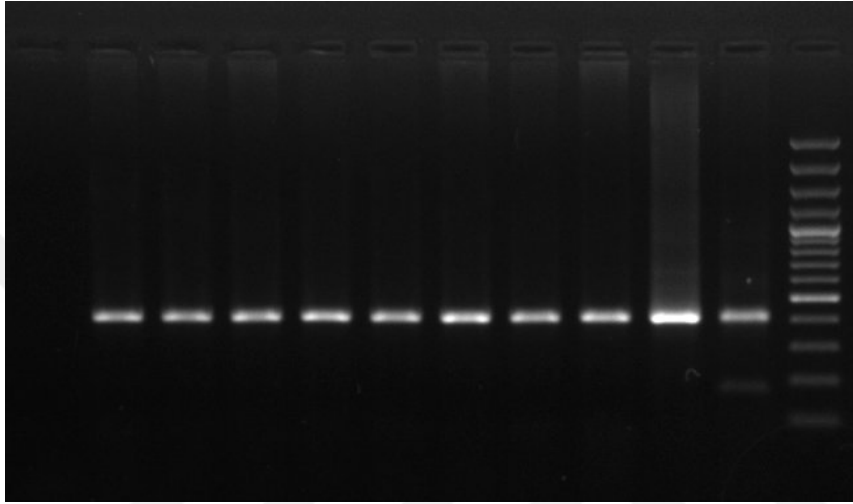


Figure 18. GAPDH PCR For cDNA Synthesis Control.

cDNA conversion was controlled with conventional PCR. First well is no template control, last is positive control with gDNA template. Others are cDNA templates synthesized from RNA of E2 treated and untreated (EtOH) MCF7 cells.

APPENDIX C

Table 9. MIQE CHECK LIST

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	YES
Yield	D	NO
RNA integrity method/instrument	E	YES
RIN/RQI or Cq of 3' and 5' transcripts	E	YES
Electrophoresis traces	D	YES
Inhibition testing (Cq dilutions, spike or other)	E	YES
REVERSE TRANSCRIPTION		
Complete reaction conditions	E	YES
Amount of RNA and reaction volume	E	YES
Priming oligonucleotide (if using GSP) 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*	N/A
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
<i>In silico</i> 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
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 optimisation (from gradients)	D	NO
Specificity (gel, sequence, melt, or digest)	E	YES
For SYBR Green I, C _q 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
r ² of standard curve	E	YES
Linear dynamic range	E	YES
C _q 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
C _q method determination	E	YES
Outlier identification and disposition	E	N/A
Results of NTCs	E	YES
Justification of number and choice of reference genes	E	YES
Description of normalisation method	E	YES
Number and concordance of biological replicates	D	YES

Number and stage (RT or qPCR) of technical replicates	E	YES
Repeatability (intra-assay variation)	E	YES
Reproducibility (inter-assay variation, %CV)	D	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/



APPENDIX D

RT-qPCR REACTIONS

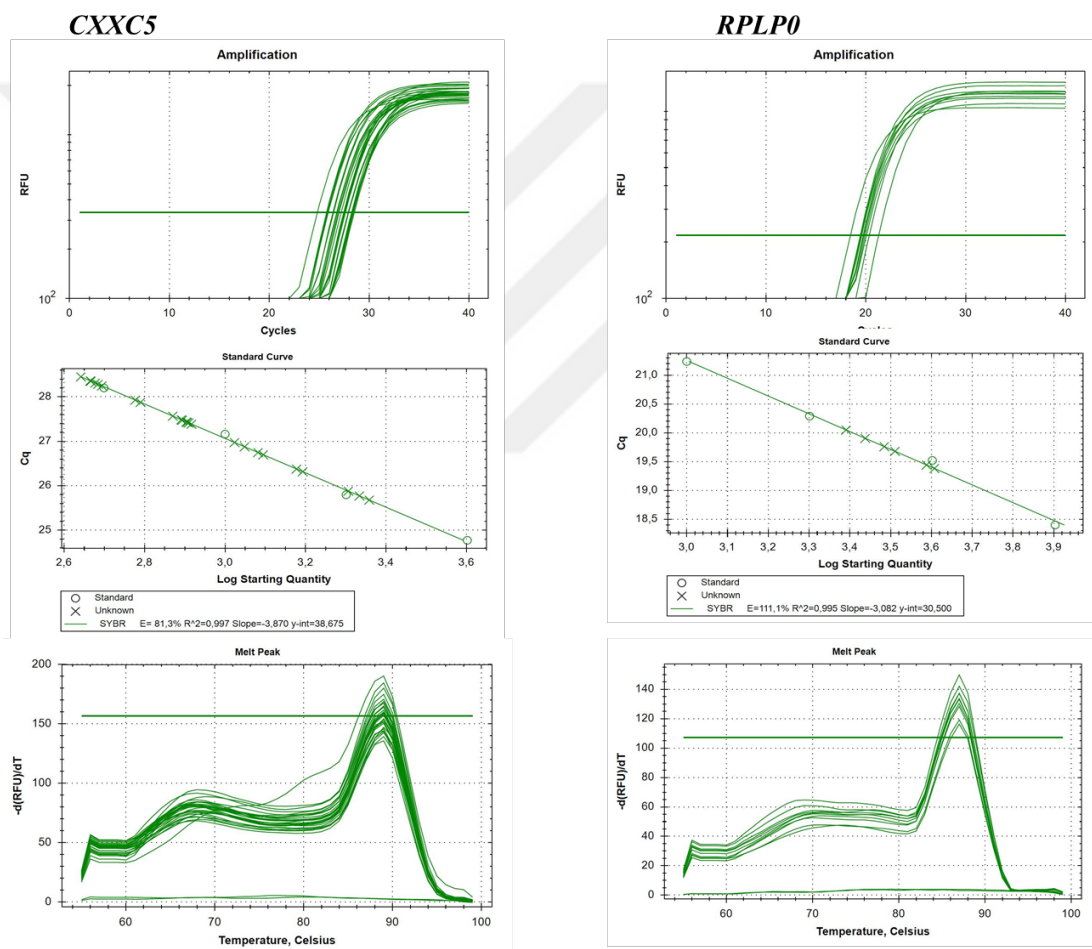


Figure 19. Amplification Curve, Melt Curve and Standard Curve of *CXXC5* and *RPLP0* qPCR Reaction



APPENDIX E

GpC METHYLASE ACTIVITY CONTROL

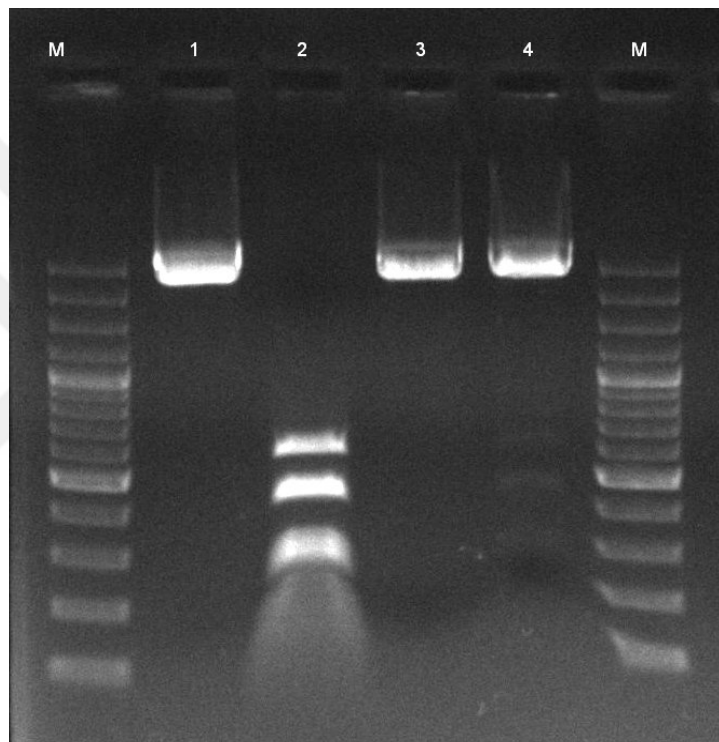


Figure 20. GpC Methylase Enzymatic Activity Control with HaeIII Restriction.

M: 1 kb marker, 1: Linearized pBSKS, 2: Linear pBSKS digested with HaeIII, 3: Linear pBSKS incubated with GpC Methylase, 4: Linear pBSKS incubated with GpC Methylase and digested with HaeIII. GpC methylation protects DNA from HaeIII digestion.



APPENDIX F

LOLLIPOP REPRESENTATIONS

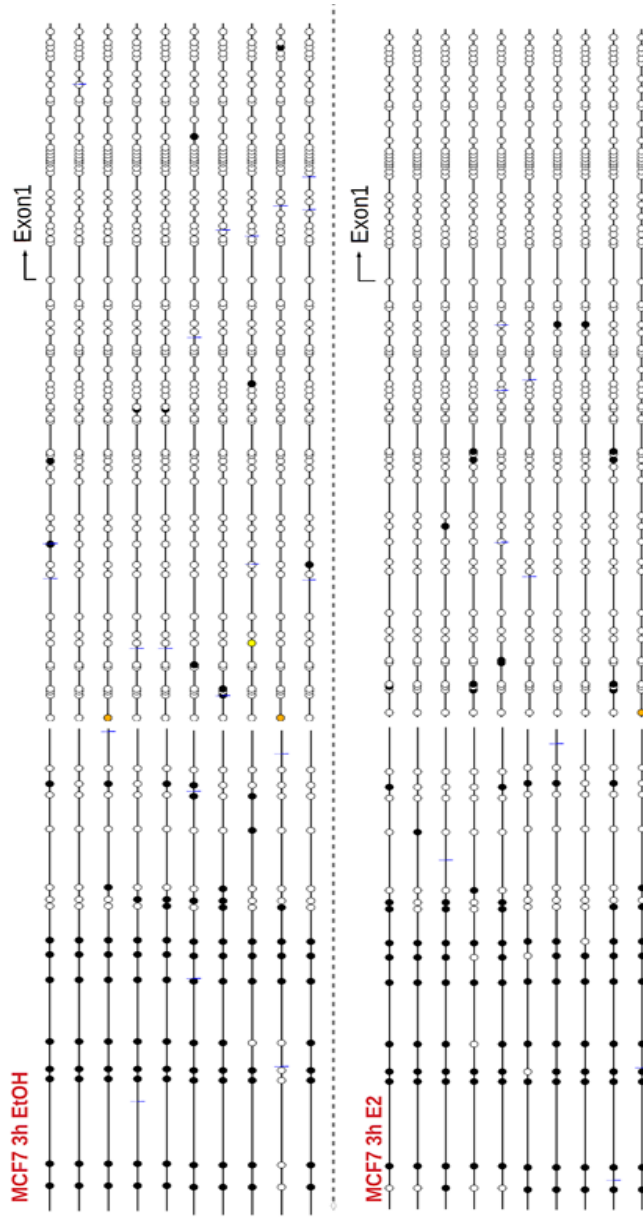


Figure 21. Lollipop Graph Representation of Endogenous Methylation of the 5' end of the putative CXXC5 Promoter from MCF7 cells. Lines indicate different clones. White circles are unmethylated CpG sites, blacks are methylated and yellow ones are CpG sites that can not be aligned. Blue marks correspond cytosine which are not included CpG dinucleotide and should have been converted to thymine.

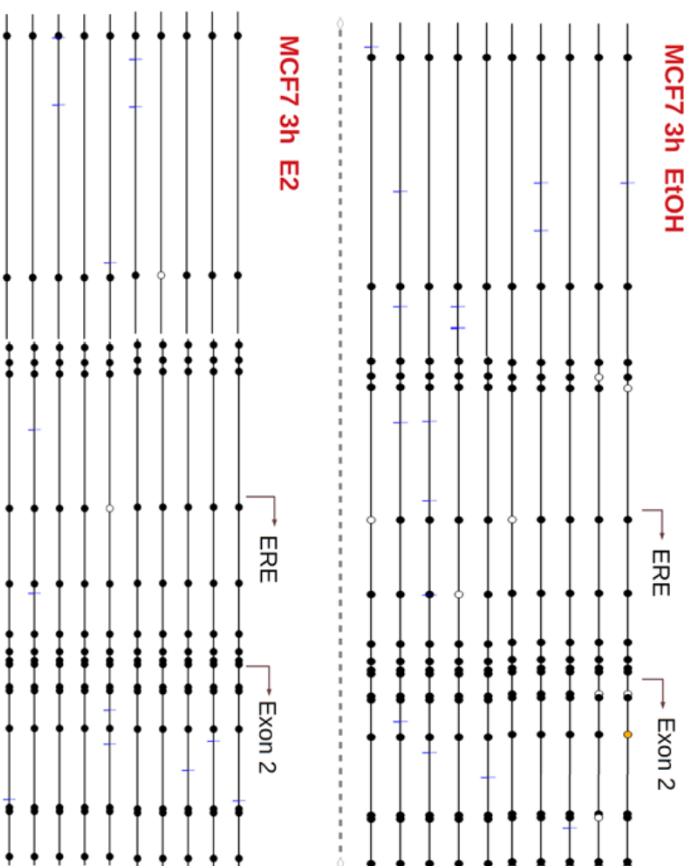


Figure 22. Lolipop Graph Representation of Endogenous Methylation of the Putative Enhancer of CXXC5 from MCF7 cells. Lines indicate different clones. White circles are unmethylated CpG sites, blacks are methylated and yellow ones are CpG sites that can not be aligned. Blue marks correspond cytosine which are not included CpG dinucleotide and should have been converted to thymine.

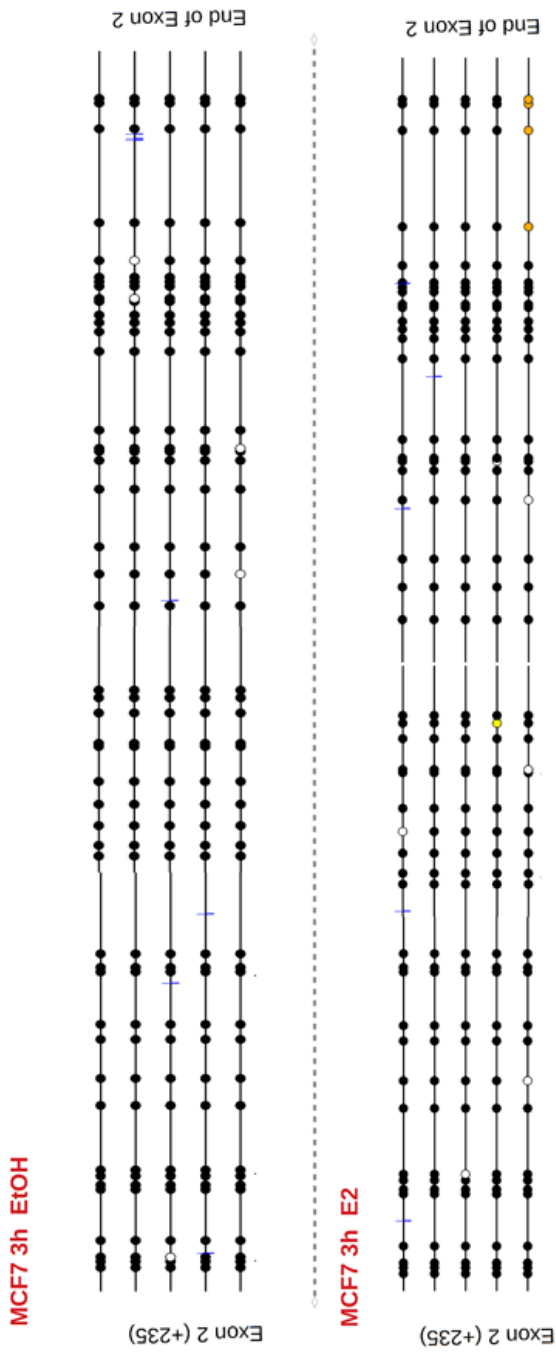


Figure 23. Lollipop Graph Representation of Endogenous Methylation of Exon 2 from MCF7 cells. Lines indicate different clones. White circles are unmethylated CpG sites, blacks are methylated and yellow ones can not be aligned. Blue marks correspond cytosine which are not included CpG dinucleotide and should have been converted to thymine. To confirm that the coding region is maintained as methylated, bisulfite PCR amplification to the end of exon 2 was conducted and 5 clones were sequenced for E2 treated and untreated (EtOH) samples.

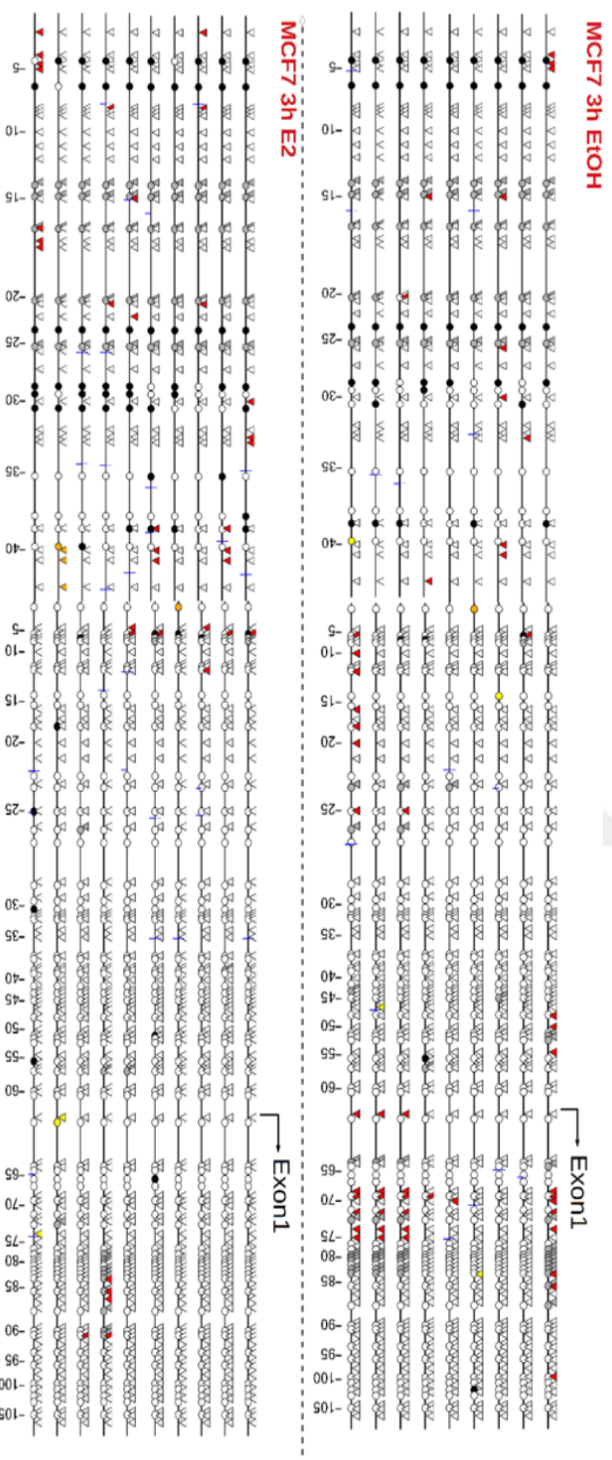


Figure 24. Lollipop Graph Representation of NOME sequences of CXXC5' Putative Promoter from MCF7 cells. Lines indicate different clones. White circles are unmethylated CpG sites and black ones are methylated. White triangles are unmethylated GpC sites. Red triangles are GpC sites methylated exogenously. Grey triangle next to a circle indicates methylated GpCpG sites. Blue marks correspond to cytosines which are not included CpG or GpC dinucleotides and should have been converted to thymine.

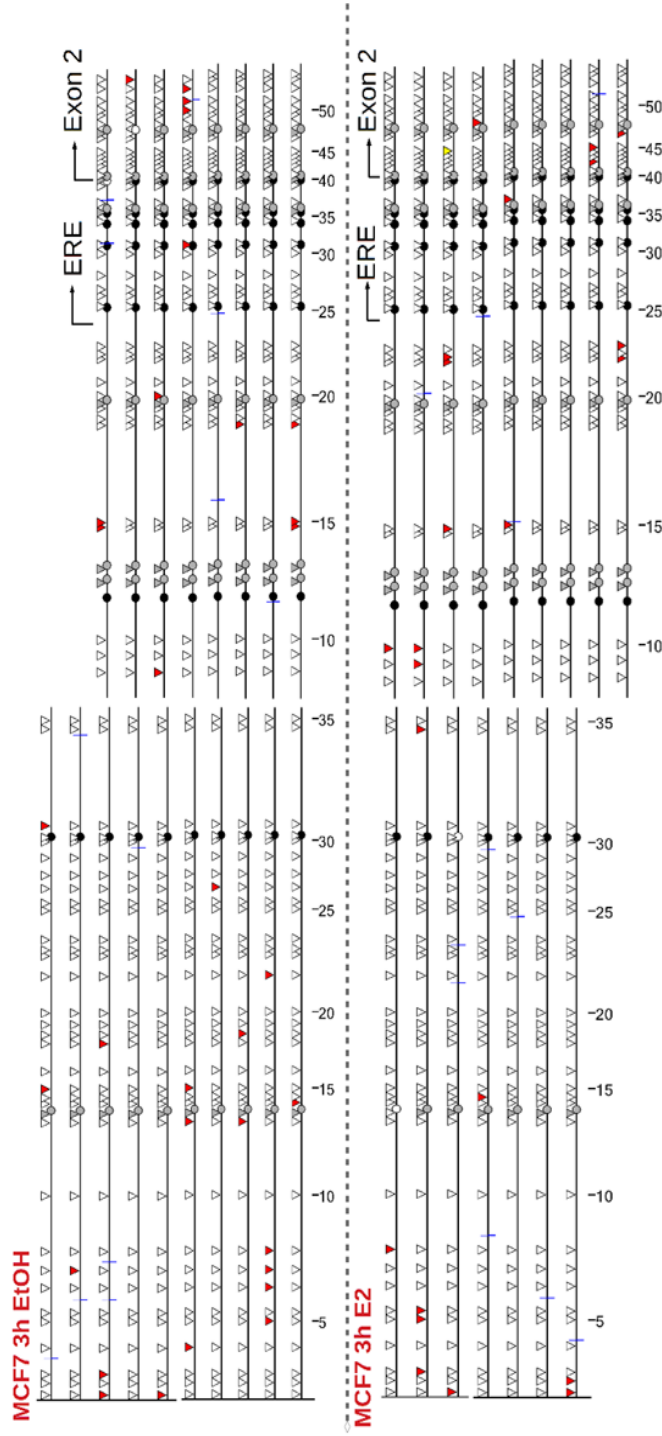


Figure 25. Lollipop Graph Representation of NOME sequences of CXXC5 Putative Enhancer from MCF7 cells. Lines indicate different clones. White circles are unmethylated CpG sites and black ones are methylated. White triangles are unmethylated GpC sites. Red triangles are GpC sites methylated exogenously. Grey triangle next to a circle indicates methylated GpCpG sites. Blue marks correspond to cytosines which are not included CpG or GpC dinucleotides and should have been converted to thymine.

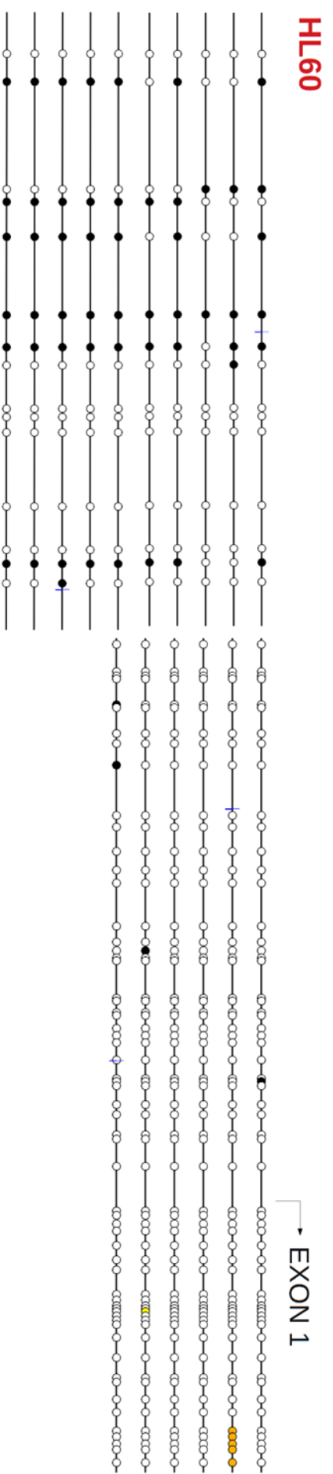


Figure 26. Lollipop Graph Representation of Endogenous Methylation of CXXC 5' Putative Promoter from HL60 cells. Lines indicate different clones. White circles are unmethylated CpG sites, blacks are methylated and yellow ones are CpG sites that can not be aligned. Blue marks correspond cytosine which are not included CpG dinucleotide and should have been converted to thymine.

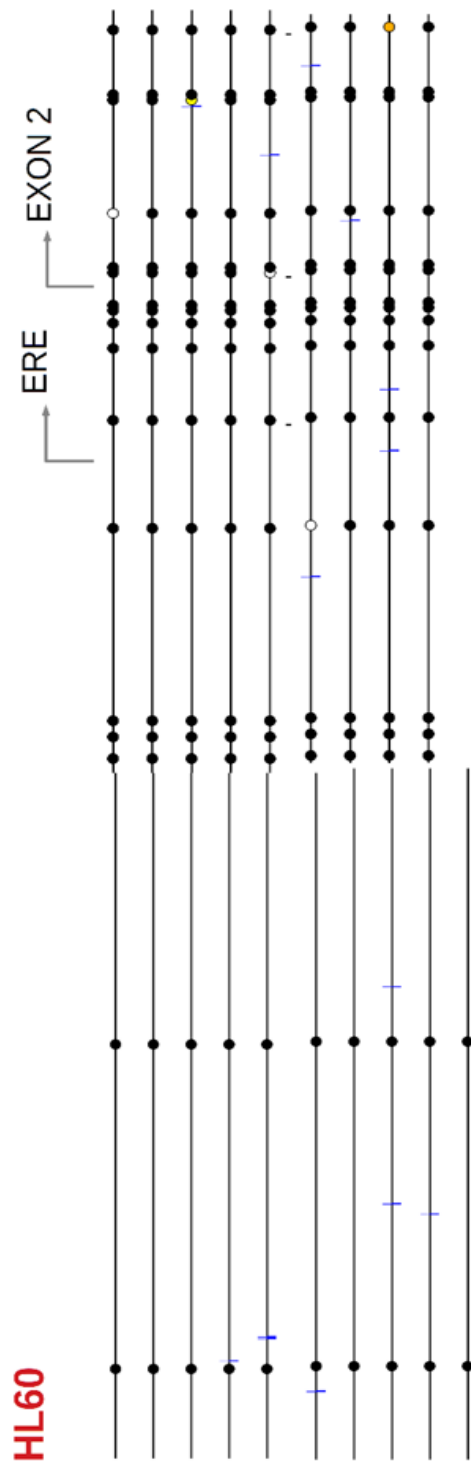


Figure 27. Lollipop Graph Representation of Endogenous Methylation of CXXC5 Putative Enhancer. Lines indicate different clones. White circles are unmethylated CpG sites, blacks are methylated and yellow ones are CpG sites that can not be aligned. Blue marks correspond cytosine which are not included CpG dinucleotide and should have been converted to thymine.

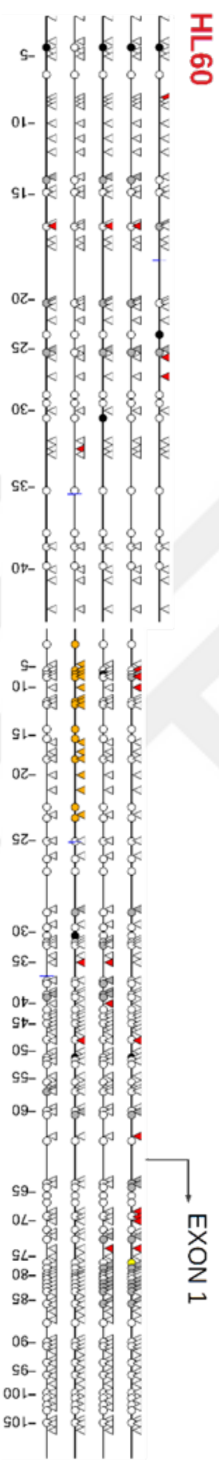


Figure 28. Lollipop Graph Representation of NOME sequences of CXXC5 Putative Promoter. Lines indicate different clones. White circles are unmethylated CpG sites and black ones are methylated. White triangles are unmethylated GpC sites. Red triangles are GpC sites methylated exogenously. Grey triangle next to a circle indicates methylated GpCpG sites. Blue marks correspond to cytosines which are not included CpG or GpC dinucleotides and should have been converted to thymine.



Figure 29. Lollipop Graph Representation of NOME sequences of CXXC5 Putative Enhancer. Lines indicate different clones. White circles are unmethylated CpG sites and black ones are methylated. White triangles are unmethylated GpC sites. Red triangles are GpC sites methylated exogenously. Grey triangle next to a circle indicates methylated GpCpG sites. Blue marks correspond to cytosines which are not included CpG or GpC dinucleotides and should have been converted to thymine.