Stony Brook University

Epigenetic Regulations and Promoter Characterization of *CERIG* (Cancer Endoplasmic

Reticulum Gene-KIAA1199)

A Dissertation Presented

by

Cem KUSCU

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Molecular and Cellular Biology

Stony Brook University

May 2012

Stony Brook University

The Graduate School

Cem KUSCU

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Jian Cao, MD
(Dissertation Advisor)
Assistant Professor, Department of Medicine

Howard B. Fleit, PhD
(Chairperson of Defense)
Associate Professor, Department of Pathology

J.Peter Gergen, PhD
Professor, Department of Biochemistry and Cell Biology

Stanley Zucker, MD
Professor, Department of Medicine

This dissertation is accepted by the Graduate School

Charles Taber
Interim Dean of the Graduate School

Abstract of the Dissertation

Epigenetic Regulations and Promoter Characterization of *CERIG* (Cancer Endoplasmic

Reticulum Gene-KIAA1199)

by

Cem KUSCU

Doctor of Philosophy

in

Molecular and Cellular Biology

Stony Brook University

2012

Although human genome project is complete, the biological function of most genes remains obscure. In this thesis project, the regulatory mechanisms of a novel endoplasmic reticulum (ER) resident protein identified by a PCR-subtraction hybridization technique, designated as Cancer ER Invasion Gene (*CERIG*), is investigated. CERIG is found to be highly expressed in human invasive breast cancer specimens examined by immunohistochemistry and real-time RT-PCR. This up-regulation of CERIG is associated with poor prognosis of patients with breast cancer assessed by a DNA microarray data-mining approach from three publicly available cohorts containing a total of 696 breast cancer patients.

In study of the regulation mechanism of CERIG, a 3.3 kb fragment of human genomic DNA containing the 5'-flanking sequence of the CERIG is found to possess both suppressive and activating elements. Employing a deletion mutagenesis approach, a 1.4 kb proximal region is defined as the basic CERIG promoter containing a TATA-box close to the transcription start site. A combination of 5'-primer extension approach with a bioinformatics analysis from the Cap-Analysis Gene Expression database reveals a single transcription start site in the human CERIG gene. Bioinformatics analysis suggests that the 1.4 kb CERIG promoter contains putative activating regulatory elements, including activator protein-1(AP-1), Twist-1, and NF-κB sites.

iii

Sequential deletion and site-direct mutagenesis analysis demonstrate that the AP-1 and distal NF-κB sites are required for CERIG gene expression. Further analyses using an electrophoretic mobility-shift assay and chromatin immunoprecipitation confirmed the requirement of these *cis*- and *trans*-acting elements in controlling *CERIG* gene expression. In further analysis of the regulatory mechanism of CERIG in cancer progression, we examine whether CERIG mRNA expression is regulated by the DNA methylation and/or histone H3 modification. Aberrant DNA methylation of CERIG 5'-untranslated region is identified as an important regulatory mechanism in the deregulation of CERIG. *In vitro* promoter methylation experiments and the demethylating reagent, 5'azacytidine, are used to determine the role of DNA methylation in CERIG expression. Pyrosequencing analysis revealed demethylation of CERIG in human breast cancer specimen that correlated with high expression of CERIG. The role of chromatin modifications in CERIG expression is also determined. Correlation between the level of tri-methylated H3K4 (activation marker) and the higher level of CERIG expression has been demonstrated in breast cancer cells. Substitution of H3K4me3 with H3K27me3 on histone tail has been found in cancer cells or primary cell lines in which CERIG expression declines. Additionally, the effects of hypoxia was analyzed on both epigenetic mechanisms for CERIG expression and determined that hypoxia increases the activation marker and decreases the repression marker on the CERIG promoter without changing the level of DNA methylation. Taken together, this study uncovers the regulatory mechanism of CERIG in breast cancer progression and suggests that CERIG may be used as a prognostic marker and potential therapeutic target in prevention of cancer metastasis.

To my family, Ceyda and Canan KUSCU

Nadire, Bahri and Cigdem KUSCU

Table of Contents

List of Figures

List of Tables

List of Abbreviations

5'aza	5'Azacytidine
BSP	Bisulfite Sequencing Primers
CAGE	Cap Analysis Gene Expression
CERIG	Cancer Endoplasmic Reticulum Invasive Gene
ChIP	Chromatin Immunoprecipitation
CNS	Central Nervous System
ConA	Concanavalin A
DCIS	Ductal Carcinoma In Situ
DNMT	DNA Methyltransferase
ECM	Extracellular Matrix
EMSA	Electrophoretic Mobility Shift Assay
EMT	Epithelial-To-Mesenchymal Transition
ER	Endoplasmic Reticulum
FFPE	Formalin Fixed Paraformaldehyde
FRT	Flip Recombinase Target
GSE	Genomic Spatial Event
H&E	Hematoxylin-Eosin
IHC	Immunohistochemistry
ILC	Invasive Lobular Carcinoma
LCM	Laser Capture Microdissection
MBD	Methyl-CpG-Binding Domain
MMP	Matrixmetalloproteinase
MSP	Methylation Specific PCR
PCR	Polymerase Chain Reaction
PTM	Posttranslational Modifications
TSS	Transcription Start Site

Acknowledgments

I would like to thank Dr.Jian Cao for his mentorships during my research project. His academic and moral supports help me a lot to finish my thesis at Stony Brook University.

I would like to thank my colleagues in Dr.Cao Lab; Dr. Hoang Lan Nguyen, Dr. Antoine Dufour, Dr. Jian Li, Niki Calabrase, Kevin Zarrabi, Ashleigh Pulkoski-Gross, and Deborah Kim for their discussions and friendships during my PhD life.

I would like to thank Dr.Stanley Zucker for his help on writing a paper.

I would like to thank all of my friends at Turkish Cultural Center Long Island for their friendship during my last seven years in the United States.

I would like to thank my former room-mates; Ibrahim Unal,Kerim Gulyuz, Selcuk Eren and Mustafa Kalafat for their kind friendships.

Chapter 1: Background and Significance

1.1 Role of metastasis in cancer

Cancer is still a major health problem for people all over the world, and it is the cause of 25% of deaths in the United States according to the 2010 Cancer Statistics (Jemal, Siegel, Xu, & Ward, 2010). We can describe cancer as a disease caused by the aberrant cell growth of the body's own cells, called tumors. However, not all tumors are cancerous, such as benign tumors, which do not grow out of control, do not invade adjacent tissue. In the case of cancer, tumor cells invade the neighboring tissue or spread to secondary part of the body by using the lymphatic system or bloodstream. Today, we know that most cancer related deaths happen due to migration and invasion of cancer cells into nearby tissue and other distant parts of the body, a process referred to as metastasis. Hanahan and Wienberg classified metastasis as one of the six hallmarks of cancer in the most cited cancer article of 2000 (Hanahan & Weinberg, 2000). And, they made us remember one of the facts about cancer; metastasis accounts for 90% of deaths of cancer patients (Sporn, 1996). Metastasis, which is the major obstacle for current treatment strategies, consists of 8 different steps as depicted in **Figure 1.1.**
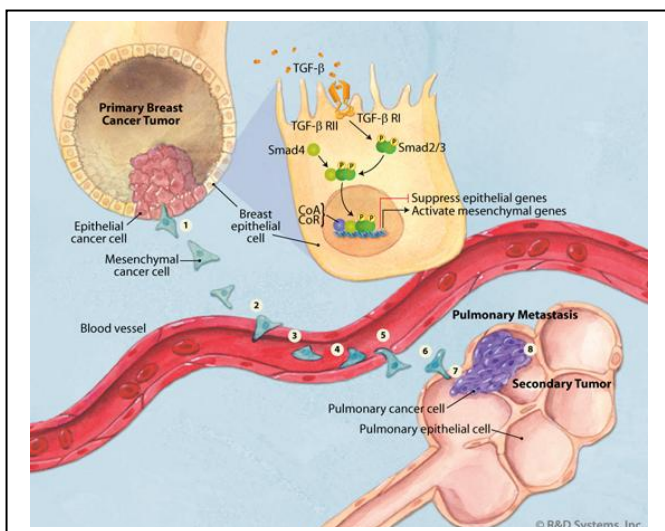
Figure 1.1: Steps of Metastasis. Epithelial cancer cells in the milk duct of breast tissue gain more migratory mesenchymal phenotype via epithelial-to-

mesenchymal transition (EMT). Inhibition of epithelial gene expression and activation of mesenchymal gene expression via TGF-beta signaling induces the EMT. Highly motile mesenchymal-like breast cancer cells invade pulmonary epithelia and proliferate as secondary tumors. This process requires EMT (1), mesenchymal cell intravasation (2), migration through the vasculature (3), adherence (4), extravasation (5), invasion of a secondary tissue (6), mesenchymal-to-epithelial transition (7), and distal proliferation (8)(adapted from R&D system, image #cb10i3_tgf-beta).

During metastasis, epithelial cells change their morphology into mesenchymal phenotype by decrease of molecules of epithelial and increase of molecules of mesenchymal. For example, some epithelial marker genes such as E-cadherin and Laminin-1 are suppressed during Epithelial-to-Mesenchymal Transition (EMT),but mesenchymal marker genes like N-cadherin and Vimentin are turned on (Kalluri & Weinberg, 2009). Another class of gene taking role in this process called matrix metalloproteinases (MMPs) that have already been recognized as important molecules for cancer dissemination because of their role in the degradation of extracellular matrix (ECM). Two gelatinases, MMP-2 and MMP-9 are listed under mesenchymal marker gene because of their role during intravasation, migration and extravasation steps of metastasis (Fassina et al., 2012). One of the key players during the reorganization of the ECM was also defined as MT1-MMP (MMP14) because of its role in the activation of latent MMPs such as direct activation of MMP-2 or indirect activation of MMP-9(Cao, Chiarelli, Kozarekar, & Adler, 2005). MT1-MMP, overexpressed in different cancer types, is localized to the plasma membrane and activates pro-MMP-2. Several studies have reported the role of Concanavalin A (ConA), which is also known as T-cell stimulator (Dwyer and Johnson, 1981), in increased MT1-MMP plasma membrane trafficking (Dwyer & Johnson, 1981; Yu, Sato, Seiki, & Thompson, 1995; Zucker, Hymowitz, Conner, DiYanni, & Cao, 2002). Concanavalin A is a lectin (carbohydrate binding protein) extracted from *Canavalia ensiformis* and consists of homotetramer with the size

2

of 106kDa. ConA changes the energy metabolism in the mouse T-cells and produces four distinct T-cell populations. It acts as mitogens and induces the expression of proto-oncogenes such as c-myc, c-fos and c-myb in T-cells (Reed et al., 1987). It has been reported that ConA induces MT1-MMP trafficking from cytoplasmic storage pools to plasma membrane, membrane anchored MT1-MMP then executes its biological functions. However, the mechanism underlying ConA-induced MT1-MMP has not been characterized.

There are several screening approaches to address global gene expression profiling in cancer cells such as DNA microarray, functional screening of a mammalian cell expression library, subtractive hybridization, *etc*. In our study, we performed a PCR-based suppression subtractive hybridization method which has been demonstrated to be effective in isolating, normalizing, and enriching differentially expressed genes over 1,000-fold in a single round of hybridization (Diatchenko et al., 1996). This approach resulted in identification of marked upregulation of an unknown gene, named KIAA1199 reported in the HUGE(Human Unidentified Gene-Encoded) database where cDNA clones larger than 4kb have been identified and characterized (Kikuno et al., 2004) **(Fig 1.2)**. Contrary to our experimental expectation of identifying genes relevant to protease function in cancer cells, this gene product exhibits no apparent effect on proteolytic activity based on our observation that silencing of *KIAA1199* in HT1080 cell, a fibrosarcoma cell line expressing endogenous MMP-14 and MMP-2 activity, did not change the activation status of MMP-2 by MT1-MMP. Since KIAA1199 is highly expressed in human cancers and the mechanism of upregulated KIAA1199 in cancer remains unclear, my thesis dissertation is focused on characterization of regulatory mechanism of KIAA1199 expression.

Better understanding of the molecular and cellular mechanisms by which cells invade into the extracellular matrix and migrate through vasculature is crucial for the treatment of cancer (Cao et al., 2004).
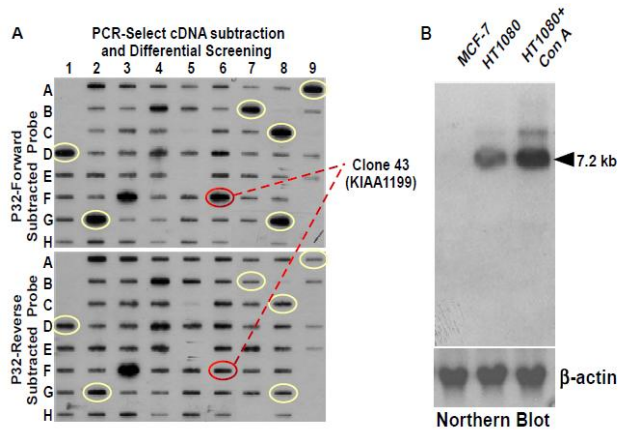
Figure 1.2) *PCR*-**based subtractive hybridization analyses A**) A *PCR*-based subtractive hybridization analysis was used to identify specific changes in gene expression in response to Concanavalin-A stimulation in HT1080 cells. *KIAA1199* (4.6 fold increase) was identified as one of the 38 genes significantly upregulated in HT1080 cells treated with ConA as compared to untreated cells. **B)** The original identified *KIAA1199* fragment was determined to be 724-bp in length. Initial sequence analysis of the cDNA fragment revealed a high degree of similarity (99%) to a functionally unknown cDNA clone DKFZp586O0118 from the Consortium of the German Genome Project (Genbank accession # AL049389). To determine the mRNA transcript size of *KIAA1199*, the *KIAA1199* fragment was used as a probe for Northern blotting analysis. A single *KIAA1199* transcript with a length of ~7.2 kb is identified in HT1080 cells, with five-fold increased expression noted after ConA treatment. No *KIAA1199* mRNA is detected in non-invasive MCF-7 human breast cancer cells.

1.2 General knowledge about *KIAA1199*

As HUGE database identified, *KIAA1199* is a relatively large cDNA with a size of 7080 bp. Based on bioinformatics analysis, the open reading frame of KIAA1199 contains 4086 bp between 261 and 4346 nucleotide of mRNA. The gene coding for human *KIAA1199* is located on chromosome 15 at the region of 15q25.1. According to the recent human genome sequence of UCSC (University of California Santa Cruz), called Feb.2009 (GRCh37/hg19 version), the exact position of *KIAA1199* in chromosome 15 is between 81,071,712 and 81,243,999, i.e. it covers 172.4 kb region in the human genome. It has 29 exons, 28 of them are coding exons because the start site of translation (ATG) is located in the second exon. Of note, one interesting thing about the structure of *KIAA1199* is the presence of a very large intron between the first and second exons. The first intron of KIAA1199 is 97 kb. **Figure 1.3** summarizes the exon and intron composition of KIAA1199. Another structural feature of KIAA1199 is the presence of a CpG island around the promoter region with high GC content and CG pairs.
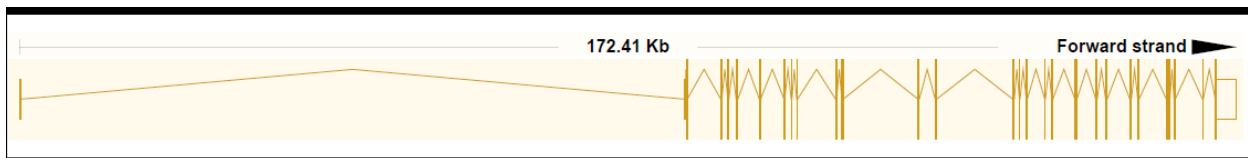
Figure 1.3: Genomic organization of *KIAA1199*. It is located on the (+) strand of chromosome 15. Exon and intron structure of *KIAA1199* are represented on a ratio-base. Exons are demonstrated by vertical bars and introns between the exons are shown by the lines. Notably, first intron with size of 97kb is larger than the half of the entire *KIAA1199* mRNA.

Highly conserved sequence of human *KIAA1199* with other organisms including mammals and vertebrates at both the DNA and protein level predicts a potential important biological function. **Table 1.1** shows the rate of identity for KIAA1199.

Pairwise Alignment Scores

Species	Gene Symbol	Identity (%) Protein	DNA
Homo sapiens	**KIAA1199**		
vs. Pan troglodytes	KIAA1199	99.3	99.5
vs. Canis lupus familiaris	LOC488766	89.4	87.1
vs. Bos taurus	LOC519047	91.4	88.7
vs. Mus musculus	9930013L23Rik	91.3	86.8
vs. Rattus norvegicus	RGD1305254	91.0	86.0
vs. Gallus gallus	KIAA1199	76.6	74.5
vs. Danio rerio	LOC100003660	67.2	67.1

Table 1.1 Pair wise Alignment Scores. Human KIAA1199 protein or DNA sequence is used as a reference to calculate percent identity of protein/DNA between human and other higher eukaryotic organisms.

The *KIAA1199* gene encodes a protein composed of 1361 a.a with an estimated molecular weight of 163 kDa based on bioinformatics tools (*ExPASy*). KIAA1199 was shown to have a G8 domain containing eight conserved glycine residues and five β -strand pairs as well as two GG domains composed of seven β -strand pairs and two α -helices (Guo, Cheng, Zhao, & Yu, 2006; He et al., 2006). However, the function of these domains remains to be understood. G8 is a novel protein domain existing between a.a. residues 44 to 166 of the KIAA1199 protein and has 5 repeated β-strand pairs and 1 α-helix in its secondary structure according to the Jpred server prediction (http://www.compbio.dundee.ac.uk/~wwwjpred /submit.html) (He et al., 2006). Moreover, four small motifs, called PbH1 (parallel beta-helix repeats) were located in the followings regions of the protein; 572-594, 595-617, 719-741, 798-819 (**Fig 1.4**). However, the presence of the G8 domain and the PbH1 motifs in the KIAA1199 protein did not give us any

6

hint as to potential function of KIAA1199.Further analysis did not reveal any other homology with known functions.

position in KIAA1199	length of domain/motif	name of domain or motif	
44 – 166	123	G8	
572 – 594	23	PbH1 1	
595 – 617	23	PbH1 2	
719 – 741	23	PbH1 3	
798 – 819	22	PbH1 4	

Figure 1.4) Domains of KIAA1199 protein. G8 domain is identified between 44 and 166 residues of the N'terminus of KIAA1199 protein. Four parallel beta-helix repeats containing 23 aminoacids except the last one which has 22 aminoacids are located in the middle of protein according to the secondary structure.

To explore the tissue distribution of *KIAA1199* under physiological conditions, the expression of *KIAA1199* was analyzed in adult tissues. The Human RNA master blot was hybridized using the *KIAA1199* fragment as a probe. As shown in **Figure 1.5A**, abundant expression of *KIAA1199* is primarily detected in the spinal cord and lymph nodes. This result is in agreement with DNA microarray data examined by mining the DNA microarray database of NCBI/GEO profiles (Yanai et al., 2005) (**Fig 1.5B**). In another microarray study, primary expression of *KIAA1199* was also found in the spinal cord and brain in addition to the placenta (**Fig 1.5C**). Collectively, our result and data mining support the primary role of KIAA1199 in central nervous system (CNS) related tissue (Su et al., 2002) (**Fig 1.5C**).
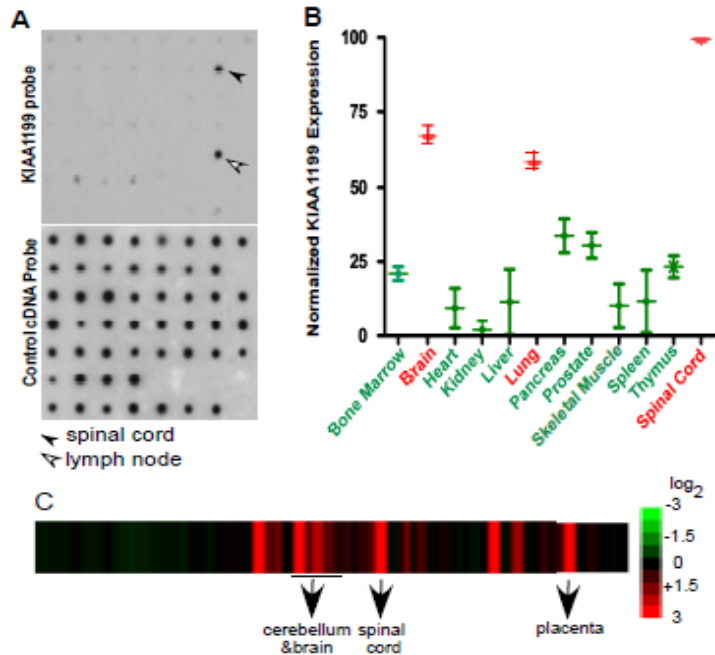
Figure 1.5) *KIAA1199* **expression in adult tissues. A)** Expression of *KIAA1199* in adult tissues. The Human Total RNA Master Panel II (Clontech) containing a series of total RNAs from normal human organs was hybridized with ^{32}P-labeled KIAA1199 probe. After exposing the x-ray film, the same membrane was reprobed using the control cDNA probe provided by the manufacturer. The expression of *KIAA1199* in normal tissues is limited to the spinal cord and lymph node. **B)** High expression of *KIAA1199* in spinal cord identified by data mining approach. By analysis of *KIAA1199* expression in 12 normal human tissues (each pooled from 10-25 individuals) using Yanai et al's data set, *KIAA1199* is most highly expressed in the spinal cord (P-value, 5.7E-14) as compared to other normal human tissues; expression in brain and lung is also increased. **C)** Primary human tissues and organs are surveyed by using the U133plus2 Affymetrix microarray for gene expression profiles. Particularly, expression of *KIAA1199* was chosen and represented for 47 tissues (Su et al., 2002).

1.3 KIAA1199 and Cancer

In the literature , the first study on KIAA1199 was published in 2003 and identified it as a product of a human inner-ear specific gene in which three point mutations are shown to cause nonsyndromic hearing loss in some Japanese families (Abe, Usami, & Nakamura, 2003; Usami

et al., 2008). Three years later, the role of KIAA1199 in the cellular mortality of normal human cells was identified (Michishita, Garces, Barrett, & Horikawa, 2006). In the study by Michishita, KIAA1199 was also shown to be upregulated in some breast cancer cell lines including MDA-MB-231 and MDA-MB-435. *KIAA1199* has recently been demonstrated to be upregulated in human gastric cancer (Matsuzaki et al., 2009) and colorectal cancer (Birkenkamp-Demtroder et al., 2011; Sabates-Bellver et al., 2007) (**Fig 1.6**). Moreover, Galamb *et al* demonstrated that the upregulation of *KIAA1199* gene occurred in the colon adenocarcinoma cells with respect to the normal cell in microarray study. They recapitulated their results in real-time PCR and Immunohistochemistry staining. In their study, they treated the HT-29 colon cell line with anti-cancer agent, NS398, and then they observed the donwregulation of *KIAA1199* as well as cadherin 3 whose aberrant expression has been linked to the breast, prostate and cervical adenocarcinomas (Shimoyama et al., 1989). In addition, high expression levels of *KIAA1199* correlates with low survival rates in several cancer types such as gastric (Matsuzaki et al., 2009) and breast; however, its function in cancer progression is still unknown.
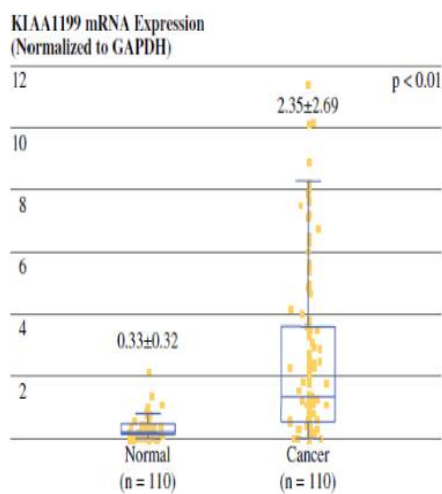
(Shinji Matsuzaki *et al.* 2009)　　　(K Birkenkamp-Demtroder *et al.* 2011)
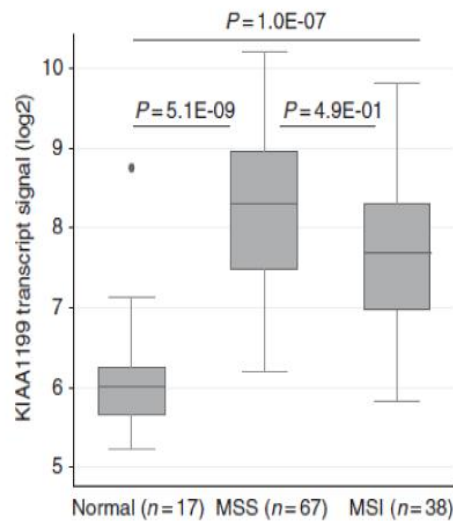
9

Figure 1.6) *KIAA1199* **expression in gastric and colon cancer. Left)** *KIAA1199* mRNA expression in normal and cancer tissues from gastric cancer patients is determined by real-time RT-PCR. Mean expression levels of *KIAA1199* in cancer tissue were significantly higher than levels in normal tissue (P < 0.01, Student's t-test) (Matsuzaki et al., 2009) **Right)** The role of *KIAA1199* in progression and clinical outcome of colorectal cancer is determined at mRNA level (normal mucosas, MSS: microsatellite stable and MSI: instable and adenocarcinomas) (Birkenkamp-Demtroder et al., 2011).

1.4 Subcellular localization of KIAA1199 in endoplasmic reticulum (ER)

Although subcellular localization of KIAA1199 has been demonstrated to be within inside the nucleus of colon cancer cells by Birkenkamp (Birkenkamp-Demtroder et al., 2011), Matsuzaki mainly found expression of the protein in the cytoplasm of human gastric cancer cells (Matsuzaki et al., 2009). To predict the localization of KIAA1199, two different bioinformatics programs, Expasy (http://au.expasy.org) and pSORTII (http://psort.hgc.jp/form2. html) were employed. Conclusive subcellular localization for the KIAA1199 product by this approach was not determined, but pSORTII suggested ER (endoplasmic reticulum) retention of KIAA1199 because of the presence of four lysine (K) residues at the C-terminus of the KIAA1199 protein. In Jackson's analysis, the KKXX motif at the C-terminus of a protein was accepted as an endoplasmic reticulum retention signal in addition to the major ER retention signal, KDEL, which was not detected in KIAA1199 (Jackson, Nilsson, & Peterson, 1990).

We demonstrated that localization of KIAA1199 was detected in the endoplasmic reticulum (ER) by using several experimental approaches. First, differential centrifugation was employed to evaluate overall distribution of KIAA1199 by subcellular fractionation. Myc-tagged KIAA1199 and MT1-MMP (used as positive control) were detected in the crude cell membrane pellet (100,000 x g) with minimal protein identified in the supernatant (cytosol). Secretory

MMP-28, which is used as negative control was detected in the cytoplasmic fraction. These results suggested that KIAA1199 is associated with cellular membranes/organelles (**Fig 1.7A**). To investigate the dynamics of KIAA1199 in the ER, a fluorescent recovery after photo bleaching (FRAP) experiment was performed in living COS-1 cells expressing KIAA1199-GFP. The recovery is rapid with an effective diffusion coefficient (D_{eff}) value of 0.0207 μm^2/s and diffusion at the edges of the photo bleached box recovered before the center (**Fig 1.7B**). This value indicated the high turnover rate of KIAA1199 in the endoplasmic reticulum. Second, to specifically determine the subcellular localization of KIAA119, a chimeric protein was generated by fusing a fluorescent marker encoding green fluorescent protein (GFP) to the C-terminus of KIAA1199. A polygonal network of interconnected tube-like structures with lengths ranging from 1 to 2.5 μm in the KIAA1199-GFP transfected cells indicates ER localization of KIAA1199 since the endoplasmic reticulum is composed of flattened sheets and tubules that branch to generate a polygonal network (Puhka, Vihinen, Joensuu, & Jokitalo, 2007). Calreticulin, also known as calregulin, was used as an ER marker because of its presence in the peripheral ER sheets and interconnected ER tubules, and it was found to be co-localized with KIAA1199. MT1-MMP and GFP transfected cells were used as negative controls and they did not show a similar phenotype when compared to KIAA1199 transfected cells (**Fig 1.7C**). Third, to determine the distribution of endogenous KIAA1199 in cells, a polyclonal antibody against recombinant KIAA1199 protein from amino acid Gly[1108] to Thr[1340] was raised. Employing this polyclonal antibody, immunofluorescence was performed to examine co-localization of endogenous KIAA1199 with an ER marker, Protein Disulphide Isomerase (PDI), in MDA-MB-435 human cancer cells. Endogenous KIAA1199 and PDI co-localized in the same subcellular

compartments (**Fig 1.7D**). Since PDI is recognized as an ER resident protein (Schweizer, Matter,

Ketcham, & Hauri, 1991), distribution of KIAA1199 occurs mainly in the ER compartment.
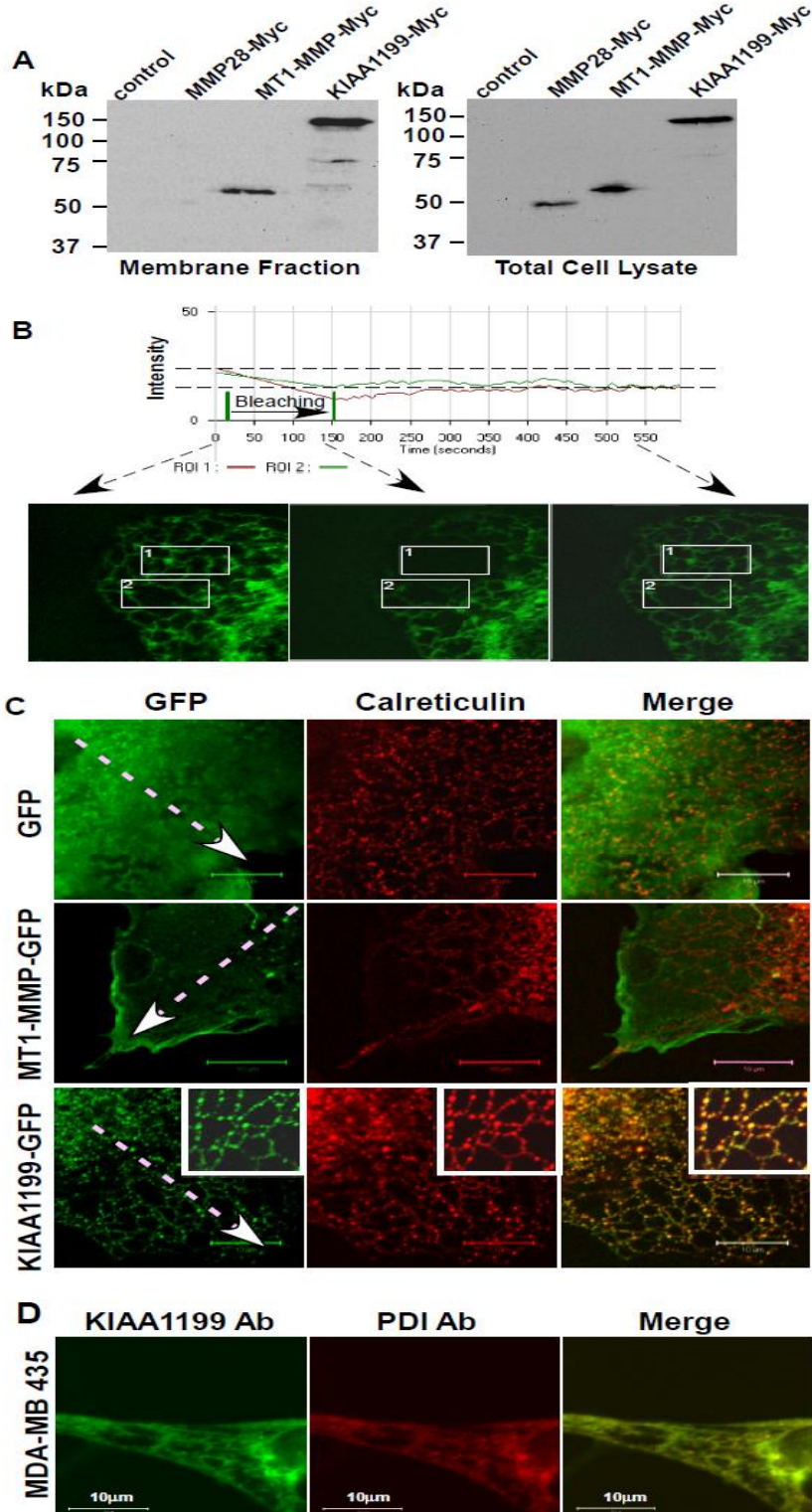
Figure 1.7) Subcellular localization of KIAA1199. A) COS-1 cells transfected with plasmids expressing myc-tagged full length KIAA1199, matrix metalloproteinase-28 (secreted MMP control), and MT1-MMP (plasma membrane protein control) were nitrogen cavitated followed by sequential centrifugation as described previously (Zucker et al., 1998).**B)** Photobleaching experiment demonstrate the rate of KIAA1199 movement in the endoplasmic reticulum. **C)** *COS-1 monkey kidney epithelial cells* transfected with GFP, MT1-MMP-GFP, and KIAA1199-GFP chimeric cDNAs were infected with a baculovirus containing an ER marker, Calreticulin. After post-infection for 48 hours, the cells were fixed and examined using confocal microscopy. Dashed arrow lines represent orientation from nuclear to plasma membrane. Part of a cell is displayed to closely look ER staining. Protein localization was superimposed with ER marker. Inserts represent polygonal *meshwork*. Bar: 10 µm. **D)** Immunostaining of endogenous KIAA1199: MDA-MB-435 cells were immune stained with anti-KIAA1199 antibody (Left Panel) and ER marker PDI antibody (Middle Panel) followed by incubating with the secondary antibodies. Co-localization of endogenous KIAA1199 with the ER marker was observed under a confocal fluorescent microscopy (Right Panel). Part of a cell is displayed.

Besides the above mentioned experimental approaches showing the localization of KIAA1199 in ER, interaction of the KIAA1199 with ER-resident protein called BiP was also shown in our laboratory by using SNAP-pull down experiment to further improve the localization of KIAA1199 in endoplasmic reticulum.

1.5 *CERIG* has a role in cancer cell migration and invasion

In our recent studies, a potential role for *KIAA1199* was identified in cancer cell invasion and migration and found it localizes to the endoplasmic reticulum. Hence, it was renamed as *CERIG* (*C*ancer *E*ndoplasmic *R*eticulum *I*nvasive *G*ene). For the rest of the thesis, the new name of *KIAA1199, CERIG* is used. The role of *CERIG* in cell invasion and migration has been determined by both loss-of-function and gain-of-function experiments. Expression of *CERIG* was inhibited by taking advantage of the shRNA approach (**Fig 1.8A**). Stable expression of

CERIG shRNA in MDA-MB-435 changed their morphology from typical mesenchymal-like cells to a polarized, epithelial cell shape (**Fig 1.8B**). No morphological change was noted in control MDA-MB-435 cells infected with luciferase shRNA. Further, staining of the cells with Texas red-labeled phalloidin demonstrated that F-actin exhibited a cortical ring-like staining pattern in the peripheral area of the *CERIG*-silenced MDA-MB-435 cells, whereas the control cells displayed increased numbers of stress fibers throughout the cells (**Fig 1.8B**). Since this type of morphological change and actin redistribution are often associated with cell migration, the *CERIG*-silenced MDA-MB-435 cells were evaluated for their migratory ability using a transwell chamber migration assay. As anticipated, silencing of endogenous *CERIG* in the cells resulted in decreased cell migration (**Fig 1.8 C, D**).
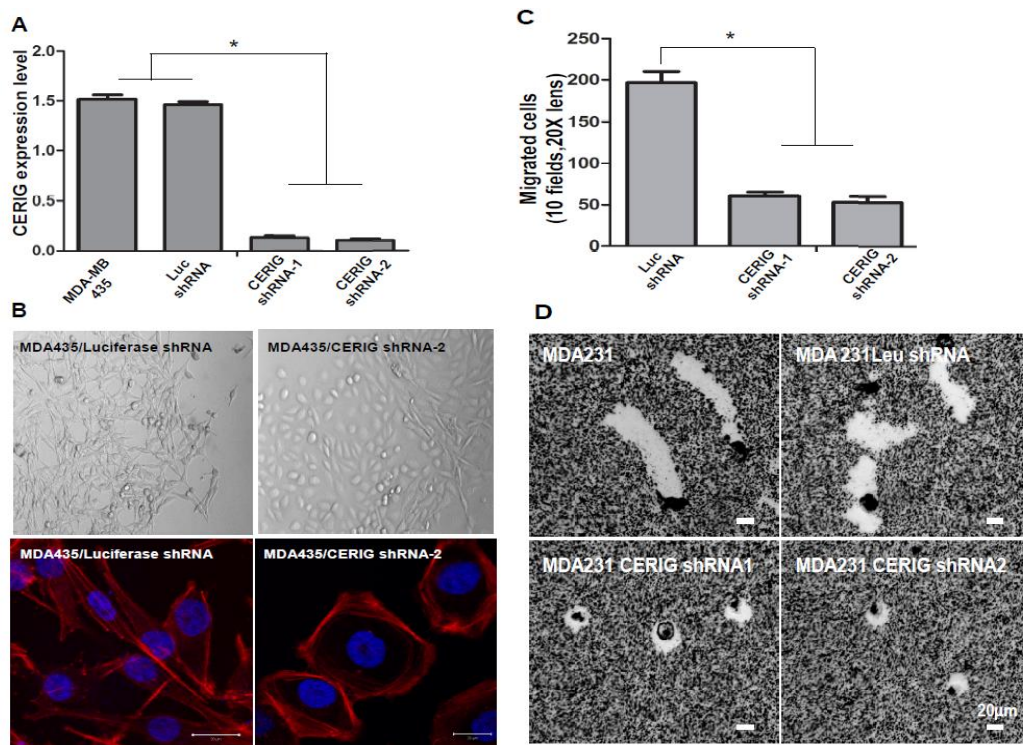
Figure 1.8) Loss of function assay for *CERIG*. A) *CERIG* expression is inhibited more than 90% by using two effective shRNA constructs. Luciferase shRNA is used as a negative control. **B)** Morphology changes during stable infection: MDA-MB-435 cancer cells were infected by retrovirus containing shRNAs for *CERIG* and luciferase (control). Down-regulation of *CERIG* gradually changed the mesenchymal appearance of the cells to polarized epithelial cell morphology (top panel). F-actin staining: The MDA-MB-435 cells expressing *CERIG* shRNA as well as control shRNA were stained with Texas Red-phalloidin. Distinct staining was observed in *CERIG* silenced cells and control cells (bottom panel). **C)** MDA-MB-435 cells expressing *CERIG* shRNAs or control shRNA were examined by a transwell chamber migration assay. Silencing *CERIG* in the cells decreased cell migration. **D)** Effects of *CERIG* shRNA are shown by using gold-colloidal migration experiment. Migrated cells remove the gold particles on a collagen coated slides, but downregulation of *CERIG* expression inhibits the migratory attribute of the MDA-MB-231 cells.

Since cell migration is a critical determinant of cancer invasion, next, the role of *CERIG* was tested in cancer invasion by employing our three-dimensional (3D) invasion assay (Cao et al., 2008). Silencing of *CERIG* in MDA-MB-435 cells resulted in decreased cell invasion into surrounding matrices (**Fig 1.9A**). To determine the impact of *CERIG* on tumor growth and metastasis, orthotopic mouse model that has previously been used to study the link between cell migration/invasion and metastasis was employed. MDA-MB-435 cells expressing *CERIG* shRNAs or GFP control shRNA were selected and maintained as pools of resistant cells to avoid artifacts due to the use of single clones and to evaluate a range of expression levels simultaneously. The cells were orthotopically injected into the mammary fat pads of immunodeficient mice and tumorigenicity was monitored. Tumor growth in MDA-MB-435 cells stably expressing *CERIG* shRNAs was diminished as compared to MDA-MB-435 and GFP shRNA control groups over 12 weeks of observation (**Fig 1.9B**). At necropsy, the lungs of mice were evaluated for metastases by histological analyses. Metastases were observed in 10/10

(100%) of mice from control groups as compared to 6/10 (60%) of *CERIG* shRNA mice (Fig

9B). In contrast, 6/10 (60%) of *CERIG* shRNA mice developed micrometastasis.
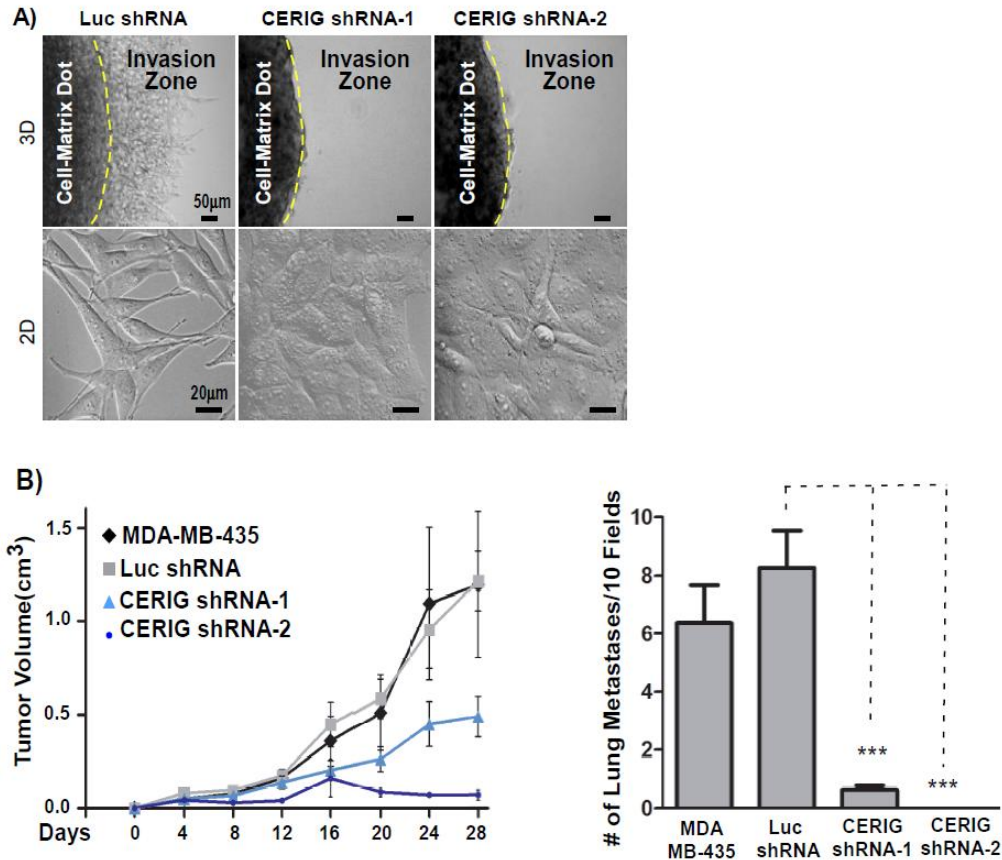
Figure 1.9) Role of *CERIG* in invasion. A) 3D invasion assay: MDA-MB-435 cells expressing luciferase (Luc) shRNA and two most effective *CERIG* shRNAs, shRNA-1 and -2 were examined by the 3D invasion assay after 8 days in culture. Silencing *CERIG* in the cells inhibited cell invasive ability, but does not alter cell proliferation (Upper Panel). Distinct morphology in two dimensional (2D) culture was noticed among these cell types (Bottom Panel). **B)** Tumor growth rate and lung metastasis: MDA-MB-435 cells stably infected with luciferase shRNA control and two *CERIG* shRNAs targeting different region of *CERIG* were orthotopically injected in mammary fat pad of 4-5-week-old immunodeficient female mice. Tumor size was measured every two days and lung metastasis was examined at the end of the experiments. Silencing *CERIG* in

MDA-MB-435 cells significantly retarded tumor growth (Left Panel) and lung metastasis (Right Panel).

To further confirm the role of *CERIG* in cell migration, a gain-of-function approach was performed by transiently overexpressing *CERIG* in non-invasive cancer cells. MCF-7 cells display weak migratory capability. These cells were transiently transfected with the *CERIG* cDNA and cell migratory ability was determined by the transwell migration assay. Vector cDNA and MT1-MMP cDNA were used as negative and positive controls, respectively (Cao et al., 2004). Overexpression of *CERIG* in MCF-7 cells markedly induced cell migration, comparable to MT1-MMP-transfected cells (**Fig 1.10**).
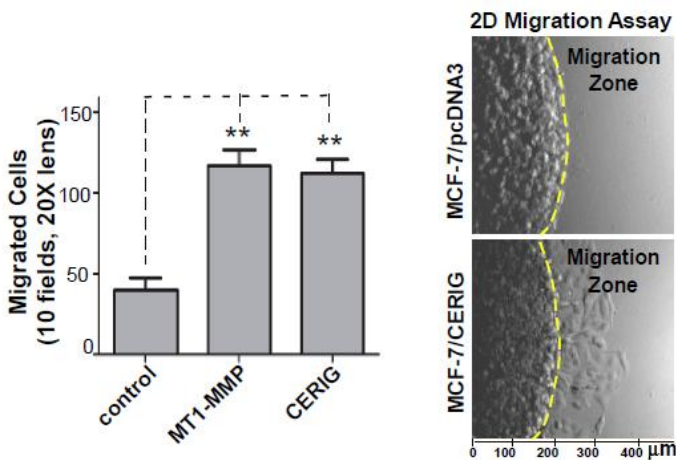
Figure 1.10) Gain of function assay for *CERIG*. Migration assays in *CERIG* transfected MCF-7 cells: Less invasive MCF-7 cells transfected with *CERIG* cDNA as well as controls were examined by a transwell chamber migration assay. MT1-MMP was used as a positive control (Left Panel). Effects of *CERIG* overexpression in MCF-7 cells via collagen dot migration assay (Right Panel). Ectopic expression of *CERIG* increased cell migration as well as MT1-MMP transfected cells.

The results obtained from both loss- and gain-of-function experiments indicate that *CERIG* is a key molecule involved in cancer cell migration and invasion.

Chapter 2: Expression Level of *CERIG* in Breast Cancer

2.1 Introduction

Breast cancer is the most common form of cancer among women and it is the second most deadly cancer after lung cancer (Jemal et al., 2010). It originates from breast tissue and is named according to the source of the tumor cells within the breast tissue. Ductal carcinoma originates from ducts of the breast whereas lobular carcinoma originates from lobules that supply the ducts with milk (Sariego, 2010). Invasive epithelial tumor cells in the invasive ductal carcinoma are no longer encircled by a myoepithelial or basal cell layer as in the normal duct. While basement membrane is degraded in invasive breast cancer, its existence around monotonous proliferated tumor cells becomes a hallmark of the ductal carcinoma in situ (DCIS), which is described as an early form of cancer. The major characteristic difference between the invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC) is the spatial organization of tumor cells within the stroma. While tumor cells in the invasive ductal carcinoma generally cluster together, tumor cells in the invasive lobular carcinoma are small and discohesive, and they show a single file pattern. Moreover, many breast cancer cell lines used in regular tissue culture study were derived from invasive ductal carcinoma, such as MCF-7 and MDA-MB-231. Several studies have also demonstrated aggressive phenotype of MDA-MB-231 compared to MCF-7 (Liu, Brattain, & Appert, 1997; Matteucci, Bendinelli, & Desiderio, 2009). Hence, these two cancer cell lines were chosen to study breast cancer for *CERIG* gene expression.

Since tissue samples taken from cancer patients are heterogeneous, and therefore are composed of a mixture of tumor cells and surrounding stromal cells, experimental techniques for the analysis of cancer cells has been difficult. However, improvements have been made over the years, including Laser Capture Microdissection (LCM), which has been used for the analysis of

histopathology samples since 1996 (Emmert-Buck et al., 1996). One can isolate specific cells of interest from histology slides using this technique. Upregulation of *CERIG* has been shown in multiple cancer types, including gastric, colon and breast. In these studies, the mRNA level of *CERIG* was approximately two-fold higher in the cancer tissues, but the major problem of these types of studies is the potential contamination problem of stroma cells into epithelial cells of cancer tissue. Therefore, Laser Capture Microdissection might give us a chance to isolate mRNA from cells that can be chosen individually. By using this technique, one can get more trustable data. So far, there is no specific study in which breast cancer specimens were used to test for *CERIG* expression; the only information known regarding *CERIG* in breast cancer is from data mining.

Another improvement in the molecular biology is the daily-based application of microarray data, particularly from clinical samples. Miniaturized DNA microarrays have been used for genome-wide studies for more than a decade (Lashkari et al., 1997). Taking advantage of this technique, Van de Vijver and his colleague Van't Veer analyzed vast number of primary breast tumors for the first time and they identified the gene expression signature for a strong prediction of a short interval time of distant metastasis (poor prognosis signature) in lymph node negative patients (van 't Veer et al., 2002; van de Vijver et al., 2002). They characterized 70 genes that are significantly upregulated in the patients with metastasis such as cyclin E2, MCM6, MMP-9, VEGF receptor. However, *KIAA1199/CERIG* was not included in this list. After their studies, many microarray studies were performed using large number of cancer patients. Today, most of the DNA microarray data is publicly available in certain databases, particularly in the NCBI with their specific GSE (Genomic Spatial Event) numbers. Using the data generated from these DNA microarrays, scientists have tried to make a correlation between a certain gene of

interest and the survival rate of patients as an indicator for the role of the gene in cancer progression. One of the well-known analyses performed using these datasets or cohorts is correlation studies, which aim at determining a correlation between the mRNA profile of patients and their survival time. In several studies, researchers focused on their gene of interest, and dichotomized the patients by using the average mean value, median value of gene expression or the quartile from the top or bottom of the group (Marcucci et al., 2005; Schmidt et al., 2011; Zhou et al., 2007). Then, they performed the survival curve analysis on these two groups of patients if enough clinical information is supported.

2.2 Results

2.2.1 mRNA expression in cancer cell lines and primary tissues

In this study, the mRNA expression level of *CERIG* in different cancer cell lines and also in primary breast cancer tissues were investigated. Real-time PCR experiments was performed for the quantification of mRNA using the HPRT-1 housekeeping gene as a reference for $\Delta\Delta$Ct calculation.

CERIG **expression is related with aggressiveness in cancer cells**

Hormone-dependent and –independent breast and prostate cancer cell lines were tested to evaluate *CERIG* expression. Using real time RT-PCR, low levels of *CERIG* in the non-invasive hormone-dependent MCF-7 breast cancer cell line and the LNCaP prostate cancer cell line were observed. In contrast, high levels of *CERIG* mRNA were observed in invasive and more aggressive MDA-MB-231 and MDA-MB-435 breast cancer cell lines and the DU145 prostate cancer cell line (**Fig 2.1**). MDA-MB-231 cells have 26 fold more *CERIG* transcripts compared to MCF-7.These findings are in agreement with a previous observation of Michishita (Michishita et

al., 2006). In addition, *CERIG* mRNA level was tested in the normal primary cell line of HUVEC (Human Umbilical Vein Endothelial Cells) and the amount of *CERIG* was found to be seven-times less than in MCF-7 cells, indicating minimal amount of *CERIG* expression in normal primary cells. Moreover, *CERIG* expression was measured in colon cancer cell lines, and was detected ten times more *CERIG* expression in HCT116 cells compared to SW480 cells. In the study of Flatmark comparing twelve colorectal cancer cell lines, HCT-116 and HT-29 cell lines produced lymph node metastasis in 70% of tumor-containing mice, whereas SW480 cells did not metastasize (Flatmark, Maelandsmo, Martinsen, Rasmussen, & Fodstad, 2004) (**Fig 2.1**). In agreement with Galamb study indicates an increase for *CERIG* expression in colon samples, significant increase in the *CERIG* expression occurs in HT-29 cells with respect to the SW480 cells (Galamb et al., 2010). Together, these data indicate a critical requirement of *CERIG* expression in the aggressiveness of cancer cell lines.
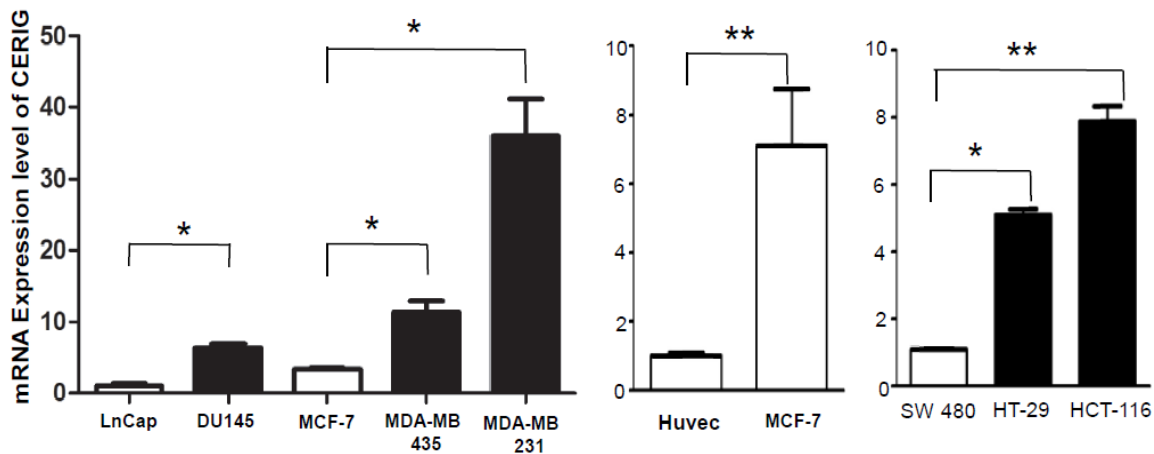
Figure 2.1) *CERIG* **expression in human prostate and breast cancer cell lines.** Human hormone dependent and independent prostate (LNCaP and DU145), and breast cancer (MCF-7, MDA-MB-231) cell lines as well as MDA-MB-435 cancer cell lines were examined by real time RT-PCR. The expression of *CERIG* was normalized by house-keeping genes (HPRT and GAPDH). Expression of *CERIG* correlated with invasive status of cancer cell lines (left panel). *CERIG* is

expressed minimally in primary cell line compared to cancer cell lines (middle panel). Higher *CERIG* expression was also observed in aggressive colon cancer cell lines (right panel). The relative levels of genes were determined using the ΔΔCT method. Each bar represents the mean ± S.E.

CERIG mRNA expression in primary breast cancer tissues

To determine the clinical relevance of *CERIG* expression in breast cancer progression, breast lesions ranging from benign to invasive tumors were examined by using archived specimens. Dr. Hu from Stony Brook Pathology Department supplied us with 10 benign, 8 invasive and 8 in-situ breast tissue samples. In bulk tissue sections, *CERIG* mRNA expression exhibited no significant difference between tissue specimens from benign controls and carcinomas. The reason might be the coexistence of epithelial and stromal tissues that dilute cell-specific gene expression in bulk tissue (Balu-Maestro et al., 2002). Therefore, after hematoxylin-eosin (H&E) staining of the samples, the tumor cells in invasive and ductal carcinoma in-situ samples were identified as well as normal epithelial cells in benign samples. Images of the tissue sections were transferred to the computer followed by the designation of tumor cells on the image via P.A.L.M MicroLaser program. Then, tumor cells were collected from the cancer specimens by laser capturing (Espina et al., 2006) (**Fig 2.2**). The same procedure was used to collect the normal epithelial cells from benign tissue.
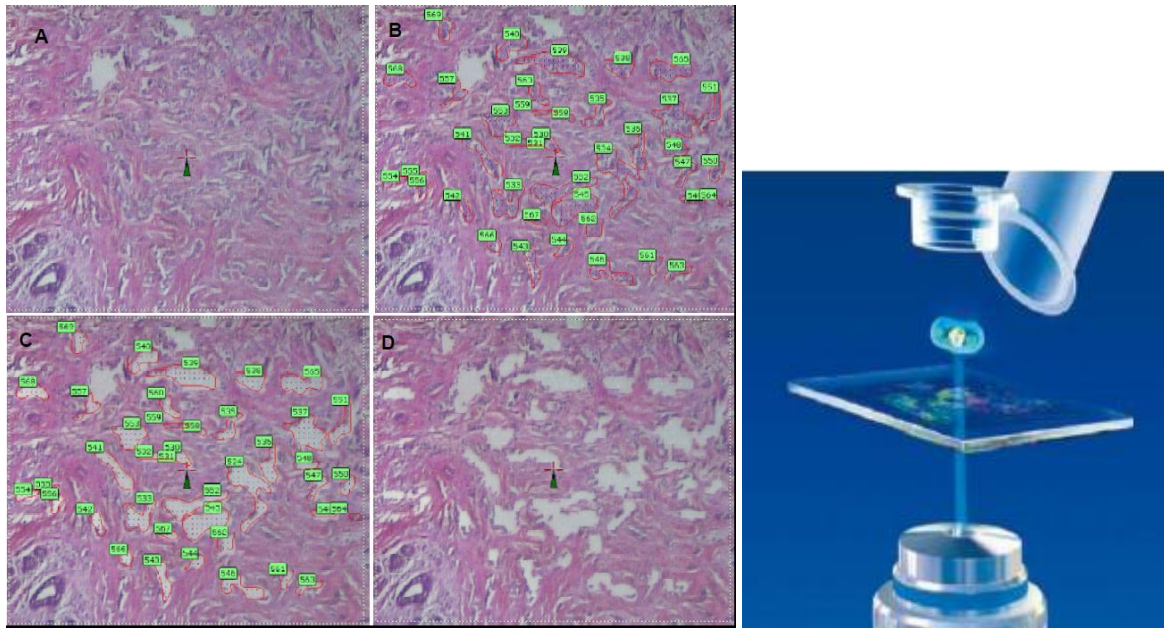
Figure 2.2) Laser Capture Microdissection (LCM) of cells. Left panel-A) Human breast cancer specimens and normal breast epithelium were stained by H&E (in-situ sample was shown in this figure). **B)** Invasive, in-situ and benign epithelial cells were designated by P.A.L.M ROBO program via circularization of cells with red lines. Designated ductal carcinoma in-situ cells are shown in B before the laser cut. **C)** Laser light hits the green arrow in the center and dissects the cells to the micro centrifuge tube. **D)** Microscopically marked cells were collected and red lines around the cells were removed. **Right panel)** Representation of cell dissection from tissue into the cap of eppendorf tube by using laser light. Image source: Carl Zeiss Mikroscopi (http://www.zeiss.dk/Mikroskopi/Produkter/AP)

Since the tissues were kept at room temperature and fixed with formalin and paraformaldehyde (FFPE), there is a risk of partial RNA degradation and loss of the poly-A tail, which is necessary for cDNA production, from the collected RNA. Therefore, low molecular weight RNA amplification kit which helps the synthesis of cDNA from partially degraded RNA by adding a poly-A tail and increasing the amount of RNA by *in-vitro* transcription was used. After the production of cDNA, samples were used for real-time RT-PCR; a 5S RNA primer was used as a house keeping gene. **Figure 2.3A** shows the results indicating the upregulation of

23

CERIG in invasive and in-situ samples with respect to the benign cells. In 80% of invasive cancer cell samples, *CERIG* mRNA was significantly upregulated when compared to benign epithelial cells. These data suggest that *CERIG* is expressed mainly in the epithelial compartment of breast cancer. There is no significant difference between the invasive and in-situ samples, but there is a tendency for higher *CERIG* expression in invasive samples. Moreover, by mining datasets from NCBI (GEO databases) and Oncomines (Cancer Profiling Database), upregulation of *CERIG* was also demonstrated in various human cancer specimens as compared to adjacent normal tissues (**Fig 2.3B**). Together these data indicated a critical requirement of *CERIG* for cancer progression.
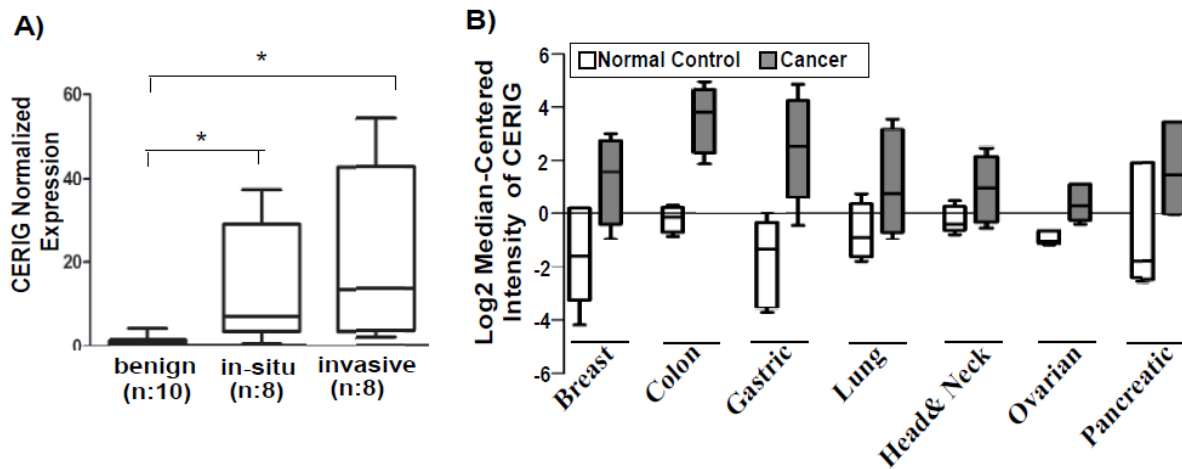
Figure 2.3) *CERIG* **mRNA expression in various cancer types A)** Archived specimens from benign (n=10), in-situ (n: 8) and invasive (n=8) breast cancers were collected and mRNA of *CERIG* in formalin-fixed, paraffin-embedded specimens was assessed. A laser capture microdissection (LCM) technique (Espina et al., 2006)was employed to collect both pure population of cancer cells and benign epithelial cells from specimens. Total RNA was isolated from LCM-collected cells by nano-RNA isolation method and mRNAs were amplified followed by Q-PCR using a specific primer set for *CERIG*. 18 times more *CERIG* expression was found in invasive breast cancer cells with respect to the normal cells (p<0.05). These results suggest that *CERIG* is expressed mainly in

the epithelial compartment of breast cancer. **B)** By mining Oncomine and GEO databases, *CERIG* expression pattern in more than 40 microarray data sets shows significant alteration (P<0.01). Representative data are presented. High *CERIG* expression in various human cancers.1-breast cancer n:27,normal n:7,p=6.97E-4 (Radvanyi et al., 2005); 2-colon cancer n:36,normal n:24, p=1.7E-4(Skrzypczak et al., 2010); 3-gastric cancer n:26,normal n:31,p=3.69E-13(D'Errico et al., 2009); 4-lung cancer n:45,normal n:65,p=8.59E-9(Hou et al., 2010); 5-head&neck cancer n:41,normal n:13,p=2.35E-7(Ginos et al., 2004); 6-ovarian cancer n:6, normal n:4,p=5.92E-4(Adib et al., 2004); 7-pancreatic cancer n:11,normal n:11,p=0.001 (Grutzmann et al., 2004).

2.2.2 Protein Expression and Immunohistochemistry

After the detection of mRNA upregulation in breast cancer cells and primary tissues, the protein expression level was investigated in the same samples. Polyclonal antibody against a recombinant CERIG protein from amino acid Gly^{1108} to Thr^{1340} was produced in rabbit. Using western blotting, the generated antibody was shown to recognize the recombinant protein and protein isolated from COS-1 cells transfected with full-length *CERIG* cDNA.

Western Blot of CERIG in cancer cell lines

In agreement with the real time RT-PCR data, endogenous CERIG protein was detected in MDA-MB-231, MDA-MB-435 at around 155 kDa, but not in MCF-7. As a positive control, CERIG is overexpressed in COS-1 cells, and our polyclonal antibody was shown to detect the ectopic expression of CERIG (**Fig 2.4**).
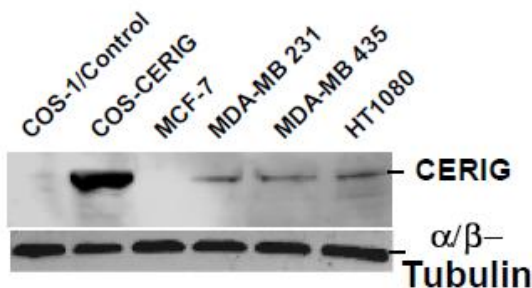
Figure 2.4) Western blot of CERIG. Using recombinant CERIG (rCERIG from Gly^{1108} to Thr^{1340}, 30 kDa), a polyclonal antibody was generated. CERIG in transfected COS-1 cells was detected as 155 kDa protein (Lane2). Endogenous CERIG was detected in aggressive cancer cell lines, but not in MCF-7 cells. Expression of α/β protein served as a control (bottom row).

Immunohistochemistry of CERIG in breast cancer tissues

After the identification of higher level of *CERIG* mRNA expression in invasive and in-situ samples compared to the normal, immunohistochemical analyses using polyclonal CERIG antibodies were performed for CERIG on invasive cancers as compared to noncancerous tissues. Negative controls also include substitution of the primary antibodies with irrelevant antibodies and blocking studies utilizing immunizing recombinant protein. CERIG immunoreactivity was present in the cytoplasm of invasive ductal carcinoma cells and invasive lobular carcinoma cells, whereas minimal or no expression of CERIG was found in benign breast or adjacent normal epithelial or stromal cells within the cancer specimen (**Fig 2.5**).Minimal staining was observed in negative controls. Particularly, cytoplasmic localization of CERIG was easily seen in the lobular invasive carcinoma.
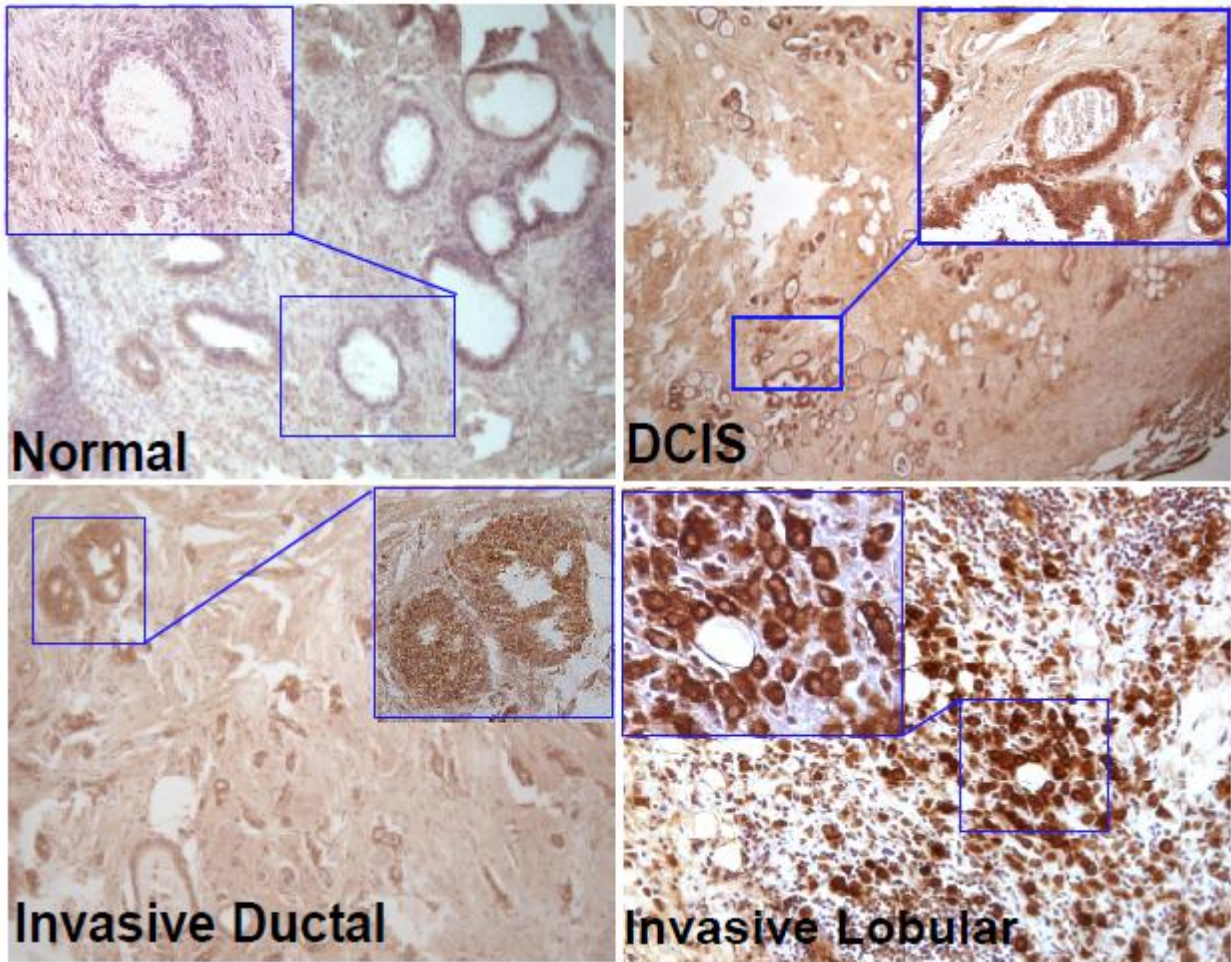
Figure 2.5) IHC staining of CERIG breast cancer samples. CERIG immunohistochemical staining demonstrates the presence of CERIG in breast cancer cells. In lobular carcinoma, CERIG localization in the cytoplasm was easily observed. Minimal staining was observed in negative controls. Enlarged picture of epithelial cells are shown in the corner of each sample.

2.2.3. Survival Curve Analysis

After the Van de Vijver study in 2002, there was a great interest in evaluating the gene expression profile of cancer patient at large scale. In 2005, three groups in Europe investigated large numbers of breast cancer patients independently, and then produced three different cohorts (Miller et al., 2005; Pawitan et al., 2005; Wang et al., 2005). To gain insight into the clinical

significance of *CERIG* in patients with breast cancer, these three publicly available DNA

microarray datasets, also known as Stockholm, Rotterdam, and Uppsala cohorts were analyzed.

Mean expression value of *CERIG* was calculated for each microarray data individually, and then

patients in these datasets were divided into two groups based on the mean value of *CERIG*

mRNA expression. After the dichotomization of patients at the mean value, patients were called

"high *CERIG* group" if their *CERIG* expression was above the mean value, and called "low

CERIG groups" if the *CERIG* expression was below the mean value. In all cohorts, low

expression of *CERIG* was found to be significantly associated with higher overall survival rate

according to the Kaplan-Meier analysis (p= 0.0348, 0.016, and 0.0213, respectively) (**Fig 2.6**).

Hence, elevated expression levels of *CERIG* in breast cancer correlates with a poor prognosis,

and expression of *CERIG* may be used as a prognostic marker.
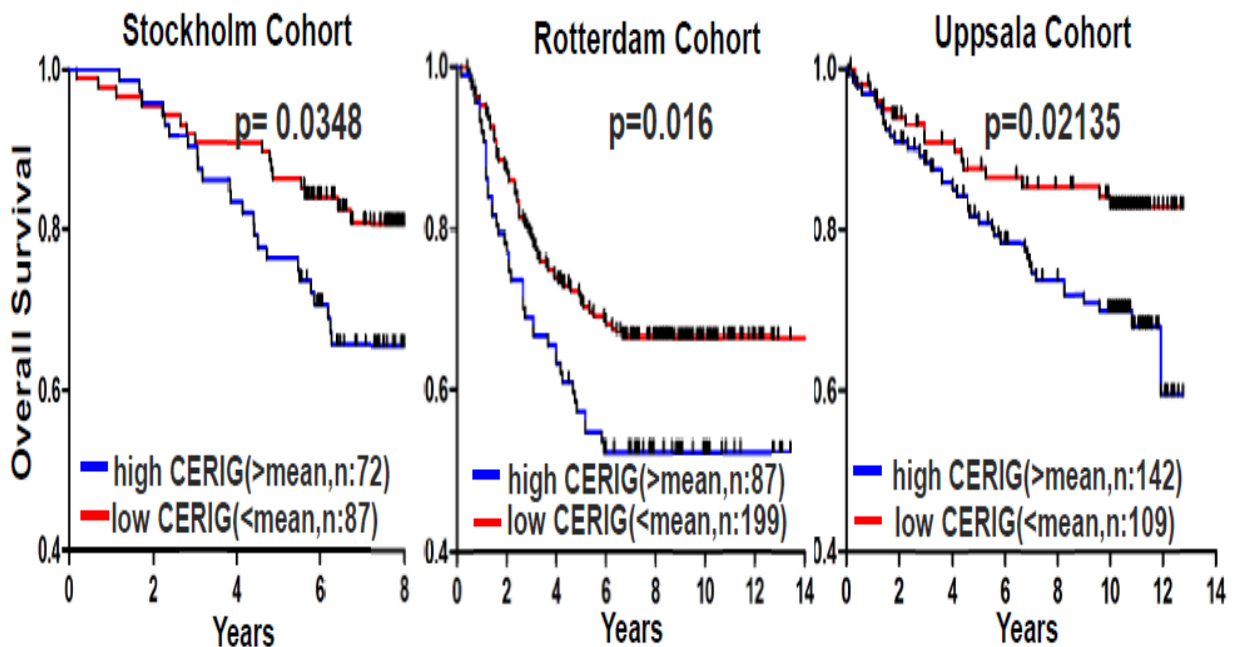
Figure 2.6) Survival curve analysis of *CERIG*. Three publicly available DNA
microarray datasets, which contain a large number of breast cancer patient
samples, are analyzed to gain insight into the clinical significance of *CERIG*

in patients with breast cancer. When patient samples are grouped based on *CERIG* mRNA expression dichotomized at the mean value in the cohorts of Stockholm, Rotterdam, and Uppsala studies (Miller et al., 2005; Pawitan et al., 2005; Wang et al., 2005), high expression of *CERIG* was found to be significantly associated with disease-free and overall survival rate by Kaplan-Meier analysis (p= 0.035, 0.016, and 0.021, respectively).

2.2.4 *CERIG* has minimal role on cell proliferation

Once the role of *CERIG* in cell migration and invasion, and its overexpression in different cancer cell lines as well as human primary cancer tissues were identified, our next aim was to understand the role of *CERIG* in cell proliferation. Therefore, MCF-7 cells that have ectopic expression of *CERIG* or GFP were used in an MTT assay. Wild type MCF-7 cells were used as a negative control. **Figure 2.7** shows that overexpression of *CERIG* did not increase the proliferative ability of epithelial breast cancer cell lines. On the other hand, MDA-MB-231 and MDA-MB-435 cells lines in which *CERIG* expression was silenced via shRNA were used to test the effect of *CERIG* silencing on the rate of proliferation. In agreement with the results of overexpression, silencing of *CERIG* expression did not change the proliferation ability of MDA-MB-231 or MDA-MB-435 cells. Even though *CERIG* silencing increased the proliferation rate of MDA-MB-435 cells, it was not significantly different from the rate of wild type MDA-MB-435. Taken together, these results indicated that *CERIG* has minimal or no role on proliferation capacity of the cells, supporting the idea that the reduced tumor volume in the *CERIG* shRNA group in the animal study was not due to an apparent effect on cell proliferation by *CERIG* shRNA cells.
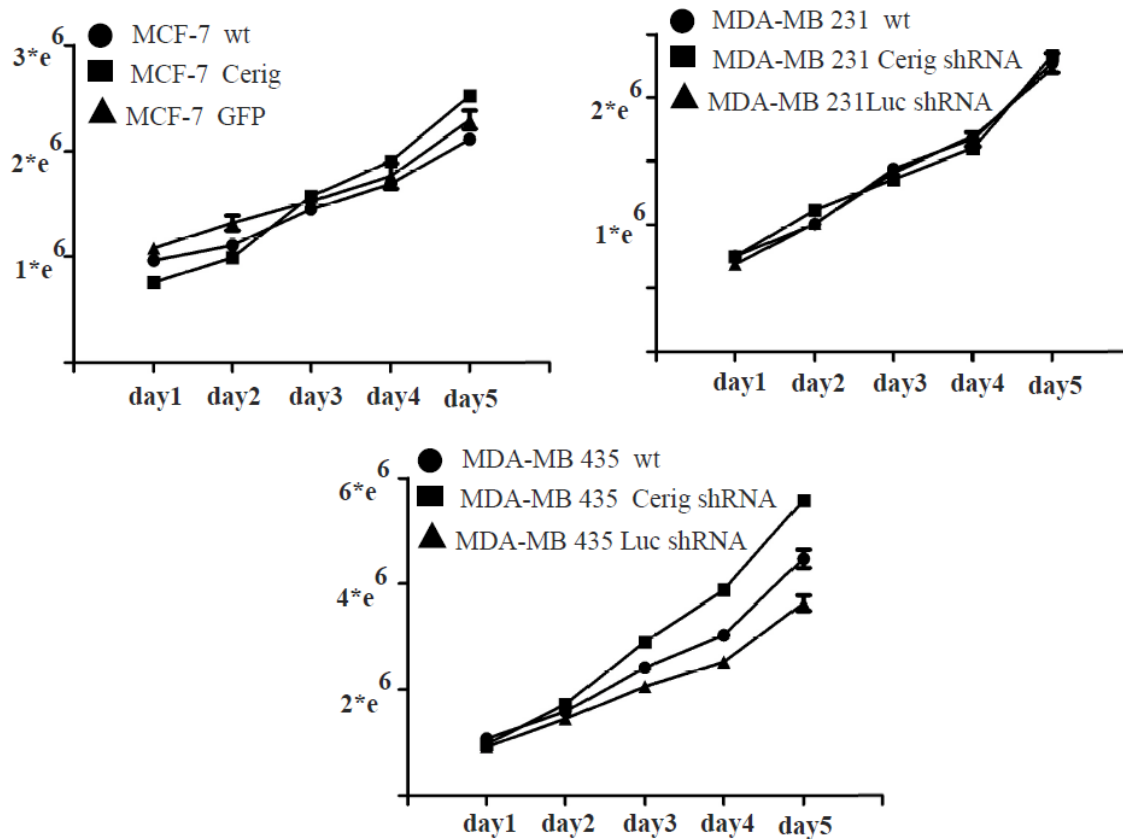
Figure 2.7) Effects of *CERIG* in cell proliferation. 5000 cells are cultured in the well of 96-well plate for each cell type. After 24 hours cells are harvested in the MTT assay and proliferation rate of each cells are measured by 24 hour interval for the next four days. Absorbance value of each cells are measured at 570 nm after overnight incubation to completely solubilize the formazan crystals. Y-axis shows the reading of luminescence.

2.3 Discussion

Human *CERIG* was originally identified as a product of an inner-ear specific gene (Abe et al., 2003), and its abundant expression was observed in spinal cord and brain tissue according to data mining of the NCBI/GEO profile (Yanai et al., 2005). In addition, *CERIG* expression, which was originally found in the central nervous system (CNS), has been shown to be upregulated in several different cancer types including breast, colon and gastric (Birkenkamp-Demtroder et al., 2011; Matsuzaki et al., 2009; Sabates-Bellver et al., 2007). Of interest, the expression profile of *CERIG* looks like other genes whose expression was originally found in the CNS, but later was reported to be upregulated in several different cancer types such as N-Cadherin in breast cancer (Nagi et al., 2005), NCAM in ovarian cancer (Zecchini et al., 2011) (Lehembre et al., 2008), and gamma-synuclein in breast carcinoma (Manivel, Muthukumaran, Kannan, & Krishna, 2011). Moreover, effects of *CERIG* expression in cell migration/invasion were demonstrated in our previous study.

In the previous study, RNA isolated from bulk tissue of cancer samples did not demonstrate significant difference of *CERIG* expression between the cancer cells and benign cells. If RNA is isolated from microdissected epithelial cells, 18-fold more *CERIG* expression in the invasive cells with respect to the control was revealed. Therefore, higher *CERIG* expression exists in the epithelial compartment of the breast cancer cells. In agreement with the mRNA expression profile, CERIG protein was also identified in the invasive and in-situ carcinoma samples compared to the minimal staining of the *CERIG* in benign tissue. These results were also supported by *CERIG* expression profile in the breast cancer and prostate cancer cells in which high *CERIG* transcription was related with the aggressiveness of the cell type. High expression of *CERIG* in colon cancer cell lines which showed higher metastasis rate to the lymph node

further supports the crucial role of *CERIG* in cancer settings. So, one can speculate that *CERIG* expression might be necessary to initiate the migration and invasion ability of the cells, and its function should stay at the same or higher level in the later stages of the tumor development. Moreover, our proliferation study further demonstrated the minimal role of *CERIG* expression in the proliferation rate of at least breast cancers cells. When mRNA level of *CERIG* and IHC staining of *CERIG* in invasive and in-situ samples are considered together, both results demonstrated the importance of *CERIG* in the early stage of breast cancer. In addition, level of *CERIG* did not change in the later stage of breast cancer since it has similar level of expression at *in-situ* samples. From survival curve analysis, high *CERIG* expression is also related with the poor prognosis, indicating the ongoing function of *CERIG* form early stage to the late stage of cancer. Collectively, our previous studies for the function of *CERIG* and recent studies exploring the expression level of *CERIG* cancer identified the upregulation of *CERIG* in human breast cancer.

2.4 Material and Methods

Cell Culture & Reagents:

COS-1 monkey epithelial, human fibrosarcoma HT1080, breast cancer MCF-7,MDA-MB-231 and MDA-MB-435 cell lines were purchased from ATCC (Manassas,VA) and were cultured in Dulbecco's Modified Eagle's Medium (DMEM) (Invitrogen) plus 10% FBS. HUVEC (*Human Umbilical Vein Endothelial Cells*) were purchased from ATCC and cultured in the human endothelial SFM media plus 20% FBS, 10 ng/ml b-FGF and 10 ng/ml EGF. SW 480 cell line was cultured in RPMI 1640 containing 10% FBS. HCT-116 and HT-29 cell lines were purchased from ATCC (Manassas, VA) and cultured in McCOY media including 10 % FBS.

RNA Isolation & Real-Time PCR

Cells were grown in 6-cm tissue culture plate for RNA isolation. Cells were collected into RLT lysis buffer containing 10% beta-mercaptoethanol. RNAs were isolated by using Qiagen RNeasy Kit according to the manufacturer's instructions. Concentration of RNA was measured by Nanodrop (ThermoScientific, DE). For real-time PCR, RNA was converted into cDNA by using iScript cDNA synthesis (BioRad) according to the manual. Real-time PCR was performed in the iQ SyberGreen super mix (BioRad) and Ct values were measured in the iCycler PCR machine (BioRad). Following primers were used for the amplification of *CERIG* and HPRT-1 (housekeeping gene). Relative expression was calculated using the ΔΔCt method including Hprt-1 as an internal control.

CERIG-for: gagatagacggctggacat, *CERIG*-rev: agctgctgtcccatatgctt

Hprt-1-for: accccacgaagtgttggata, Hprt-1 –rev: aagcagatggccacagaact

Hematoxylin and Eosin (H&E) staining

Five micron FFPE sections were collected on glass slides and incubated at 65°C for 2 hours. After drying the sections on a glass slide, sections were merged into xylene solution for 10 minutes, twice. Sections then were rehydrated in serial ethanol solution from 100% EtOH into distilled water by following order; 100% -95% -70% -50% EtOH and water. The slides were stained with nuclear dye (Mayer's hematoxylin solution) for 5 minutes followed by rinse with distilled water. Then, slides were incubated in acid alcohol solution which was prepared by mixing 0.1% HCl and 50% EtOH for 3 seconds. After washing slides in tap water, they were treated with counter stain (eosin) for 30 seconds to 1 minute. Slides then were dehydrated in several ethanol solutions in the following order; 70% -95% -100% EtOH. Finally, slides were

treated with xylene for 5 minutes, twice. In the case of LCM (laser capture microdissection), slides were not treated with xylene and not covered with coverslips.

LCM (Laser Capture Microdissection)

H&E (hematoxylin-eosin) stained samples without xylene treatment were used for cell capturing via laser. Cancer cells or benign epithelial cells were collected from the surface of slides by using P.A.L.M ROBO application on the computer. Brief procedure has following steps; UV-energy was set at 82 which can be adjusted according to the softness of the tissue. UV-Focus was set at 76 and dissection speed was adjusted to the 75 steps /second. Image was focused automatically under 5X lens. Microcentrifuge cap was filled with 40 µl of distilled water and placed upside down. The area of interest has been designated on the image. The specimens were then subjected to the dissection by hitting the laser in the designated area on the glass slide and cells were collected in the upside down microcentrifuge cap filled with water. At the end of dissection, 200 µl of ethanol was added into the microcentrifuge tube. Tubes were centrifuged at 13200 rpm for 15 minutes. After centrifugation, most of the ethanol was removed from tube and cells were used to extract the RNA or DNA.

RNA Nanoprep Kit

Since there is a limited amount of cell sections collected from LCM, Absolutely RNA Nanoprep Kit (Stratagene) was used according to the instructions of the company. In addition to the kit, sulfolane was purchased from Sigma (T22209) to increase the yield of isolated RNA. RNA was eluted in minimal volume of elution buffer, i.e 10 µl.

Low Molecular Weight RNA Amplification

To increase the yield of RNA isolated from archived sample and to produce cDNA, low molecular RNA Amplification Kit (Genisphere) was used according to the protocol. Brief procedure has following steps; Poly A polymerase was used to add poly (A) tail to the purified RNA from Nanoprep Kit. By using dT primer, first strand cDNA synthesis was performed and RNA was degraded. Terminal Deoxynucleotidly Transferase (TDT) was used to add poly (dT) tail at the 3'termini of cDNA. The T7 template was annealed to the 3'end of the cDNA. Klenow enzyme filled in the 3'end of first strand cDNA to produce a double-stranded T7 promoter. The T7 template contains a blocker to prevent second strand synthesis. Sense RNA copies of the original RNA molecules were generated by in-vitro transcription. Amplified RNA was converted into cDNA followed by real-time PCR by using the same primer pairs and experimental approach as mentioned in the real-time PCR section.

Western Blot

Polyclonal antibody was produced against *CERIG* recombinant protein from amino acid Gly^{1108} to Thr^{1340} in rabbit. Lysates were collected in the phospholysis buffer and mixed with 2x SDS gel-loading buffer. The samples were resolved by 10% polyacrylamide gel electrophoresis, and proteins were transferred to nitrocellulose membranes and probed with antibodies. After extensive washing with TBS-T (20mMTris–HCl, pH 7.6, 137mMNaCl, 0.1% Tween), the membrane was probed with anti-rabbit or anti-mouse IgG for corresponding primary antibodies and was visualized by the Western Lightning Chemiluminescence Reagent (Perkin-Elmer).

Immunohistochemistry

Invasive and benign breast sections were cut at 5μm on slides and slides were rehydrated into TBS. Slides were autoclaved on wet cycle for 5 minutes in antigen retrieval buffer, pH 4. Samples were blocked for 1 hour by using the TSA blocking buffer. Rabbit polyclonal *CERIG* antibody was diluted 1:100 in TSA blocking buffer and samples were incubated with the antibody for overnight at 4 $^{\circ}$C. Slides were then washed four times by using TBS, 5 minutes for each. Specimens were then bleached into 3% H_20_2 for 1 hour and washed with TBS buffer. Second antibody (anti-rabbit HRP) was diluted at 1:100 in TBS and incubated on samples for 1 hour. TSA amplification was performed by using TSA-XX Biotin (Molecular Probes, OR) for 10 minutes. After standard washing procedure, SA-HRP was diluted 1:100 in TBS and added on top of the samples for half hour. At the end, samples were treated with DAB to see the color change.

Survival Curve Analysis

Three independent data sets were evaluated for prognostic role of *CERIG* in breast cancer. All datasets were obtained through the NCBI/GenBank GEO data base with the following accession numbers; Stockholm Cohort: GSE 1456, Rotterdam Cohort: GSE 2034, Uppsala: GSE 3494. In the Stockholm dataset, there are 159 patients with the mean prognosis age of 58 and mean tumor size of 22 mm. This cohort was a mixture of patients with and without adjuvant therapy. There are 286 patients with ages ranging from 26-83 and without any adjuvant therapy in Rotterdam cohort. There are 251 patients in the Uppsala cohort. Affymetrix Human133A GeneChip had been used to get the expression profile in all these 3 datasets and corresponding gene probe was identified by Affymetrix identification numbers for *CERIG*;212942_s_at. Dichotomization of the patients was done by using the average expression

value of the *CERIG* for each cohort. Expression level above the mean value was accepted high *CERIG* and expression level below the mean value was accepted as low *CERIG*. Kaplan-Meier survival curve was produced for each data set by using the clinical information of the high/low dichotomized samples. Later, Log Rank (Mantel Cox) was used to determine the statistical significance.

MTT Cell Proliferation

Five thousands MCF-7, MDA-MB-231, MDA-MB-435 and same cells with the *CERIG*/GFP knockdown or upregulated one were plated in each well of the 96-well plate followed by incubation at 37 ^{o}C with 5% CO_2 for five days. After each day, cell proliferation amount was measured by using the CellTiter 96® Non-Radioactive Cell Proliferation Assay (Promega) at 570 nm in the Luminometer. Cells were plated in triplicate and average value was calculated for each specific day.

Phagokinetic Cell Migration Assay

This procedure was performed as previously described (Albrecht-Buehler, 1977). Briefly, coverslips were treated with poly-L-lysine (50 μg/ml) and then coated with 0.5% BSA followed by air drying. Coverslips were placed in 12-well plate. Freshly made colloidal gold particles were heated until the boiling then wait for 5-10 seconds. The gold particles were immediately transferred to the coverslips in the 12-well plate for at least one hour. Coverslips were rinsed with PBS at least 3 times and soaked into 70% EtOH to sterilize the gold and PBS coated coverslips. Cells were plated onto treated coverslips and incubated at 37°C tissue culture incubator for 6h-18h. After culturing, cells were fixed with 4% paraformaldehyde in PBS. Migratory cells were observed and photographed under light microscopy (Nikon). Images were

processed and measured using NIH image software (Image J). The percent of phagocytosis was analyzed by image J which calculated the area of gold particles.

Transwell Cell Migration Assay

Polycarbonate membranes of 13 mm diameter with 8 µm pore size (Neuro Probe, MD) were inserted into the Blind-Well Chemotactic chambers (Neuro Probe, MD). Prior to seeding into the transwell inserts, COS-1 or MCF-7 cells were released from plates with trypsin-EDTA followed by the addition of DMEM plus 10% FBS media. The lower chemotactic chamber was filled with DMEM plus 10% FBS (200 µl). Cells were counted using a hemacytometer (Hausser Scientific, PA). The upper chamber was filled with 100,000 cells suspended in DMEM to a final volume of 200 µl. Chambers were incubated for 6 hours and 18 hours at 37 ^{o}C and 5% CO_2 in a humidified tissue culture incubator. The cells in the upper surface were then removed from the filter with a cotton swab and washed three times with PBS. The cells remaining on the lower surface were fixed in 4% PFA/PBS solution 20 min in a 4 ^{o}C refrigerator. Cells were stained with 0.1 % crystal violet for 20 min and examined under a light microscope. The number of cells in 10 different areas of the filter was counted to obtain the number of migrating cells.

Chapter 3: Promoter Characterization of *CERIG*

3.1 Introduction

Promoter is characterized as a DNA region located generally in the 5' upstream of a gene on the same strand of DNA and it has been known to initiate the transcription. Many studies show differences in the length of the functional promoter and the interacted elements on the promoter for each gene. By using the activator/repressor binding sites and putative transcriptional start site in the genomic DNA, potential promoter sequence or promoter start site can be predicted in the genome. Eukaryotic promoters are difficult to characterize but there are many algorithms for detection of promoters in genomic DNA. Many trans-acting elements, called transcription factors or repressors, bind on the specific site, called cis-acting elements on the promoter for the regulation of a gene. To better understand the molecular mechanisms and the function of a gene, promoter characterization sheds fresh light on regulatory mechanism; however, *CERIG* promoter was not investigated in this line of research so far. Thus current work was initiated to bring first knowledge about the transcriptional regulation of *CERIG* and to help us understanding the inhibition of the *CERIG* transcription in cancer cells for the treatment.

Nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB), identified as a transcription factor that regulates the κ-light chain expression in B-lymphocytes, constitutively exists in most cell types as homodimers or heterodimers of a family of structurally related proteins (May & Ghosh, 1997). Identified NF-κB family members include NF-κB1 (p50 or p105), NF-κB2 (p52 or p100), p65 or RelA, RelB and c-Rel (Nabel & Verma, 1993). These proteins share a highly conserved structure named Rel homology domain (RHD) that allows these factors to form homo- or heterodimers (Vallabhapurapu & Karin, 2009). The major active NF-κB complex in mammalian cells is a heterodimer containing a p50 and p65 complex that

binds to the consensus DNA motif within promoters. NF-κB works as "rapid-acting" transcription factors that do not need new translation to be a functional because of the conversion of inactive state in the cytoplasm into the active state after the appropriate induction. Aberrant regulation of NF-κB has been reported in many cancers, and targeting the NF-κB signaling is a way of treatment for cancer and inflammatory diseases (Garg & Aggarwal, 2002). General consensus sequence of NF-κB identified in promoter and enhancer region looks like GGG RNN YYC C, (R=purine Y=pyrimidine).

C-jun as a part of AP-1 is another member of the "rapid-acting" transcription factors. The activator protein 1 (AP-1) plays an important role in various human diseases and regulates the expression of multiple genes essential for cell proliferation, differentiation, and apoptosis (Ameyar, Wisniewska, & Weitzman, 2003; Mullenbrock, Shah, & Cooper, 2011). AP-1 consists of dimers composed of members of the Jun, Fos and activating transcription factor protein families (Hess, Angel, & Schorpp-Kistner, 2004). The different dimer combinations recognize different sequence elements in the promoters and enhancers of target genes. TPA DNA response element (TRE; 5'-$^{T}/_{G}$GA$^{G}/_{C}$TCA-3') was identified as an AP-1 binding site on the DNA (Hess et al., 2004).One of the intronless gene, c-jun, has been defined as a proto-oncogene because of the over expression of c-jun in many cancer types and its rapid induction by external growth factors. In this study, the functional characterization of the human *CERIG* promoter as well as *cis*-acting elements and *trans*-acting factors on regulation of *CERIG* expression were characterized. Furthermore, epigenetic regulation of the *CERIG* promoter region was examined in human breast cancer specimens as well as in breast cancer cell lines. The results from this study demonstrate, for the first time, a regulatory mechanism for *CERIG* gene expression.

3.2 Results

3.2.1 Molecular cloning of the potential *CERIG* promoter

To identify the potential promoter region of *CERIG* a bioinformatics approach employing two different promoter-prediction programs (Genomatix Model Inspector and Mc Promoter-Prediction Server at Duke University) was utilized (Cartharius et al., 2005). The 5'-flanking region covering a 1.4kb region was identified as the potential *CERIG* promoter between two putative TATA-boxes. Cloning of promoter was started from 2kb further upstream with respect to the above potential promoter whether other putative activator or suppressive regulator elements in these regions might control the *CERIG* transcription. To clone the 5'-flanking region of the *CERIG* gene containing the 3.3 kb fragment, a PCR approach was employed using BAC clone 96012 (Invitrogen, Clone ID: 2215F6) as a template. A 3.3 kb fragment consisting of the 5'-untranslated region and the coding sequence of the first exon of the *CERIG* gene was amplified and cloned into pGL3-basic vector that lacks a promoter for the luciferase reporter gene. This construct was used as a template to generate the shorter chimeric luciferase constructs used to identify the minimal promoter required for activity of *CERIG* in addition to *cis*-regulatory elements within the promoter.

3.2.2 Determination of transcription start site of *CERIG*

To determine the transcription start site of *CERIG*, a non-radioactive 5'-primer extension approach (Flouriot, Pope, Kenealy, & Gannon, 1997) was used. By utilizing biotin-labeled reverse primers to reverse transcribe mRNAs extracted from MDA-MB-231 cells; a single 250 bp fragment was identified, indicating the presence of a single transcription initiation site for the *CERIG* gene (**Fig 3.1A**). To further assess the location of the transcription start site in the *CERIG* promoter, Cap Analysis Gene Expression (*CAGE*) database, which measures expression levels of

transcription start sites by sequencing large amounts of the 5'-ends of transcripts, termed CAGE

tags (Kawaji et al., 2006) was employed. Our analysis based on this approach revealed that the

CERIG promoter region consists of three transcription start sites separated by one nucleotide

(**Fig 3.1B**). However, only one major transcription start site (G) was given in the sequence of

…**G**$^{+1}$AGCTAG…

Considering the above mentioned data and the fact that a 94 kb intron separates the

promoter region and the translation start site (AUG) located within the second exon of *CERIG*,

the transcription start site was denoted as **+1** as shown in **Figure 3.1C**. Relative to the

transcription start site, a canonical TATA-box in the -31/-27 region and a GC-box in the -248/-

243 region were identified in the 5'-flanking region of the *CERIG* promoter (**Fig. 3.1C**). Our

analysis suggests that the *CERIG* promoter is a single dominant peak promoter, which is

characterized by the existence of a TATA-box and a single transcription start site (Carninci et al.,

2006).

3.2.3 Identification of minimal region(s) required for basal promoter activity of *CERIG*

To elucidate the regulatory effects of different regions within the *CERIG* promoter on

transcriptional activity, two series of reporter constructs were generated containing fragments of

the human *CERIG* promoter coupled to the firefly luciferase gene (**Fig. 3.2**). Based on the

deduced *CERIG* promoter (1.4 kb), two deletion mutants were generated by sequentially

removing fragments of approximately 900 bp from the 5'-end of *pro*-3.3 (the -3.3kb construct),

to get *pro*-2.3 and from *pro*-2.3 to form *pro*-1.4. The reporter constructs, along with a Renilla

report gene that serves as a normalization control, were transfected into cells expressing

endogenous *CERIG* (e.g. MDA-MB-231) (**Fig. 2.1B**) and non-detectable levels of *CERIG* (e.g.

MCF-7 and COS-1 cells). The 5'-flanking region containing -1.425 kb relative to +1 site is

sufficient for near maximal activity of *CERIG* promoter in both *CERIG* positive and negative

cell lines (**Fig. 3.2A**). In contrast, expression of constructs *pro*-2.3 and *pro*-3.3 in these cells

demonstrated a decrease in luciferase activity. These data suggest that activating elements are

located in *pro*-1.4, whereas repressor element(s) is located in the region from -1425bp to -3322

bp. In addition, effect of repressor element(s) on reporter gene activity is more severe in MCF-7

since these elements decreased the reporter activity 80% in MCF-7 cells, but 50% in MDA-MB-

231 and COS-1. Since upregulated *CERIG* has been observed in human cancers, the activating

elements within *pro*-1.4 have been focused to determine the minimal region required for *CERIG* promoter activity.

To identify minimal regions required for transcriptional activity of the *CERIG* promoter, an extensive mutational analysis of the -1425bp *CERIG* promoter (*pro*-1.4) was performed. A series of deletion mutants of *pro*-1.4 were generated and expressed in COS-1 cells followed by a Dual-*Luciferase*® reporter assay. Two areas were found to contribute to *CERIG* gene regulation, including -1425 to -1135, and -125 to +27, demonstrated by a significant decrease in luciferase activity upon deletion of these regions (**Fig. 3.2B**). Removal of the transcription start site with TATA-box and other important *cis*-acting sites in construct *pro* -125 resulted in a significant reduction of luciferase activity to levels observed in cells expressing empty pGL3-basic vector (**Fig. 3.2B**). However, first exon did not contribute markedly to the reporter gene activity. These observations were recapitulated in *CERIG*-positive MDA-MB-231 cells as well as in negative MCF-7 cells (**Fig. 3.2B**). Deletions from -1135 to -125 (*pro*-1135, *pro*-913, *pro*-583, *pro*-217) did not result in significant changes in promoter activity, indicating a lack of critical regulatory elements within the region spanning -1135 to -125 of the *CERIG* promoter.
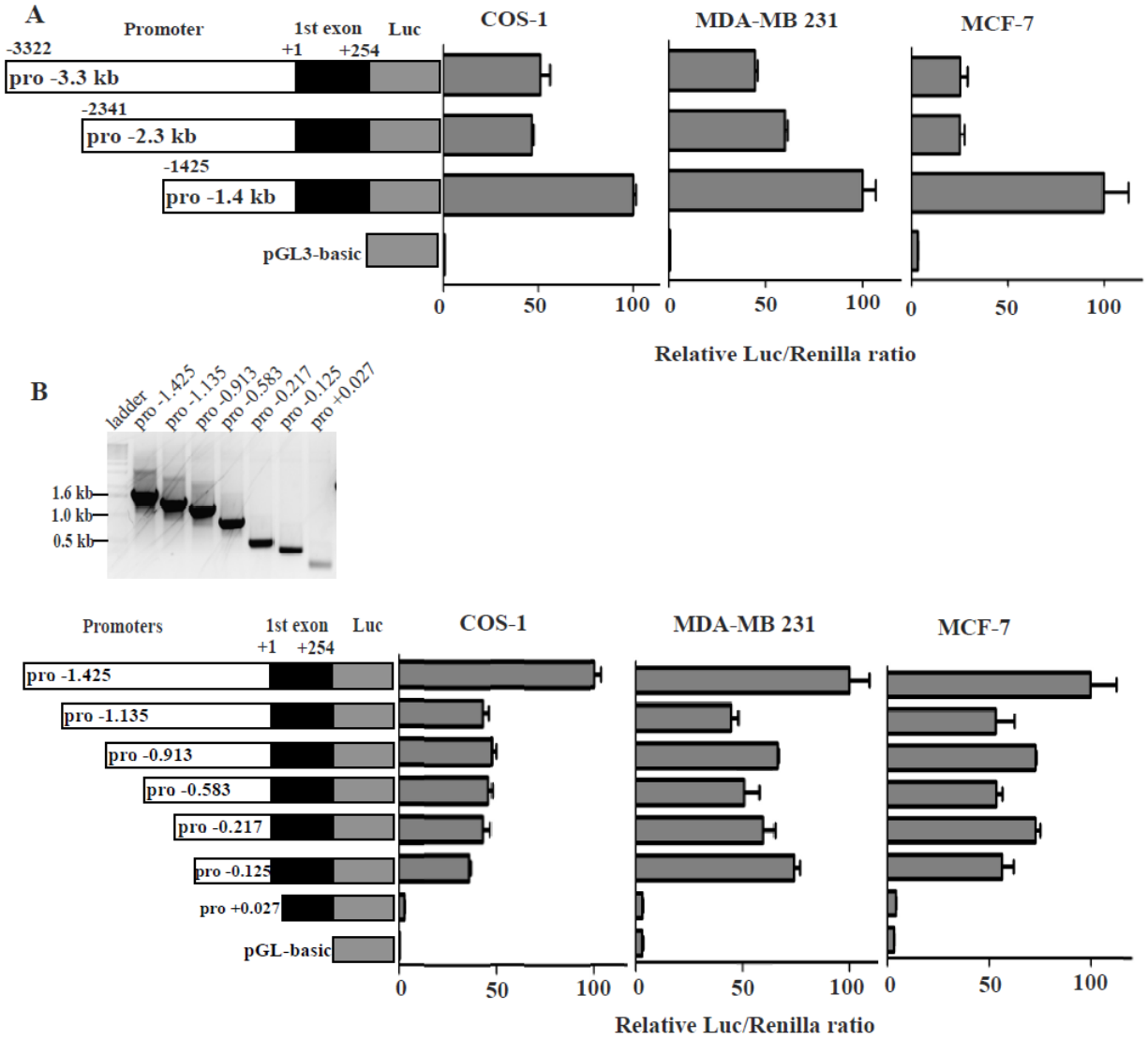
Figure 3.2) Deletion analyses of the *CERIG* promoter activity in breast cancers and COS-1. A-B) A series of the fragments of the 5'-flanking *CERIG* promoter region are schematically shown. The base positions relative to the main TSS are indicated. Each promoter-reporter construct or a promoterless plasmid, pGL3-basic, was co-transfected with Renilla-plasmid into breast cancer cell lines and COS-1. Luciferase activities were measured after 48 h and normalized for transfection efficiency. The luciferase activity of each construct is presented relative to the activity of Pro-1.4 (−1.425) construct. Three independent experiments with triplicate samples are shown for each construct. Agarose gel result of produced fragment is shown in the upper site of panel B. Error bars, mean +/_ S.E.

3.2.4 Computational analysis of the putative transcription factor-binding sites within the *CERIG* promoter

To identify putative transcription factor-binding sites within the *CERIG* promoter, two programs (MatInspector and Alibaba2) were employed to predict transcription factor binding sites based on DNA sequence. This analysis revealed many transcriptional elements in the *CERIG* promoter, but I decided to focus on AP-1, NF-κB, and Twist-1 for further evaluation. All these transcription factors are located in the activator region (*pro*1.4) of promoter with their high ci-value which indicates the chance of interaction between transcription factors and promoter (**Fig. 3.3**). AP-1 consensus binding sequence (GAGT) that is located between -48 and -45 and its flanking regions on both sides are highly conserved (100%) between human and mouse. It is also located to the close proximity of transcription start site. Since NF-κB has four putative binding sites in diverse sites of promoter and has aberrant expression in cancer progression, it needs to be analyzed. Although four putative NF-κB transcription binding sites were recognized, the first three putative NF-κB binding sites located in the proximal part of the promoter (-246/-234, -324/-312, and -704/-715), may not be involved in activation of the *CERIG* promoter based on the finding that deletion mutants containing these three NF-κB binding sites did not affect transcriptional activity of the *CERIG* promoter (**Fig. 3.2B**). In contrast, the deletion mutant encompassing the fourth putative NF-κB binding site (-1345 and -1333) lost approximately 50% transcriptional activity (**Fig. 3.2B**). In addition to AP-1 and NF-κB binding sites, two putative Twist-1 binding sites (-269/-249 and -923 /-903) were also identified. Since Twist-1 has been shown to upregulate during EMT and *CERIG* has a role in migration and invasion which are important events of EMT, knowing the possible role of Twist-1 in *CERIG* regulation might be noteworthy. However, deletion of the region containing these Twist-1 putative binding sites did

not notably change luciferase activity. However, since transcription factors are often conditionally expressed in certain pathological conditions, the role of AP-1, NF-κB, and Twist-1 in transcriptional activation of *CERIG* needs to be further examined.

3.2.5 Requirement of the AP-1 element in activation of the *CERIG* promoter

To examine the role of AP-1 in *CERIG* promoter activity, additional reporter gene constructs in which the putative AP-1 binding core sequence (GAGT) is either deleted (pro-1.4Δ^{AP-1}) or mutated to AGAC (pro-1.4^{Swap}) was produced by a PCR approach. The engineered mutants coupled to firefly luciferase cDNAs were co-transfected with Renilla cDNAs into COS-1 cells followed by a Dual-*Luciferase*® reporter assay. Mutations of the AP-1 site either by deletion or substitution approaches resulted in significant loss (90%) of luciferase activity (Fig. 5A). This observation was reproducible in *CERIG*-high (MDA-MB-231) and low (MCF-7) cells (**Fig. 3.4A**), suggesting the requirement for AP-1 in activation of the *CERIG* promoter.

```
human  CAAAGAAGGGCTGAGAATCCTGACCACCACCCCCAGTAAGGATCAAGTTCCGCTTTCTGG -1365   T---TTCTCGGG-CT--ACTCTTTTGCTGCTGAGTTAAAAAGAGATCCTGAGACTTAGCC -452
mouse  --------------------------------------------TGGTTCCAACACATG- -1385   AGGATTCTGGGAACTGAACTCTGTTTTTTCAAACCTGTAAACATCTTAGCGGTCATCTCC -496
                                                   :.*****.. : .**        :  ***** . ** .** *.* .*:*. .**

                             NFkB (4)
human  AAGAGGCCTAGAGGTGGATTGCAGGACTTCCCAGTTTAACCAGGTGACAAAGGCTTCTGC -1305   TGGGACCCCGACTCGCGG-CGATTCTGTGGCAGGCGCAGACCTTTC--TCCTGTCCTCC  -395
mouse  ------GCCCAAAG-------GCTG--CTGCCCAG-------AG--GACCTAG-------- -1356   CTAGCTTTCAAAACAAGATCTTCTCTAAGAGACCTCTAGACCTAGGCGATCCTGGATTCT -436
              ***  *.**     **:* ** ***** **  ***. :**         .*.  *.*:*..* .  : ***.:* .*    ******:  *  ***** . **

human  TTCAGCCCAGATGTGATCACAGACTTCCACAGCTTCTGCCGACCACAGCCTGGCCTTAAA -1245   ACTGCCTCTTTGCCT--CGGAGGGCGCCCCGGG--------------GCGCGAGCGGGGC -351
mouse  ----------------ACAGGGGCTTCCAAAGTCTCTG----CCATAGACTGGCCCTAAA -1316   GTTGCTATGTTGGCTTACAAAAAGTGTTACGGGTTTGAAATCTAACAGCCTGAACGTAGA -376
                       .:**  .*.****** **  **** ** ******* ****        *** *** ** *.*.* *  ****         ** **.**  **

human  GTCCAGACCTAGAGTCCCGGAGCTAGTGTGTACTACACAGAATGGAGACCTTGCCTGAGG -1185   ACG--GGCGACCCTG--------------------GCT-------CCTTGG--AGGCCGG -321
mouse  GCCTGCACTGAGATTCCTAAAGCTGGTGGGCACTGCCTAGAATGAAGACCCT-CTCGCAC -1257   GTGGTGGCGACTCTGTCAAAATCTCTTTTTTTCTGGCTGACTGTCCCTTGGGGAGGCCACT -316
       * *. ***  *  ***  .  ***. ***. .*** . *.***.*.***. *.:**      * ****** ***         ***  ****** *** *

                                                                             NFkB (2)
human  AGCTGGTTCTCAGAGAACACTGGCAGGAAGGAAGAAAGAAATGTATATTCTAGGAGGCCT -1125   GG--ACTTGCT----CTCCG-CGGTCCGCCGCGGCCGCTCCACCG----AGTCCTGGCCT -272
mouse  AGCAG-TGAGCATAGGACATTGGCAGGCAGGAAGAAA------TGTGTACTTG-AGAACT -1205   GGATACAAGCTGCGCTTCCATCACTAAGCCCTCGGAGGTCCGCAGCTTTCGTCCTAGGCT -256
       ***.* .*.  . *** .* ********. ********      *.*.*:*.:* ***     **  **:.*  ****.* . ***. .* *** *.*    *****.* **

                                                Twist-1 (1)            NFkB (1)
human  CTGCT-AACCAGGAAAGGGGGCTGAGTCACAC-TCTGAAAGGTGGAAGAGGCAGAGAGCA -1067   CCAGCGCCCTCACCTGGCTGCTGCGGGCGGGACTCTGCGCACCCGGGCTCCGCGGGCTCA -212
mouse  CTCATTAACCAGAGGAGGGGAGCTGAGCAATAAGTCTCAAAGAGGGACAGCATATAGAATT -1145   CCAGCGCCCTCACCTGGTGGCCACGG-CTTGTCCCGGCGG---CGG-CTCT-CGGGCTCA -202
       ** .* *****. .. **** . ****** . .  .**** ** **.* *. *  * ***     *************** . ***.  ** *.:* * *.. *** *** ********

human  GGTATCAACGCCCTCATTTTACAGATTAAAACAAACAAGAAAAACAAACACTCCTCACC  -1007   GGTCTGCGTCTCCCAGAGCCGCCGCCAGAGCCCGCCCCGGTCCGCCCGGGCACAGGGCC  -152
mouse  GCAACCAAAGCTGTCATCTT--AGAATTAAAATGAAATGGAAAAAAAAAATCCCCAGCACC -1087   GGTCTGCGGCTCCCGGAGCGGCCGCCAGAGCCCGCCCGGGTCCAGCCCCG---CAGG---- -148
       *  *  *** **  ****  *  ***.*  **  .* .. ****. **:* ***** *.**    ******** ***** ****  ****************** *.***** ** *

human  TCCCTGCAAAAAAACCACCCTGAGGCTCAGACAGGGGAAGTAGCTG--GCCCTGGTCACA  -949   CCCGCCCCACTTGGGCAAGGCGGGGCGCGTGGAGGGAAGTTTCATCGGGGGGCGAGGGA  -92
mouse  TTTCACCCTTTTTGCCA------GGGTCAGAGTGGATAGGTAGCTGGTGCTCAGGGCACC  -1033   -CCGCCCACCGCGGTGAGTGGGAGGACGC----TGCCCAGAGACAGGGGTGGGACAGGGG  -93
       *  *.*   * : ** **       ***.**.. .**. :********  ** .*** .*.*       ****** **..**. . * * .** ***          .:*  :** ** .*** ****.

                              Twist-1 (2)                                                       AP-1
human  CAGT--AGGTTAGCGGCAG--AGTTGAGAGGAGACTCCATATGCCCTAGGGATGTGTTGT  -893   GGGAGCTTCGGTGCCGCGCGGGCGTCAGGGGCTCCTCGGGGGGGAGGAGTCAG----TAAAA -36
mouse  CAGTGTTGAGGAGATGCTAGATGCTCTGGGTGGGCTTACTTATTCCTATGGCAGAGG-AT  -974   GGACG-----GGATGAGTGCGGGCGCCGGGGGCTCTGGTGGGGAGGAGTCAGGCCCGGAAG  -37
       ****   .*  .  ***.  *.* :. ** .** **  . * :: .** **  .:*.     **. *  *.******. ******* ***********      **

                                                                            TATA-Box                    +1(TSS)
human  GATGAACTTTTCCTACTGGTACTGTTTCCTCCCCGCGAGGG-AATGTCTAGACCAGCCGCA -834   GTGGCTATAAAAG-CGGCCTCTCGG-CTGAGGGCCGGGAGCTAGCGCTCAAGCAGAGCC  +22
mouse  AACTCACTGGTGGCACTG-TCTTGTTTTGATCCTTGAGAACATTGTCCTGTCCAGGCACA  -915   GCGATATAAAAGCGTCGGCTCGGGCTGGCAGCTCC-GAGCTAGCTACTAAGCCAGTGA  +22
        .  *  *** . .*****  **: ***  .* ***:** *.    * ***** *.*:**** *:.***   * * *.*****    * *.*. *** *.   ****.**** ..** . ****.

human  -CCTTCTTGCT--TTGACCCTTCAGAACTTTGGCCTG-TCCAGTTAAAGAGGCACAGAG-  -779   CAGCGCGGTG-CTATCGGACAGAGCCT-----GGCGAGCGCAAGCGGCGCGGGGAGCCAG  +76
mouse  GACTACATGCAGGATAGCAAACACTTACTATGACTTAGTACCCACAAAATCTCACTGAAA  -855   GCGCGCGGAGTCGAGCGAGGAGCGCCTTCTGCGCTGAGTGGCAGTGGTCGGTGCTGCTG  +82
        .* : ***:   . *.* *  :*  *  **  * *: *: ** .* ****: *.* **.      .******:* *  ** ** ****       * *** * .** **     ** *   *:*

human  ----CCCTCCT-----------ACCCACAGGGAATGTTTCTAACTTAC---CAAGCACCT  -737   -----CGGGGCTGAGCGCGGCCAGGGTCTGAACCCAGATTTCCCAGACTAGCTACCACTC  +131
mouse  TCCTCTCTCCAGGCTCATGGGAGCCCAATGGTGAAGTCCCTTGTTTATGGTCAAGTAACT  -795   GAGCACTGTGCCGAGTGAGGTCGGAATCTGCACCCCCACTTTCCG---CCAGCTCCCGTCC +139
             * ****            ***** **   **** **** *: .   ****.*.** *       * ***** .*  *.** *** . ***** ****** *    .*** ** .*

                   NFkB (3)
human  ATTATGTGCTTAGAACTTGGACTT-TCCTGATCTCATTTAAGCCTCACCATTCTAGTGAG  -678   CGCTTGCCCACGCCCCGGGAGCTCGCGGCGCCTGGCGGTCAGCGACCAGACGTCCGGGGC  +191
mouse  GGCATGTGCAGAAAACTGACACATGTGAAGAAGTAACTCTACTATCCGTGTGTGTGTGTG  -735   CACTTGCTTACGCTCAGG-AGCTCAA--CGCGTGG-GATCAGATACTAGAC-TACGGG-A  +193
        * .***** * ***** .**. *  *:* * *:. .* . . :. *** .  .  :*:*:    *.***** .***** ***.  **.* ** **.***. ** ***** *  *.**** *

human  GAAGCATCACAG-AACCCATTTTCCTGATG--AGGAAACTGAGACTGGGAGGGGTTAAG-  -622   CGCTGCGCTCCTGGCCCGCGAGGCGTGACACTGTCTCGGCTACAGACCCAGAGGTAAGAG  +251
mouse  TGTGTATGTGTGTGTCCCCTAAGACAGAGGTCAAGCTGCTTGGAGTGTTATTTGTTTTTG  -675   TGCTGCACTCCAGATCCCCGAGGCGTGGCAGGGTCTCGG-TCATCAACCTGAGGGAGGAA  +252
        *:* *:*.:*.* * :*** :  .*.. **   *.* ** . ** **.*. *  ****:*      ***** ****.** **.*  ****** ****** * ***  .**.:. *.** .*.**

human  -ATCCTG--TACACAGTTGTAACTGGCATG-TCTGAATCCAGTTTCAAACCAAGGTGCCT  -566   ------GGC  +254
mouse  AGACCGGGTATCTCTATTTTACCCGGACTGGCCAGCAAGCCCCAACAAGCCACC-TGTCT  -616   GCTGTTGGA  +261
        :** *  :*:* * *  :*  *  ** *** **  **:.:**: *** **:* ** ** **        **.

human  TTCAGTGAACACCCAAGCATCTTCCTCTCGGATTAACTGCTAGGCCAATACCAAAATCTC  -506
mouse  CCACCTCCCCAGCTCTGGAATTACAAATTAGGGCTATTGCCCACCCAAGGCATTTTTATG  -556
        . . .** *:. **  .* .  *::*. *:. .*. :* . *: .***. * .:::* .*
```

Figure 3.3 Sequence alignment of the *CERIG* promoter across human and mouse genomes and putative transcription factor-binding sites. Putative transcription factor-binding motifs and TATA-box are shown in rectangular box. The *asterisks* mark the fully conserved sequences across the species.

To determine whether there is direct binding of the AP-1 transcription factor to this site, an electrophoretic mobility shift assay (EMSA) was employed. Incubation of nuclear extracts from MDA-MB-231 cells, which express a high level of endogenous *CERIG*, with a biotinylated double-stranded oligonucleotide containing the consensus sequence for the AP-1 binding site derived from the *CERIG* promoter, produced two shifted bands (Shift-1 and Shift-2) (**Fig. 3.4B**). These shifted bands did not occur in the presence of additional non-biotinylated probes (cold probes). To further analyze these shifted bands, a probe encoding a mutated AP-1 consensus sequence was generated and incubated with the nuclear extracts of MDA-MB-231 cells followed by EMSA. Incubation of the nuclear extract with the biotinylated-mutant probe abolished the Shift-2 band but had no effect on the Shift-1 band, suggesting that the Shift-2 band represents a specific interaction between AP-1 and its corresponding binding sequence (**Fig. 3.4C lane 4**). An antibody supershift experiment was also performed by sequentially adding the anti-c-Jun (AP-1) antibody to the binding reaction in the nuclear extracts containing biotinylated AP-1 probe. Co-incubation of antibody against c-Jun with nuclear extracts and AP-1 probe diminished formation of the Shift-2 band and resulted in formation of a super shift band (**Fig. 3.4C, lane 6**). This is due to the formation of a ternary complex consisting of the AP-1 probe, the AP-1 transcription factor, and the anti-AP-1 antibody.
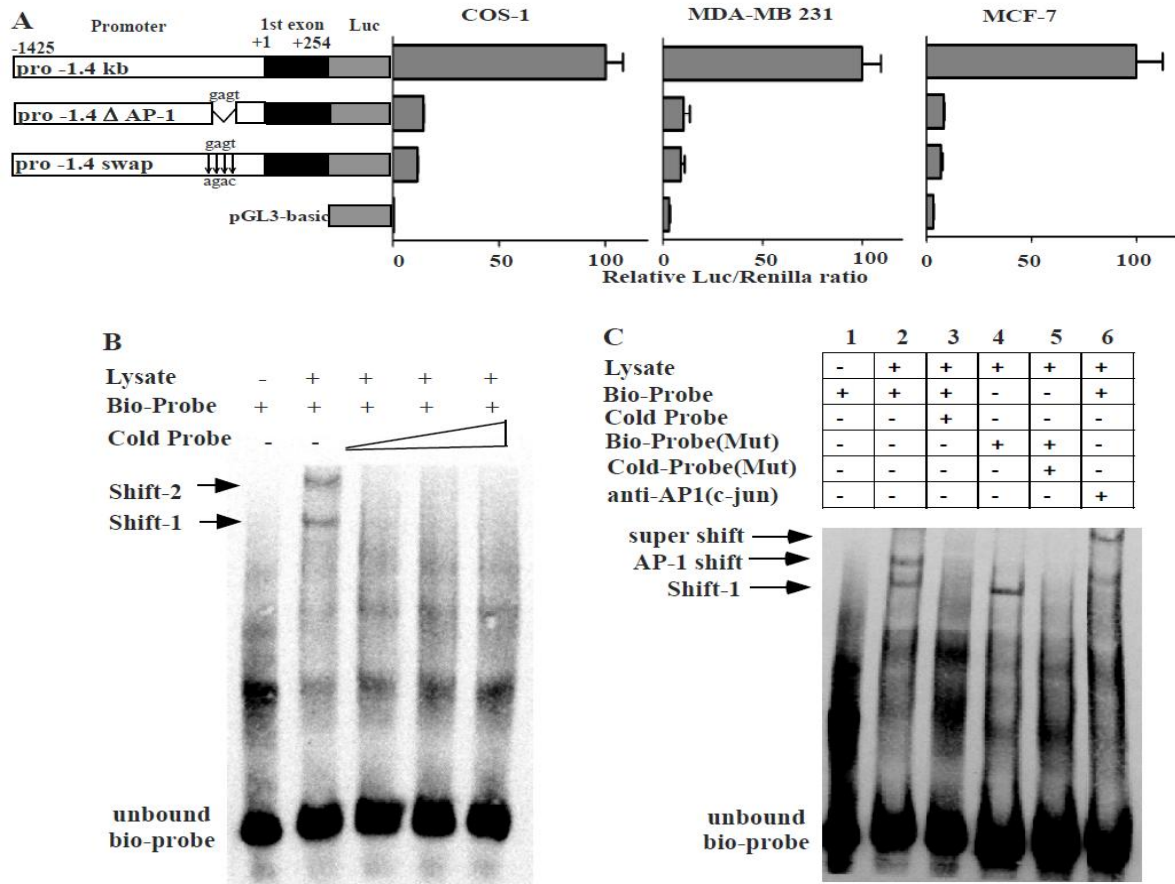
Figure 3.4) Mutation analyses of the AP-1 motifs and EMSA study of AP-1.A) The relative position of the putative transcription-binding motif is illustrated. A site-directed mutagenesis and deletion was carried out in the consensus motif of AP-1 (GAGT) in the pro-1.4 promoter construct. The mutations are shown with arrows. The promoter activities of the mutation and deletion were then determined relative to the activity of the wild type pro-1.4 promoter construct in MCF-7, MDA-MB-231 and COS-1 cells. **B)** Nuclear extracts was obtained from MDA-MB-231.A double-stranded 50-bp oligonucleotide that includes AP-1 binding site was generated as a probe. The labeled probe was incubated with the nuclear extracts either alone (*lane 2*) or in competition with 50-200 fold excess amounts of unlabeled probe (*lane 3-5*). Some of the specific DNA-protein complexes formed are indicated as Shift-1 and Shift-2(*lane 2*).**C)** The mutant probe failed to compete against the wild type probe for the specific binding (*lane 4*). To identify the DNA-binding proteins in a super-shift assay, antibody against c-Jun was added to the binding reactions and incubated for 20 min at room temperature before labeled probes were added (*lane 6*).

To further assess if endogenous AP-1 binds to the AP-1 consensus sequence within the *CERIG* promoter, a chromatin immunoprecipitation (ChIP) assay was utilized. Since the AP-1 consensus sequence is known to be recognized by a transcriptional complex composed of either a homodimer of c-Jun protein molecules or a heterodimer of c-Jun and c-Fos molecules, anti-c-Jun antibody was used to precipitate AP-1 followed by real time RT-PCR. Cross-linked chromatin was prepared from MDA-MB-231 cells. The *CERIG* promoter region containing the AP-1 binding site was precipitated using either the anti-c-Jun antibody or IgG control. Two primer sets: Set A spanning the AP-1 region from -125 to -30 and Set B spanning an unrelated region in the first intron from +900 to +1080 were designed for quantitative real time PCR analysis of either anti-AP-1 or anti-IgG antibody-precipitated chromatins. As shown in **Figure 3.5**, precipitation with anti-AP-1 antibody, but not IgG control, resulted in amplification of the region encompassing the AP-1 site. No PCR amplification of the immunoprecipitated chromatin was detected with primer Set B for the region that does not have the AP-1 consensus sequence, thus confirming the specificity of the results. Together, these data indicate that the *CERIG* promoter contains a functional AP-1 binding site that plays a critical role in transcriptional activation of the *CERIG* promoter.
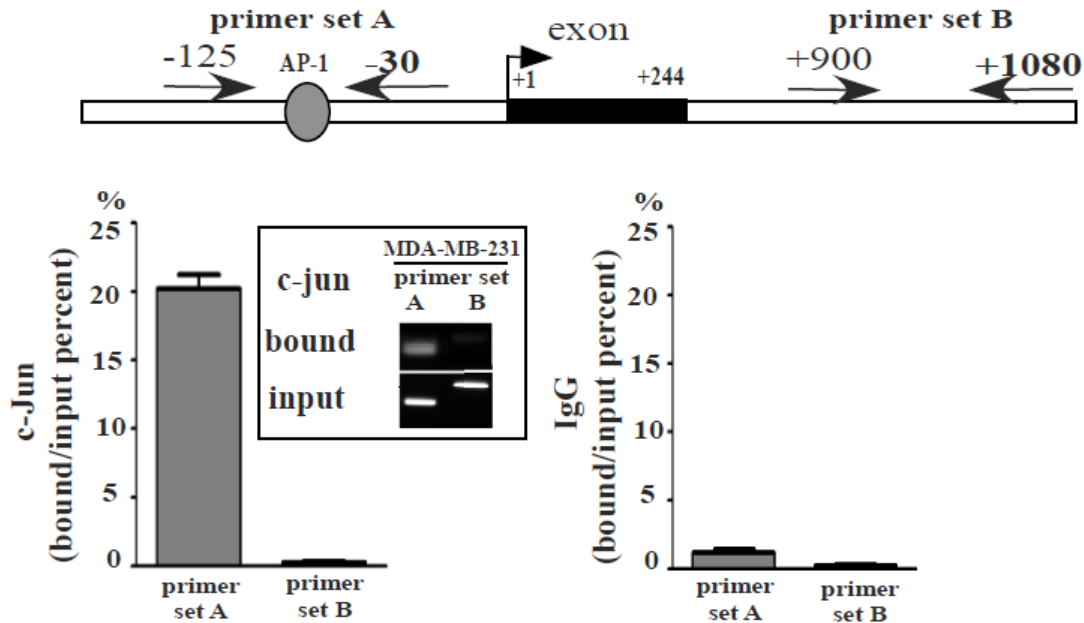
Figure 3.5) Chromatin immunoprecipitation analyses of AP-1 association with the *CERIG promoter* region. The association of endogenous AP-1 and the promoter sequence was examined in intact MDA-MB-231 cells by PCR amplification of the sequence between -125 and -30, following chromatin immunoprecipitation using specific c-Jun antibody. 20% of DNA was pull-down by AP-1 in its binding region. A weak association of AP-1 was detected in the unrelated intron region of *CERIG*. Normal rabbit IgG was used as a negative control. Results are calculated according to the bound/input ratio.

### 3.2.6 Involvement of NF-κB in transcriptional activity of the *CERIG* promoter

NF-κB and Twist-1 transcription factors have been demonstrated to play important roles in epithelial-to-mesenchymal transition (*EMT*), an early and critical step for cancer invasion and metastasis (Huber et al., 2004; Maier et al., 2010; Yang et al., 2010). NF-κB and Twist-1 were overexpressed in cells to examine whether the putative NF-κB and Twist-1 sites are required for *CERIG* transcription activity. COS-1 cells were transiently co-transfected with *CERIG* promoter pro-1.4 along with either NF-κB p65, a subunit of the NF-κB transcription complex, or Twist-1 cDNAs followed by a luciferase activity assay. In agreement with our observation derived from

the study with deletion mutants (**Fig. 3.2B**), overexpression of NF-κB p-65, but not Twist-1

cDNAs, significantly enhanced *CERIG* transcription activity as compared to vector control with

the *pro*-1.4 promoter construct (**Fig. 3.6A**). Since two lines of data (**Fig. 3.2B** and **Fig. 3.6A**)

suggested that Twist-1 was not involved in induction of *CERIG*, only NF-κB was further

examined to determine which putative site(s) is involved in *CERIG* promoter activity. To this

end, a site-direct mutagenesis approach was employed to substitute key consensus nucleotides

within the deduced binding sites based on predicted high Consensus Index score (Ci-value) as

indicated by MatInspector. The mutant plasmids were co-transfected into MDA-MB-231 cells

along with Renilla plasmid followed by a Dual-*Luciferase*® reporter assay. No significant

change in luciferase activity was observed upon mutation of the three NF-κB sites in the

proximal region of *CERIG* promoter relevant to +1 (data not shown). However, substitution of

the distal NF-κB binding site reduced the promoter activity by 75% compared to the full length

1.4 kb promoter (**Fig 3.6B**). In addition, overexpression of NF-κB p65 cDNA in cells expressing

NF-κB mut-IV failed to enhance *CERIG* promoter activity (Fig. **3.6A**), suggesting the fourth NF-

κB binding site may interact with NF-κB.

To assess whether there is a direct interaction between NF-κB and the NF-κB consensus

sequence, similar EMSA assay was done. When MDA-MB-231 cells treated with TNF-α, which

induces nuclear translocation of NF-κB (Chen & Goeddel, 2002), were lysed and protein extracts

incubated with a biotinylated-probe covering a 25 bp region spanning the fourth NF-κB site, a

single band was detected (**Fig. 3.6C**). The band disappeared when the mutated NF-κB sequence

or an excess of cold probe were used (**Fig. 3.6C**, **lane 6** and **4**, respectively). When an antibody

super shift assay was performed using an antibody against NF-κB p65, a tri-molecular super shift

complex was formed (**Fig. 3.6C**, **lane 5**), further supporting the hypothesis that NF-κB plays a direct role in *CERIG* activation.

To further confirm the association of NF-κB with its binding consensus sequence in the *CERIG* distal promoter region, I transfected HeLa cells with GFP or NF-κB p65 cDNAs and performed ChIP analysis in HeLa cells, which showed higher transient transfection efficiency compared to the MDA-MB-231. Cross-linked chromatin from the transfected HeLa cells was immunoprecipitated using anti-NF-κB p65 and anti-IgG (control) antibodies. Three primer pairs for *CERIG* were designed for real time PCR as illustrated in **Figure 3.6B**: Set A covering the fourth NF-κB binding site, Set B covering the first and second NF-κB binding sites, and Set C covering an unrelated region in the first intron region. As expected, a three-fold increase in the bound over input ratio was seen in NF-κB p65 transfected cells as compared to GFP control (**Fig. 3.6D**). No difference was noted using primer Set B and C or using normal IgG control. Taken together, these data confirm that only the fourth NF-κB consensus sequence within the distal regulatory region of the *CERIG* promoter is accessible to NF-κB
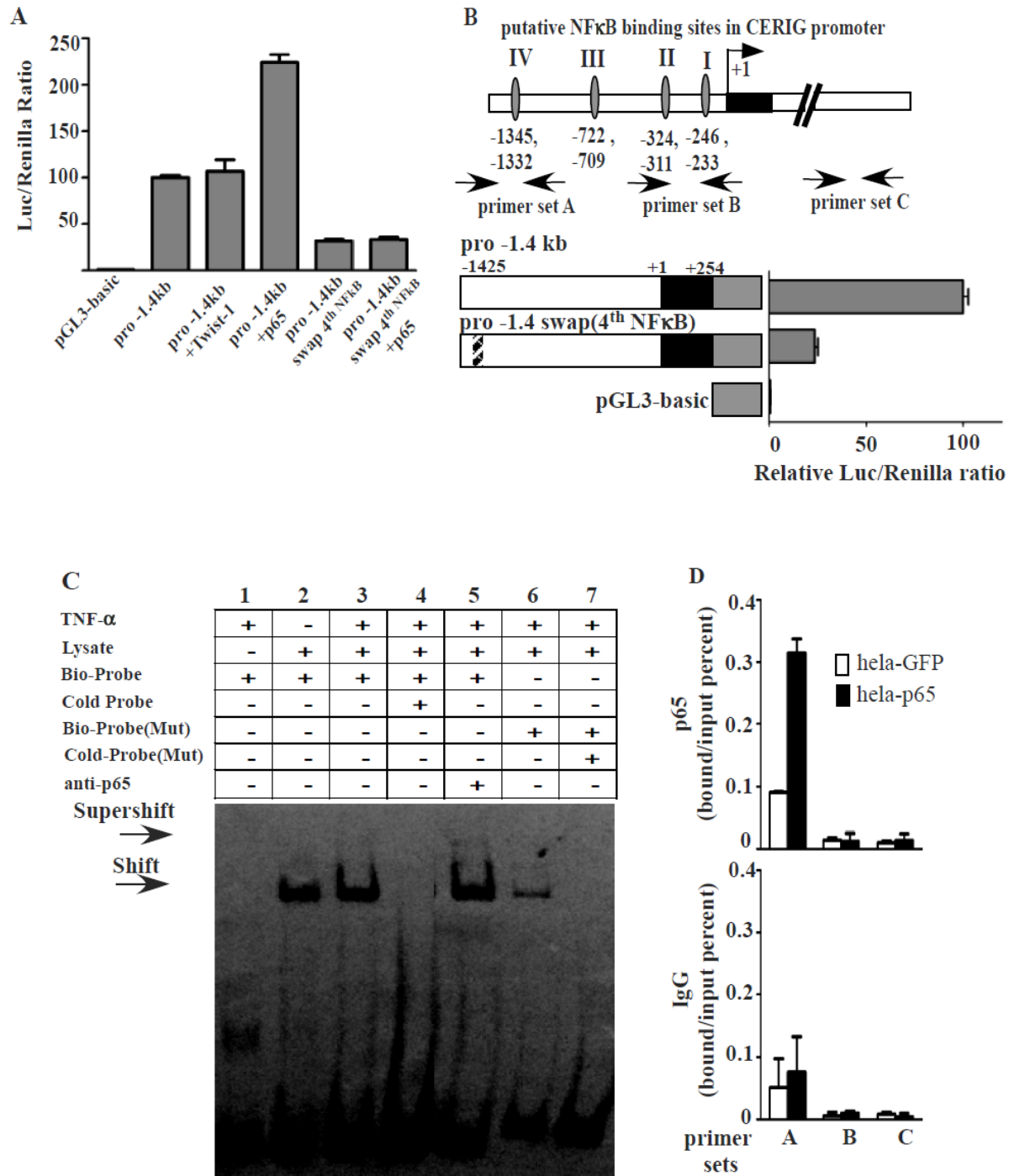
Figure 3.6) Characterization of NFκB as a regulatory element in distal part of promoter A) Twist-1 and NFκB (p65 subunit) cDNA were cotransfected with pro-1.4 kb promoter to see their effects on reporter gene activity. NFκB cDNA was also used with the pro-1.4 swap (4th NFκB) construct. B) The promoter activities of the mutation were then determined relative to the activity of

the wild type pro-1.4 promoter construct in MCF-7, MDA-MB-231 and COS-1 cells. Positions of primers used in the ChIP assay are shown on the constructs. Primer A amplifies the 4th NFκB binding site, primer B amplifies the 1st and 2nd NFκB binding sites, and primer C amplifies the unrelated region in the intron. C) Nuclear extracts was obtained from MDA-MB-231 treated with TNF-α. A double-stranded 50-bp oligonucleotide that includes 4th NFκB binding site was generated as a probe. The labeled probe was incubated with the nuclear extracts either alone (*lane 3*) or in competition with 100 fold excess amounts of unlabeled probe (*lane 4*). Specific DNA-protein complex formed is indicated as Shift (*lane 3*). To identify the DNA-binding proteins in a super-shift assay, antibody against p65 was added to the binding reactions and incubated for 20 min at room temperature before labeled probes were added (*lane 5*). The mutant probe decreased the binding capacity of NFκB indicating the specific interaction between DNA and protein (*lane 4*).D) ChIP assay identifies the interaction of NFκB (p65) in the fourth putative binding site only.

3.3 Discussion

Although the Human Genome Project is now completed, many obscured genes, such as *KIAA1199* or *CERIG,* remain to be characterized. Our recent efforts in characterizing the role of *CERIG* in cancer progression have led to the demonstration that *CERIG* plays a critical role in inducing cell migration and EMT, leading to cancer cell invasion and metastasis (manuscript in preparation). However, the function of *CERIG* in cancer progression has not been previously reported. Our functional characterization, in addition to the high expression in human cancers, strongly suggests that *CERIG* plays an important role in cancer dissemination.

To characterize the transcriptional regulation of *CERIG*, firstly TATA-box, the GC-box and the transcription start site within the proximal region of the promoter were identified by *in silico* analysis. Our results suggest that the *CERIG* promoter is a classical promoter with its relatively simple organization and tight regulation limited to a few structural elements such as

the canonical TATA- and GC-boxes, and the AP-1 and NF-κB *cis*-binding elements. Twenty four percent of human genes have a TATA-box in their promoter and these genes have been characterized with their high regulation by biotic and stress stimuli (Yang, Bolotin, Jiang, Sladek, & Martinez, 2007). Moreover, TATA-dependent promoters have been related to tissue or context specific expression, which is in concordance with the abundant expression of *CERIG* in the CNS (Sandelin et al., 2007). However, the tissue distribution mechanism of *CERIG* remains to be fully understood. Another property of TATA-box containing genes is the involvement of several transcription start sites (TSS) (Shiraki et al., 2003). In order to reveal the number of TSS in the *CERIG* promoter, 5' primer extension analysis was performed. Observation of a single band implied the presence of a single TSS for the *CERIG* promoter, but there are limitations to this technique when it comes to the separation of cDNAs that differ by only a few nucleotides. Therefore, this finding was compared with the findings from the internet-based Fantom3 database, called *CAGE* (*C*ap *A*nalysis of *G*ene *E*xpression) *Analysis Viewer* that supplies information about the structure and type of any given promoter (Shiraki et al., 2003). Although it is very common to see several TSSs for TATA-containing promoters, only one dominant TSS was determined by CAGE analysis with two neighboring guanines (G) upstream of the TSS predicted as well but with much less chance of being an alternative transcription start site. These results supported the classification of *CERIG* as a SP (single dominant peak) of promoter according to the Carninci promoter classification method in which four types of promoters were identified (Carninci et al., 2006). Hence, it is obvious that there is no other major TSS around with a 500 bp of the +1 site. In addition, the transcription start site of *KIAA1199* is conserved between human and mouse.

Another common structure in the *CERIG* 5'-flanking region is a GC-box that is located in the -248 /-243 region of the *CERIG* promoter. Since effective GC-boxes are generally found 70-80 bp upstream of the TSS (Charron, DeCerbo, & Wright, 2003), it was hypothesized that the GC-box of *CERIG* may not be required for regulation of *CERIG* promoter activity. This hypothesis was supported by our serial deletion mutant study. Deletion of the region in pro -0.583 containing the GC-box did not change the promoter activity (**Fig 3.2**). Methylation of cytosine residues is another important feature of GC-boxes in the regulation of genes (Miyajima, Furihata, & Chiba, 2009; Sunahori, Juang, & Tsokos, 2009), but no methylation of cytosines in the GC-box of the *CERIG* promoter was observed. All these findings suggest that there is little or no function of the GC-box in *CERIG* regulation. Further functional analysis could be done by using Sp-1, which binds to GC-boxes (Kang & Chen, 2009; Muckenfuss et al., 2007), to further understand the possible importance of the GC-box within the *CERIG* promoter. Similarly, it was found that Twist-1, a transcription factor involved in EMT conversion in epithelial cancer cells (Martin & Cano, 2010), was not required for regulation of *CERIG*, although putative *cis*-elements were present. Our results suggest that the *CERIG* promoter is a classical promoter with its relatively simple organization and tight regulation limited to a few structural elements such as the canonical TATA- and GC-boxes, and the AP-1 and NF-κB *cis*-binding elements.

After the identification of the TSS and several basic transcription initiator motifs in the *CERIG* promoter, identification of important DNA fragments of the promoter i.e *cis*-acting sites and the *trans*-acting factors associated with these regions are taken into consideration. Through a series of deletion studies in epithelial breast cancer cells, three important DNA fragments were revealed for *CERIG* promoter activity; 1) an inhibitory region between -2341 bp and -1425 bp, 2) a proximal activator region between -125 bp and +27 bp and 3) a distal activator region between

-1425 bp and -1135 bp. The strongest promoter activity with the 1.6 kb promoter containing the

region between -1425 bp and + 254 bp was obtained; therefore the following discussions of

promoter activity are based on this promoter region. First, the size of the promoter fragment was

extended from 1.6 to 2.5 and 3.5 kb including the first exon and a significant reduction occurred

in the promoter activity, indicating the presence of negative regulatory elements in the region of

upstream from -1425bp. This promoter activity was decreased by more than 50% when 2.3 kb

CERIG 5'-flanking regions were examined in COS-1 cells as well as in MCF-7 and MDA-MB-

231 breast cancer cells, suggesting the presence of negative regulatory *cis*-acting elements in the

upstream region between -1425 and -2341. Due to the difficulty in determining mechanisms that

regulate gene expression, computational biologists are devoting much of their effort to predicting

the binding sites of transcription factors within a given gene (Tompa et al., 2005). Taking

advantage of several bioinformatics tools such as Genomatix, Alibaba and Transfect, several

putative transcription factor bind sites within both the inhibitory and activator regions of *CERIG*

were identified. Interestingly, the inhibitory domain between -2341 bp and -1425 bp appeared to

contain multiple negative glucocorticoid response elements (nGRE) in the -2204/-2190,-2189/-

2175 and -1780/-1766 regions. The repressive effect of nGRE on promoter activity has been

known for many years (Sakai et al., 1988). In addition, glucocorticoids, which bind to nGRE,

have been used in the treatment of some cancers, including breast, by interfering with the

abnormal mechanism of cancer cells (Kassi & Moutsatsou, 2011). When the effects of this

inhibitory region on promoter activity in two different breast cancer cells were compared, the

negative regulatory region with the GRE binding sites decreased the promoter activity of the

reporter gene to a higher degree in ER (estrogen receptor) positive MCF-7 cells as compared to

ER-negative MDA-MB-231 cells. In addition, a positive effect was documented for the use of

59

glucocorticoids as a treatment for ER(+) breast cancer patients (Pan, Kocherginsky, & Conzen, 2011). These results suggest a potential role for the nGRE *cis*-element and glucocorticoid in *CERIG* regulation. Since there is no further characterization of this distal inhibitory region, possible roles of the nGRE binding site for *CERIG* activity remains to be defined.

Second, the fragment between -125 and +27 was identified as a basal promoter as deletion of this fragment reduced the promoter activity drastically in all cells tested. Our study demonstrated a pivotal role of AP-1 in controlling *CERIG* expression since the activity of the pro-1.4 kb *CERIG* promoter was repressed by more than 70% by either deleting or swapping the AP-1 consensus sequence. A direct association of AP-1 with the putative binding site was further confirmed by performing various experiments, including EMSA, a supershift assay, and a ChIP assay. Transcriptional regulation of *CERIG* by the AP-1 complex is of great interest, because collagenase, stromelysin, and metalloproteinases which all play a role in tissue invasion and metastasis, have also been documented as potential target genes of AP-1 during carcinogenesis (Leaner et al., 2009). For example, interaction of the AP-1 complex with the promoters of MMP-2 and MMP-9 has been shown to play a very critical role in tumor invasion (Hasegawa, Senga, Ito, Iwamoto, & Hamaguchi, 2009). Besides the role of AP-1 in cancer dissemination, this complex has been reported to play a role in other biological functions such as cellular proliferation, differentiation and apoptosis, depending on the composition of AP-1 complex. One of the AP-1 constituents, called c-Jun, has been reported to trigger AP-1 mediated cancer progression (Hess et al., 2004). In this respect, taking into consideration the role of c-Jun in cancer progression, c-Jun was further investigated to show the interaction between AP-1 and the *CERIG* promoter. Confirmation of the link between the AP-1 complex and *CERIG* might reveal a novel strategy for therapeutic approaches that focus on inhibition of c-Jun in breast cancer

patients expressing high levels of *CERIG*. There are two shifted bands in EMSA and the one has been shown to be interacted with AP-1 caused the upper shift band or shift 2. Since the 50-bp oligonucleotide probe contains the TATA-box sequence that has been shown to bind to TATA-box binding protein (TBP) (Kornberg, 2007), it is assumed that the Shift-1 band may result from the binding between TBP and the probe based on the fact that the molecular weight of TBP is less than Fos/Jun. This observation raises the possibility that the interaction of the probe with TBP caused the shift-1, or the lower shift, in the EMSA study.

This project was started by finding an increase in the *CERIG* transcription after mitogen treatment, called ConA. Interestingly, Kovacs identified an increase in the DNA-binding of AP-1 to the KA receptor, which is non-NMDA ionotropic receptors, after the treatment of ConA. They decreased the concentration of ConA to 50 μg/ml, and they still found obvious effect of AP-1 binding on DNA (Kovacs, Cebers, & Liljequist, 2000). Moreover, proliferation of rat thymocytes and Jurkat T-cells has been linked to the enhancement of AP-1 DNA binding which occurred after ConA treatment (Kvanta, Kontny, Jondal, Okret, & Fredholm, 1992; Sikora, Grassilli, Bellesia, Troiano, & Franceschi, 1993). Due to the characterization of AP-1 in *CERIG* regulation in our study, upregulation of *CERIG* by ConA might be linked to the ConA-AP1 axis. However, receptor system and signaling cascade taking a role in the ConA induced AP-1 activation needs to be determined.

The third important region in the *CERIG* promoter was identified in the distal part of the promoter. Although it is not very common, there is some evidence for the localization of responsive elements in the distal part of promoters (Jonk, Itoh, Heldin, ten Dijke, & Kruijer, 1998; Ponticos, Harvey, Ikeda, Abraham, & Bou-Gharios, 2009). In our study, four putative NFκB binding sites ranging from the proximal to the distal part of the promoter were identified.

Site-directed mutagenesis identified the distal site, between -1345 and -1333, as the most critical binding site because substitution in its region induced the greatest decrease of promoter activity. Moreover, alignment of the human and mouse genome depicted the fourth NFκB binding site as the most conserved sequence among the four sites. Association of NFκB with the distal part of the promoter was further confirmed by EMSA and ChIP analysis. In both assays, antibody against p65, which is the major component of NFκB, was used to show the specific interaction between the DNA and protein. Identification of the NFκB site as an activator element in the *CERIG* promoter is significant because NFκB is essential for epithelial cancer EMT and metastasis (Escarcega, Fuentes-Alexandro, Garcia-Carrasco, Gatica, & Zamora, 2007; Garg & Aggarwal, 2002). Interestingly, in several of these cases, both AP-1 and NFκB have been shown to play a role in tumorigenesis together such as in colon cancer and breast cancer (Giancotti, 2006; Vaiopoulos, Papachroni, & Papavassiliou, 2010). Therefore, revealing the vital role of both AP-1 and NFκB in *CERIG* transcription further highlights the importance of *CERIG* in cancer progression and strengthens the need to expand our understanding of the relationships between these molecules and *CERIG* upregulation in order to help us develop novel cancer preventive methods.

3.4 Materials and Methods

Reagent:

Oligo primers were purchased from Operon (Huntsville, AL). pGL3-basic and pRL-Luciferase plasmids were purchased from Promega. Human p65 and c-Jun antibodies were purchased from Millipore (Massachusetts, MA).

Identification of promoter

Genomatix, The Markov Chain Promoter Prediction Server from Duke University, and Transfect were used as a statistical tool for the prediction of transcription start sites and putative promoter and promoter binding elements in eukaryotic DNA.

Non-radioactive 5' extension study

mRNA was isolated from MDA-MB-231 by using the Qiagen RNA Isolation kit and 5 µg RNA was converted into cDNA by using specific reverse primer, rev: 5'bio-GCCCTCTTACCTCTGGGTCT 3'. SuperscriptII reverse transcriptase was used in reverse transcription. After cDNA production, RNA was degraded with 2 µl of RNaseA (0.5 mg/ml) for overnight at 65 °C. cDNA was precipitated by using cold Ethanol DNA precipitation method. Purified cDNA was loaded in 6 % polyacrylamide gel in TBE and run for 1.5 hours at voltage 100. Then, gel is transferred to the nylon membrane for 1 hour in TBE buffer. DNA is cross-linked to the membrane by using UV-exposure. Biotin-labeled DNA was detected by chemiluminescence kit of Thermo Scientific.

Construction of plasmids

pGL3-basic and pRL-Luciferase were purchased from Promega. Various *CERIG* promoters were cloned from BAC clone 96012 (Invitrogen, Clone ID: 2215F6) by using the PCR approach. The list of primers was given in **Table 3.1**. KpnI (GGTACC) restriction site was added to the 5' of forward primer. In the reverse primer, BglII restriction site (AGATCT) was used. PCR reaction was done in the presence of 0.5M GC-rich resolution buffer (Roche) with following conditions; 95 ^{o}C for 5 min, (94 ^{o}C for 30 s, 56 ^{o}C for 30s and 72 ^{o}C for 30s) for 30 times, 72 ^{o}C for 7 min.

Primer number	Construct name	Primer Sequence
2814	Pro-3.3 kb	5'ATGGTACCCATGGAAAAATGCTTGGCTA
2813	Pro-2.3 kb	5'ATGGTACCCAGGCCAATCTACTGCACAA
2676	Pro-1.4 kb	5'ATGGTACCCAAAGAAGGGCTGAGAATCC
2677	Pro-1.135	5'ATGGTACCTAGGAGGCCTCTGCTAACCA
2678	Pro-0.913	5'ATGGTACCATGCCCTAGGGATGTGTTGT
2679	Pro-0.583	5'ATGGTACCTTCAAACCAAGGTGCCTTTC
2680	Pro-0.217	5'ATGGTACCGCTCAGGTCTGCGTCTCC
2681	Pro-0.125	5'ATGGTACCGCGTGGAGGGAAGTTTCAT

2682	Pro+0.027	5'ATGGTACCGCGGTGCTATCGGACAGAG
2683	Common reverse primer	5'ATAGATCTGCCCTCTTACCTCTGGGTCT

Table 3.1 List of primers used for the construction of the different size of

***CERIG* promoters.**

For the deletion and mutation studies, Quick-change site-directed mutagenesis kit from Stratagene was purchased. Pro-1.4 kb promoter containing plasmid was used as a template and all mutation and deletion were done on this template by using primers listed in **Table 3.2**. Reactions were performed according to the instructions of the company.

Primer Number	Construct name	Direction	Primer Sequence
2761	**Pro-1.4 kb ΔAP-1**	For	5'GGGCTCCTCGGGGGAGTAAAAGTGGCTATA
2762	**Pro-1.4 kb ΔAP-1**	Rev	5'TATAGCCACTTTTACTCCCCCGAGGAGCCC
2759	**Pro-1.4 kb swap AP-1**	For	5'GTCAGGGGCTCCTCGGGGGAGAGACCAGTAAAAGTGGCTATAAAAGC
2760	**Pro-1.4 kb swap AP-1**	Rev	5'GCTTTTATAGCCACTTTTACTGGTCTCTCCCCCGAGGAGCCCCTGAC
2748	**Pro-1.4 kb swap 1st NFκB**	For	5'CACCTGGCTGCTGCGGGCAAAGCTTCACGCACCCGGGCTCCGCGG

2749	**Pro-1.4 kb swap 1st NFκB**	Rev	5'CCGCGGAGCCCGGGTGCGTGAAGCTTTGCCCGCAGCAGC CAGGTG
2746	**Pro-1.4 kb swap 2nd NFκB**	For	5'CTGGCTCCTTGGAGGCCCGAAAGCTCATTCTCCGCGGTC CGCCGCGG
2747	**Pro-1.4 kb swap 2nd NFκB**	Rev	5'CCGCGGCGGACCGCGGAGAATGAGCTTTCGGGCCTCCAA GGAGCCAG
2809	**Pro-1.4 kb swap 3rd NFκB**	For	5'ACCAAGCACCTATTATGTGCTTAGAACTTAAACCCCTTT GATCTCATTTAAGCCTCACCA
2810	**Pro-1.4 kb swap 3rd NFκB**	Rev	5'TGGTGAGGCTTAAATGAGATCAAAGGGGTTTAAGTTCTA AGCACATAATAGGTGCTTGGT
2805	**Pro-1.4 kb swap 4th NFκB**	For	5'CTGGAAGAGGCCTAGAGGTGGATTGCGAAGCCCTCCAGT TTAACCAGGTGACAAAGGCTT
2806	**Pro-1.4 kb swap 4th NFκB**	Rev	5'AAGCCTTTGTCACCTGGTTAAACTGGAGGGCTTCGCAAT CCACCTCTAGGCCTCTTCCAG

Table 3.2 List of primers used for mutation and deletion of *CERIG* promoter

Dual-Luciferase Assay

Target cells were plated at 50% confluent a day before the transfection. Next day, NaCl transfection method was used with 1.5 µg luciferase plasmid in 3 cm culture plate or appropriate amount of DNA for 96-well plate. 1.5 ul of Renilla plasmid (1 ng/ µl) was used in every transfection reaction as an internal control. Following day, media was changed with the fresh one. After 48 hours of transfection, Luciferase and Renilla activity were measured by using the Dual-Glo Luciferase Assay System (Promega) according to the company's instructions.

Preparation of Nuclear Lysate for EMSA

The basic procedure was described earlier (Wadman et al., 1997). In brief, 5*10e6 cells were washed with cold PBS twice and resuspended in 500 µl of buffer A containing 10 mM HEPES at pH:7.9,1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT plus 1:100 protease inhibitors. NP-40 or Triton x-100 was added to a final concentration of 0.5% and cells were vortexed for 10 seconds. Lysates resuspended in 150 µl of Buffer B containing 20 mM HEPES at pH:7.9,1.5 mM $MgCl_2$, 420 mM NaCl, 0.2 mM EDTA, 25% v/v Glycerol and 1:100 protease inhibitors. Nuclear extracts were recovered from supernatant.

Electrophoretic Mobility Shift Assay (EMSA)

Light Shift Chemiluminescent EMSA kit and Biotin 3' end DNA Labeling Kit were purchased from Thermo Scientific. In brief, following procedure was performed for EMSA assay. Oligos that are listed in **Table 3.3** were labeled at their 3' termini by using the Biotin 3' end DNA Labeling Kit according to the manual. Oligos were then annealed by incubation at 90 °C for 3 min followed by one hour incubation at 37 °C. MDA-MB-231 nuclear lysate was mixed with the appropriate biotinylated oligos. For the cold probe or antibody containing reaction, cold

probe or antibody was added to the reaction 10 min before the addition of biotinylated probe.

Binding reaction was performed for 30 min and loaded into 6% polyacrylamide gel in 0.5X TBE-buffer for two hours at voltage 100. Then, gel was transferred to the nylon membrane for one hour in 0.5X TBE-buffer. DNA was cross-linked to the membrane by using UV-light cross-linker via auto exposure twice. Biotin-labeled DNA on the nylon membrane was detected by Chemiluminescence EMSA Kit (Thermo Scientific, MA).

Oligo number	Probe Name	Oligo sequence
2788	AP-1 probe	5'CGTCAGGGGCTCCTCGGGGGAGGAGTCAGTAAAAGTGGCTATAAAAGCGG
2789	AP-1 probe	5' CCGCTTTTATAGCCACTTTTACTGACTCCTCCCCCGAGGAGCCCCTGACG
2759	AP-1 swap probe	5'-GTCAGGGGCTCCTCGGGGGAGAGACCAGTAAAAGTGGCTATAAAAGC
2760	AP-1 swap probe	5'-GCTTTTATAGCCACTTTTACTGGTCTCTCCCCCGAGGAGCCCCTGAC
2817	NFκB probe	5'AGAGGCCTAGAGGTGGATTGCAGGACTTCCCAGTTTAACCAGGTGACAAA
2818	NFκB probe	TTTGTCACCTGGTTAAACTGGGAAGTCCTGCAATCCACCTCTAGGCCTCT
2805	NFκB swap probe	5'CTGGAAGAGGCCTAGAGGTGGATTGCGAAGCCCTCCAGTTTAACCAGGTGACAAAGGCTT
2806	NFκB swap probe	5'AAGCCTTTGTCACCTGGTTAAACTGGAGGGCTTCGCAATCCACCTCTAGGCCTCTTCCAG

Table 3.3 List of oligos used in the EMSA study

Chromatin immunoprecipitation (ChIP)

MDA-MB-231 cells were cultured in two 15-cm culture plate. Cells in confluent plates were crosslinked with 0.75% (v/v) final concentration of formaldehyde at room temperature for 10 minutes by slow rotation. 125 mM final concentration of glycine was used for 5 minutes at room temperature to stop the crosslinking reaction between protein and DNA. Cells were washed twice with cold PBS and collected into cold PBS containing protease inhibitors. After centrifugation at 1000g for 5 min, supernatant was dissolved in 1 ml FA lysis buffer. Lysates were sonicated to get the average DNA size around 500 bp. 50 µl of sample was taken as an input DNA, and then DNA was purified with proteinase K digestion method and measured by nanodrop. Lysate containing 25 µg sonicated DNA was immunoprecipitated with 2.5 µg antibody in RIPA buffer containing 25 µl of protein A beads at 4 ^{o}C overnight. Beads were then washed four times with wash buffer containing 0.1 % SDS, 1 % Triton X-100, and 2 mM EDTA (pH8), 150 mM NaCl and 20 mM Tris-HCl at pH8. Beads were washed with "final wash buffer" which has 500 mM NaCl rather 200 mM NaCl concentration in the wash buffer. Finally, DNA was eluted in the elution buffer containing 1% SDS and 100 mM $NaHCO_3$ at 30 ^{o}C for 15 minutes. Eluted DNA and input DNA were used in the real-time PCR. For HeLa cell, cells were transfected with 10 µg DNA of GFP or p65 (NFκB) DNA by using NaCl transient transfection method. After two days, cells were treated and collected as described for MDA-MB-231.

Chapter 4: Epigenetic Regulation of *CERIG*

4.1 Introduction

Epigenetics consists of two separate words, "epi" and "genetics". "Epi" means "above" or "in addition" according to Greek language, so epigenetics is known as "nongenetic cellular memory" or the information on top of the traditional molecular basis for inheritance. This term was coined by C.H. Waddington (1942) in his studies on development. Since then, many definitions have been made for epigenetics and some mechanisms of epigenetics are still under debate today. According to the recent identification of epigenetics in a review in Science 2010, epigenetics mechanisms are divided into two major groups, called trans-epigenetic signals and cis-epigenetic signals (Bonasio, Tu, & Reinberg, 2010). Trans-epigenetic mechanisms are primarily established by feedback loops and transcription factors (TF) as in the case of the self-propagation of a gene whose product works as a transcription factor to activate itself. Some small RNAs and noncoding (nc) RNAs (miRNA) are also considered examples of trans-epigenetic mechanisms (Moazed, 2009). Trans-epigenetic states are mainly reported in single-cell eukaryotes and prokaryotes. Contrary to trans-signals, cis-epigenetic mechanisms are described as direct interactions of its players with the chromosome resulting in heritable modifications upon cell division. DNA methylation by various DNA methyltransferases and some covalent modifications of histones, which constitute the protein part of chromatin, are considered the key players of cis-epigenetic mechanisms.

There are three criteria to be considered as an epigenetic signals; (i) propagation mechanism; that is, replication of molecular signature after cell division; (ii) self-sustaining transmission to the progeny; and (iii) role in transcriptional regulation (Bonasio et al., 2010). In this respect, DNA methylation meets all these requirements. However, for histone

posttranslational modifications (PTM), each alteration should be considered individually because some of them are not considered to change the transcription level and are not transmitted to the progeny.

Figure 4.1 represents the how DNA methylation and histone modifications occur on the chromosome.
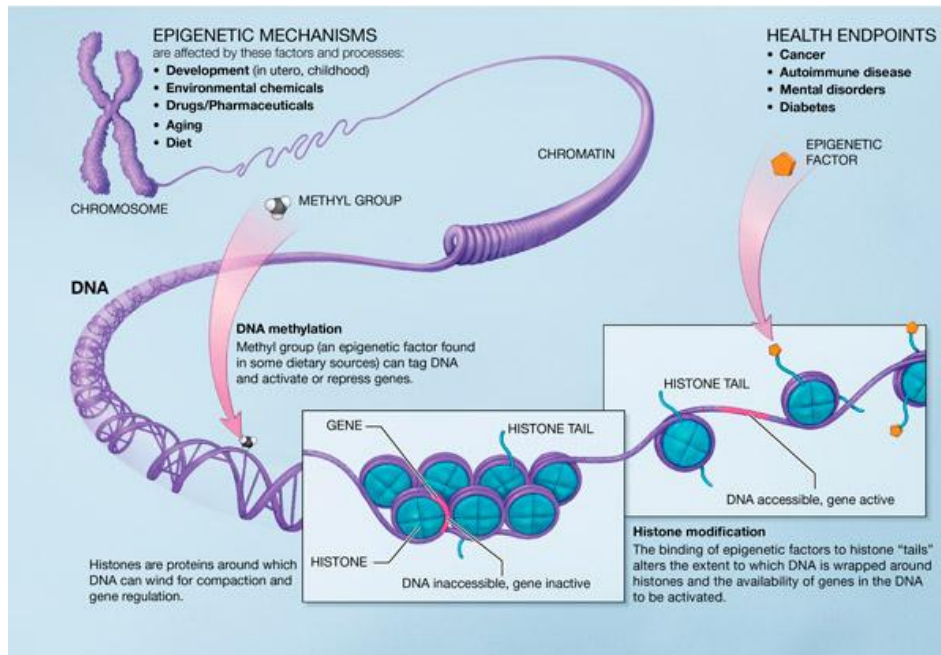

Figure 4.1 Two main mechanisms of epigenetic regulation are shown with their target regions on the chromatin. **Source:** http://nihroadmap.nih.gov/EPIGENOMICS/images/epigeneticmechanisms.jpg

DNA methylation: DNA methylation usually occurs in the 5' position of the cytosine nucleotide of a CG pair. This modification is performed by an enzyme called DNA methyltransferase (DNMTs). There are three different DNMTs, called DNMT1, 3a and 3b in humans (**Fig4.2**).
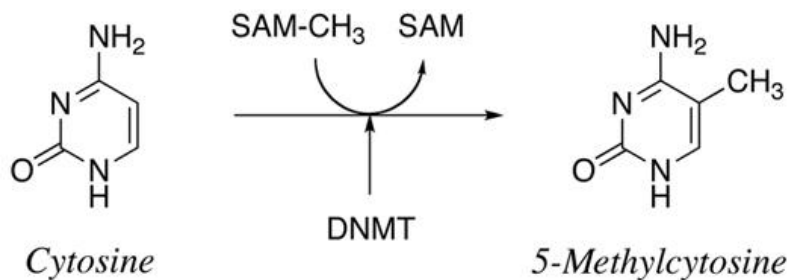
Figure 4.2 Synthesis of 5-methylctosine from cytosine by methyltransferase (DNMT). DNMT performs the transfer of a methyl group (CH$_3$) from S-adenosylmethionine (SAM) to the 5-carbon position of cytosine. **Source:** http://www-medchem.ch.cam.ac.uk/

In normal cells, most of the DNA methylation takes place in repetitive genomic regions, such as LINES (long interspersed transposable elements) and SINES (short interspersed transposable elements) to maintain genomic integrity (Robertson, 2005). Besides random and diverse methylation of cytosines in these regions, DNA methylation is concentrated in CpG islands (CGI), which are described with the following formula; minimum 200-bp stretch of DNA with a minimum C+G content of 50% and an Obs$_{CpG}$/Exp$_{CpG}$ in excess of 0.6 (Gardiner-Garden & Frommer, 1987). Based on this criterion, there are approximately 29,000 CpG islands (CGI) in the human genome. Besides the localization within promoters of genes, they were shown to be in the intergenic and intragenic (intron) regions (Saxonov, Berg, & Brutlag, 2006). Half of all human promoters have been reported to have CGIs, but these CGIs within promoters are generally not methylated. Methylated CpG islands in promoters and intragenic regions can inhibit transcription directly by blocking the access of specific transcription factors to their binding sites, and indirectly by recruiting methyl-CpG-binding domain (MBD) proteins that further recruit histone modifying and chromatin-remodeling complexes which condense the chromosome resulting in repression of the gene (Portela & Esteller, 2010). In addition to the

basic role of DNA methylation in the development of organisms and cellular differentiation via altering gene expression profiles, aberrant DNA methylation patterns have also been linked to nearly all types of cancer (Jaenisch & Bird, 2003). The role of methylation in cancer progression is primarily considered as a molecular instrument for hypermethylation of promoters in order to silence tumor suppressor genes, and major interest is given in that direction (Esteller, 2002; Hayslip & Montero, 2006). On the other hand, hypomethylation of oncogenes or growth-related genes is another mechanism related with cancer progression. Almost 30 years ago, Feinberg and Vogelstein demonstrated for the first time that hypomethylation in the promoters of oncogenes, such as c-Ha-ras and c-Ki-ras, induced the formation of tumors in colon and lung tissues (Feinberg & Vogelstein, 1983).

Histone Modifications: Histone proteins play an important role for packaging of DNA into the condensed structure known as nucleosomes in which ~ 146 bp of DNA wrap around the octamer structure of histone proteins. Besides the function of histones in the spatial organization of genetic information, their tails state the "histone code" via various posttranslational modifications (PTMs), ultimately playing a role in gene regulation. These covalent modifications are acetylation, methylation, phosphorylation, ubiquitination, sumoylation, proline isomerization, and ADP ribosylation (Cohen, Poreba, Kamieniarz, & Schneider, 2011; Kouzarides, 2007). Histone acetyltransferases (HATs) & deacetylases (HDACs), methyltransferases (HMTs) & demethylases (HDMs), kinases, phosphatases, ubiquitin ligases & deubiquitinases, SUMO ligases & proteases add or remove the above mentioned modifications on the histone tails (Bannister & Kouzarides, 2011). However, only acetylation and methylation on particular residues of histone tails are considered as an example of a *cis*-epigenetic mechanism because of

73

their role in the conformational change of chromatin to either euchromatin or heterochromatin formation. Euchromatin, which is also known as "open state", favors transcription and is characterized by acetylation of lysine residues at histone tails and trimethylation of lysines on following positions H3K4, H3K36 and H3K79. On the other hand, heterochromatin, which is known as "closed state", inhibits transcription and is defined by deacetylation of lysines and tri-methylation of lysines on H3K9, H3K27 and H4K20 (Karlic, Chung, Lasserre, Vlahovicek, & Vingron, 2010; Li, Carey, & Workman, 2007). Enzymes modifying histone tails have also been linked to carcinogenesis by changing the open/closed states of chromatin (Esteller, 2007). In the aspect of epigenetic regulation, two modifications have been extensively studied and are characterized as an activator marker and repressive marker, H3K4me3 and H3K27me3, respectively (Martin & Zhang, 2005). While lysine methyltransferases bring methyl groups to lysines, for example SET1 and MLL1 have been reported to add methyl groups to H3K4 as an activator signal (Wang et al., 2009) and EZH2 incorporates tri-methyl groups to H3K27 as a repressive modification (Ezhkova et al., 2011), lysine demethylases remove methyl groups from lysine residues on the histone tails, for example RBP2(JARID1A) was reported to remove a tri-methyl group from H3K4me3 which works as a repressive mechanism (Christensen et al., 2007; Klose et al., 2007) , and JMJD3 removes the methyl groups from H3K27me3 as a activator mechanism (Agger et al., 2007). Enzymes playing a role in methyl group modifications within the H3 tail are summarized below (**Fig 4.3**).
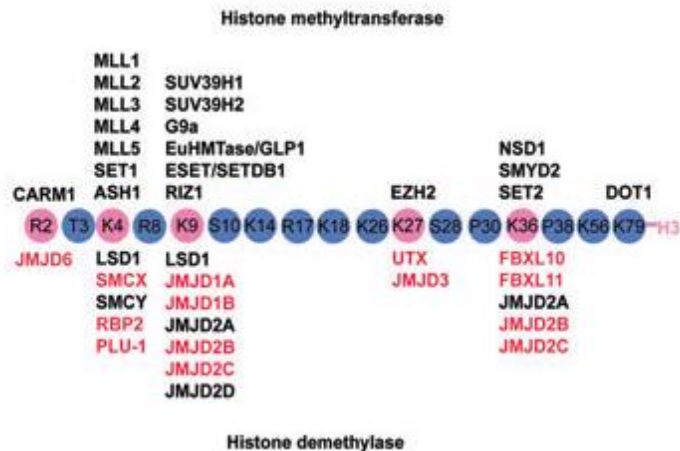
Figure 4.3) List of histone methyltransferases and histone demethylases for H3 tail. While enzymes above the lysine residues transfer methyl groups, enzymes listed below the lysine residues removes methyl groups. Activities of some enzymes labeled with pink color are under the control of hypoxia (Yang et al., 2009).

miRNA: miRNAs are small non coding RNAs of ~22 nucleotides that inhibit the translation of proteins by helping the degradation of mRNA (Lu et al., 2005). miRNAs not only play a role in normal cell physiology, but also play a role in the progression of cancer due to their mis-expression. Today, 1000 miRNAs have been predicted *in-silico*, and 60% of mRNA are potential targets of miRNA (Friedman, Farh, Burge, & Bartel, 2009). miRNAs can be considered as key players in epigenetic remodeling and as targets of epigenetic regulation at the same time (Wiklund, Kjems, & Clark, 2010). However, understanding the complete interaction between these two new expanding fields needs to be discovered.

As a summary, epigenetics is considered as a mechanism for changing gene expression without altering the DNA sequence. Therefore, DNA methylation, histone modifications, and miRNAs are accepted as the major players in the epigenetic remodeling that is heritable, reversible and transmittable. In addition to its role in normal physiology, there is an increasing interest to discover the link between epigenetic mechanisms and disease formation, particularly in regards to cancer progression.

75

4.2 Results

4.2.1 DNA methylation

4.2.1.1 Identification of a CpG island in the *CERIG* regulatory region

DNA methylation affects the transcription level of a gene in two independent ways. First, the methylation status of a promoter determines direct accessibility of transcription factors to their corresponding consensus sequences in a promoter area. Second, methylation of CpGs in the promoter or intragenic region might change the status of chromatin. Since *CERIG* has been found to be highly expressed in human cancers as compared to normal tissues, methylation status of the *CERIG* promoter was examined to hypothesize the DNA methylation as a possible mechanism controlling its expression in human cancer tissues and in aggressive cancer cell lines. To determine whether methylation occurs in the *CERIG* regulatory region, the presence of predicted CpG islands was searched in the portal region surrounding the transcription start site. Employing a CpG island search program (UCSC Genome Browser), a potential 1.9 kb-long CpG island between -444bp and +1509 bp (relevant to +1 TSS site) was identified. The CpG island of *CERIG* contains a high GC content (66.7%) with an observed CpG/expected CpG ratio (Obs_{CpG}/Exp_{CpG}) of 0.729, as calculated by CpG island Searcher (http://cpgislands.usc.edu/) (Takai & Jones, 2002). Given that the defining criteria for a CpG island requires a minimum 200-bp stretch of DNA with a C+G content of 50% and an Obs_{CpG}/Exp_{CpG} in excess of 0.6 (Gardiner-Garden & Frommer, 1987), the *CERIG* promoter region may contain a functional CpG Island. Further analysis showed that there are two sub-regions in the CpG Island (CGI) based on their relatively higher GC content with respect to the rest of the island. The first sub-region was identified between -444 and +280 mainly in the proximal part, and the second one was identified between +525 and +1059 in the first intron (**Fig 4.4 A**).

4.2.1.2 5'azacytidine treatment and *in vitro* promoter methylation

To determine the contribution of the *CERIG* promoter methylation to the expression of this gene, a common demethylating agent, 5' azacytidine, was used. Because of being an analog of cytosine in DNA or cytidine in RNA, 5'azacytidine, which is known as Vidaza ®, titrates out the methyltransferase enzymes (DNMTs) and inhibits their functions by acting as a false substrate (Holliday & Ho, 2002). *CERIG* expression in MCF-7 and MDA-MB-231 cells treated with 5' azacytidine was evaluated by real time RT-PCR using specific primers which were used in section 2.2.1. MCF-7 cells treated with 1μM 5' azacytidine for 4 days resulted in a three-fold increase in *CERIG* expression, while no change was seen in MDA-MB-231 cells (**Fig 4.4 B**), suggesting a possible role for methylation in regulating *CERIG* expression.

A biochemical approach was used to further validate the role of DNA methylation in controlling *CERIG* expression. Methyltransferase M.SssI was used to chemically methylate cytosine residues within the CpG island of *CERIG* that was cloned into the pGL3-basic plasmid. This plasmid, containing the basic luciferase plus the methylated CpG island, and the basic luciferase plasmid were digested using HpaII which cuts the CCGG sequence between two cytosines only if the cytosines are not methylated. After confirming the function of M.SssI in these plasmids, the *in-vitro* methylated plasmid DNAs were transfected into COS-1 cells followed by a luciferase activity assay (Esteve, Chin, & Pradhan, 2005). Methylation of the CpG island sequence of *CERIG* using methyltransferase M.SssI reduced promoter activity by 200-fold (**Fig. 4.4 C**), but the reduction in the control plasmid was less than 10-fold, indicating a role for aberrant methylation in the deregulation of reporter gene activity. The results from the 5'azacytidine treatment and *in-vitro* promoter methylation experiments suggest an inverse correlation between CpG island methylation and *CERIG* expression.
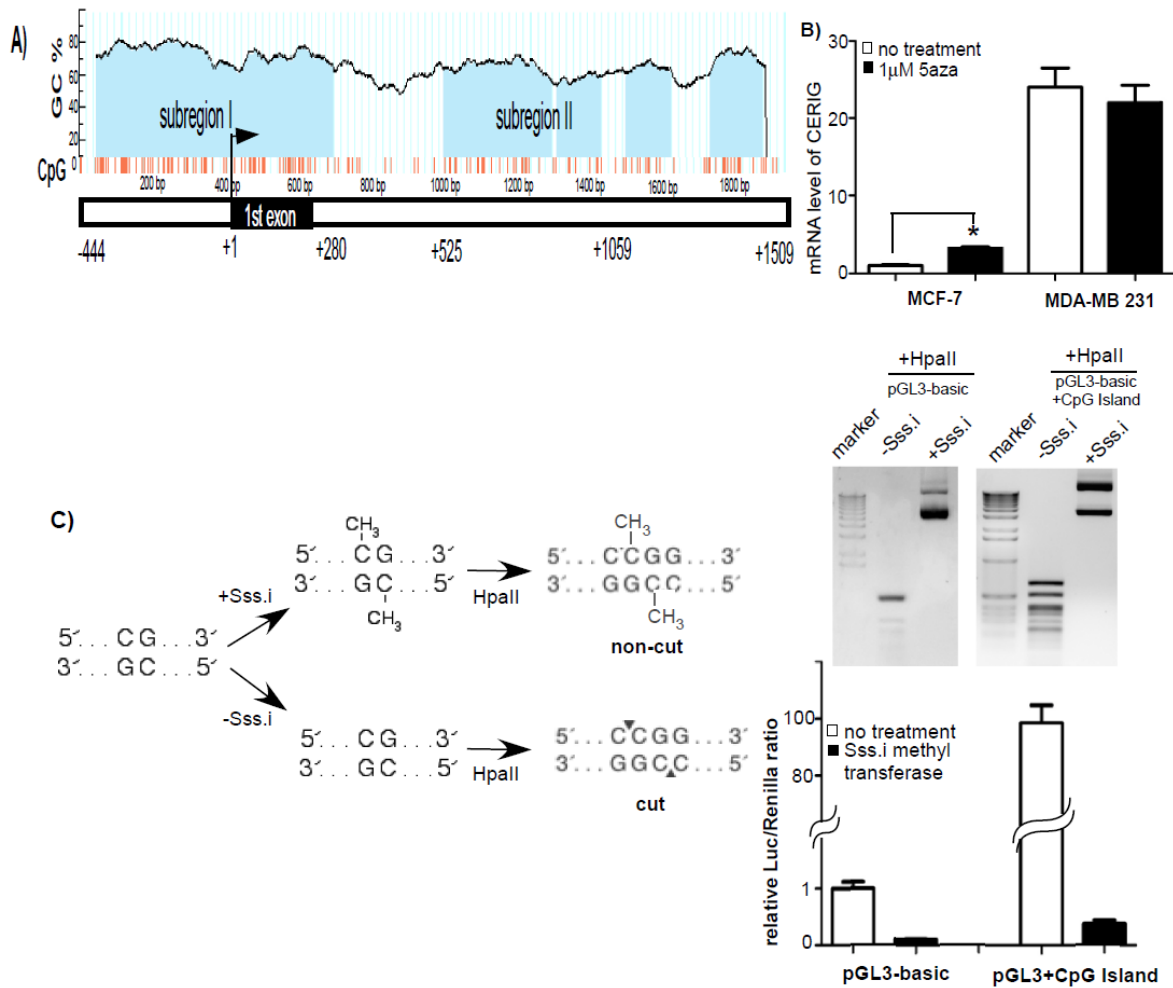
Figure 4.4) Effects of 5'azacytidine and in-vitro promoter methylation in CERIG expression. A) DNA fragment between -444 and +1509 according to the +1 site was identified as a CpG island (CGI). B) MCF-7 and MDA-MB-231 cells were treated with 1 μm 5'Azacytidine for four days followed by RNA isolation. Real-time PCR showed the increase in the activity of *CERIG* in MCF-7 cells, but not MDA-Mb-231. *CERIG* expression was normalized compared to the Hprt-1.C) pGL-3 basic with or without CpG island was in-vitro methylated followed by HpaII digestion. Methylated plasmids were transfected into COS-1 cells to investigate the effects of methylation on reporter gene activity. Methylation in the CpG island decreased the reporter gene markedly with respect to the pGL3-basic control plasmid.

4.2.1.3 Methylation Specific PCR (MSP)

Since the *CERIG* promoter contains a TATA box and a report suggested that CpG islands often do not present in the TATA-box containing promoter (Carninci et al., 2006), we examined whether the deduced CpG island within the *CERIG* promoter is methylated in minimal *CERIG* expressing cells by employing methylation-specific PCR (MSP). MSP distinguishes between unmethylated and methylated CpG islands by using two sets of primers that amplify either unmethylated or methylated sequences after bisulfite treatment, which specifically converts unmethylated cytosines to uracils (Fraga & Esteller, 2002). Therefore, methyl specific primers (M) and unmethyl specific primers (U) were designed for both regions. DNAs from MCF-7 (low expression of *CERIG*) and MDA-MB-231 (high expression of *CERIG*) cells treated with bisulfite were amplified by PCR using two sets of primers. In the CpG island region one, amplification only occurred using unmethylated primers (U) in both cell lines, suggesting no methylation at cytosine residues in this region (**Fig 4.5 A**). In contrast, PCR products were amplified using both methylated primers (M) and unmethylated primers (U) in both MCF-7 cells and MDA-MB-231 cells from subregion two. Comparing PCR products using methylated primers (M), it is notable that amplification in MCF-7 cells was enhanced compared to MDA-MB-231 cells (**Fig 4.5 A**), suggesting the possibility of elevated methylation of the CpG island in the *CERIG* regulatory region in MCF-7 cells.

4.2.1.4 Bisulfite Sequencing of CpG Island

To quantify individual methylated CpG sites, pyrosequencing was performed for both regions in MCF-7 and MDA-MB-231 cells (**Fig 4.5 B**). Bisulfate sequencing primers (BSP) whose activity does not depend on the methylation status of cytosine were designed and were used to clone the region of interest. Since fragments containing either methylated cytosines or

unmethylated cytosines can be amplified by using BSP, the ratio of methylated cytosines to unmethylated cytosines was measured within the specific region. In agreement with our methylated PCR data (**Fig 4.5 A**), no methylation was observed in the first sub-region of the CpG island. In contrast, the average methylation level of the second subregion of the CpG island in MCF-7 cells displayed a significantly high level as compared to MDA-MB-231 cells ($p<0.001$) (**Fig 4.5 C**). The methylation level of 36 CG pairs was quantified between base pairs +525 and +1059 by using a pyrosequencing approach. CG pairs in the following positions; 8 CG pairs between +932 and +1025 have been identified to show a greater difference between MCF-7 and MDA-MB-231 and have been focused on for the evaluation of clinical samples.
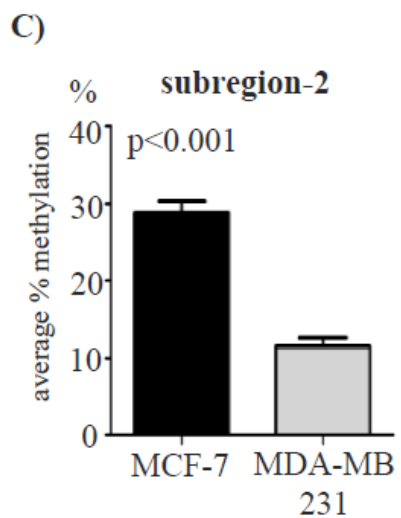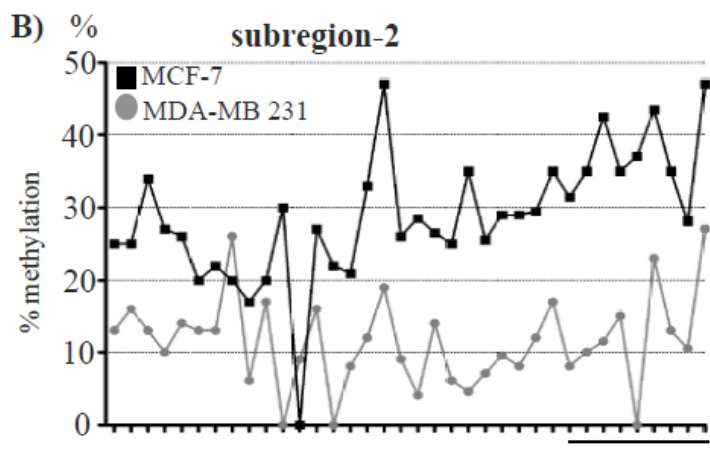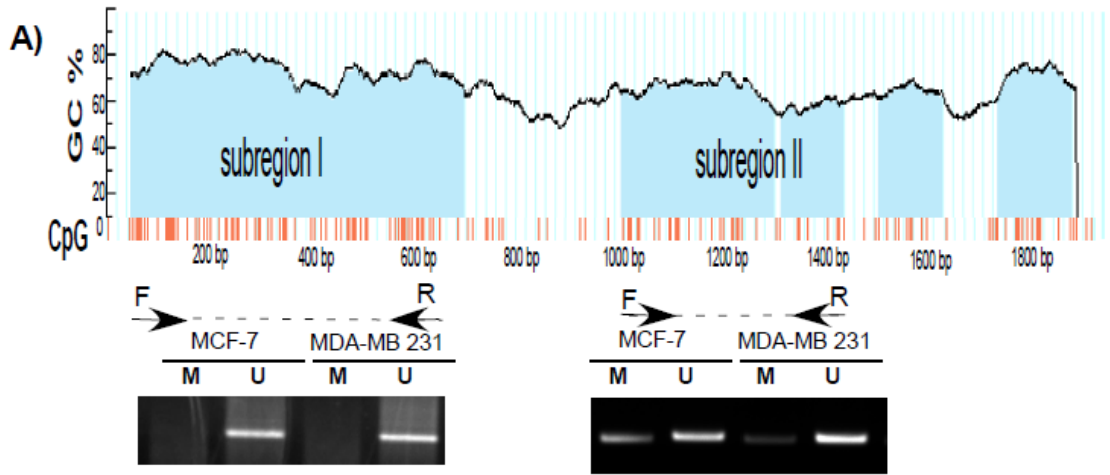
Figure 4.5) Methylation of Cytosines in *CERIG* CpG island. A) First and second subregion in the CpG island was tested by MSP-UMP primer. **B)** Methylation rate in the second subregion was quantified by pyrosequencing method. **C)** Average methylation rate was calculated for two cancer cell lines in which MCF-7 has three fold more DNA methylation compared to the MDA-MB-231 cells.

4.2.1.5 Hypomethylation of the *CERIG* regulatory region in human breast cancer specimens

Data from high throughput analysis of CpG Islands (CGI) by using Human Reference Epigenome Project (UCSC) was initially analyzed by focusing on the region of a CpG island in

chromosome 15. Higher level of methylation in the CpG Island of *CERIG* was observed in

human stem cells, sperm cells and cerebellum tissues, indicating a possible role for methylation

in tissue specific expression of *CERIG*. These high-throughput analyses further indicated that the

methylation level is markedly decreased in the CpG island in human colon carcinoma samples

(third row in **Fig 4.6**). Birkenkamp also showed less methylation in colon carcinomas with

respect to normal colon mucosas in his experiment in which *CERIG* was investigated

individually, but there is no detailed information about the position of methylated residues in the

CpG island or quantification the level of methylation status in their study (Birkenkamp-

Demtroder et al., 2011).
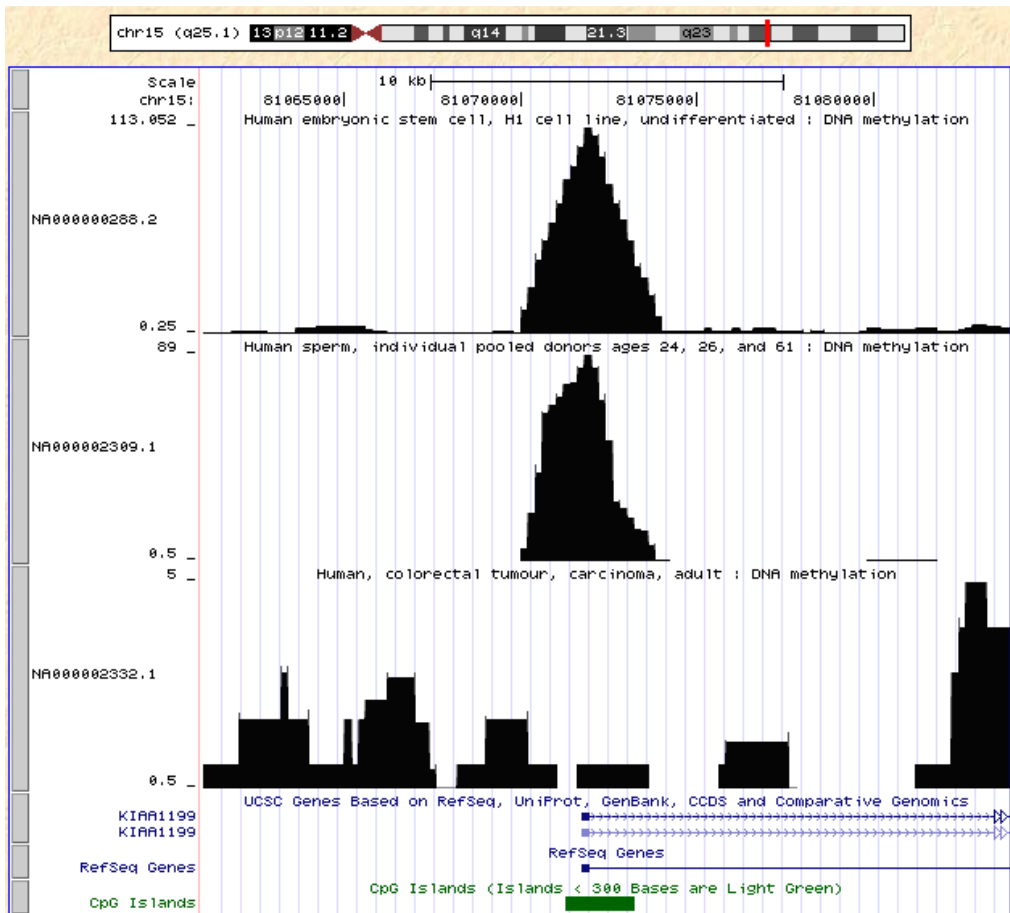
Figure 4.6) Methyl profile of *CERIG* in Epigenome. DNA methylation around the CpG island of *CERIG* was shown according to the methyl profile of Human Reference Epigenome Project (USCS).

To determine the methylation status of the *CERIG* regulatory region in human breast cancer specimens, the methylation level of CpG sites were investigated within the *CERIG* regulatory region from +932 to +1025 in 15 human breast cancer tissues and 10 normal breast tissues. To precisely determine methylation status of breast cancer tissues, a laser-capture microdissection (LCM) method was employed to collect the malignant cancer cells as well as benign epithelial cells as performed in section **2.2.1**. Real time RT-PCR confirmed the previous observation that *CERIG* mRNA is highly upregulated in human breast cancers as compared to normal tissues (**section 2.2.1**). Bisulfite-treated DNA was isolated from microdissected cells followed by pyrosequencing to quantitatively analyze the methylation status of eight CpG sites in malignant breast cancer and normal tissues. Pyrosequencing analysis revealed with high reproducibility that neighboring CpG sites in breast cancer specimens showed a low degree of average methylation as compared to normal controls *(p=0.035)* (**Fig. 4.7 A**). Representative pyrosequencing data from one patient with invasive and normal cells is shown in **Figure 4.7 B**. The prevalence of *CERIG* promoter demethylation in breast tumors is strongly consistent with the enhanced expression of *CERIG* in human breast cancer tissues and aggressive breast cancer cell lines.
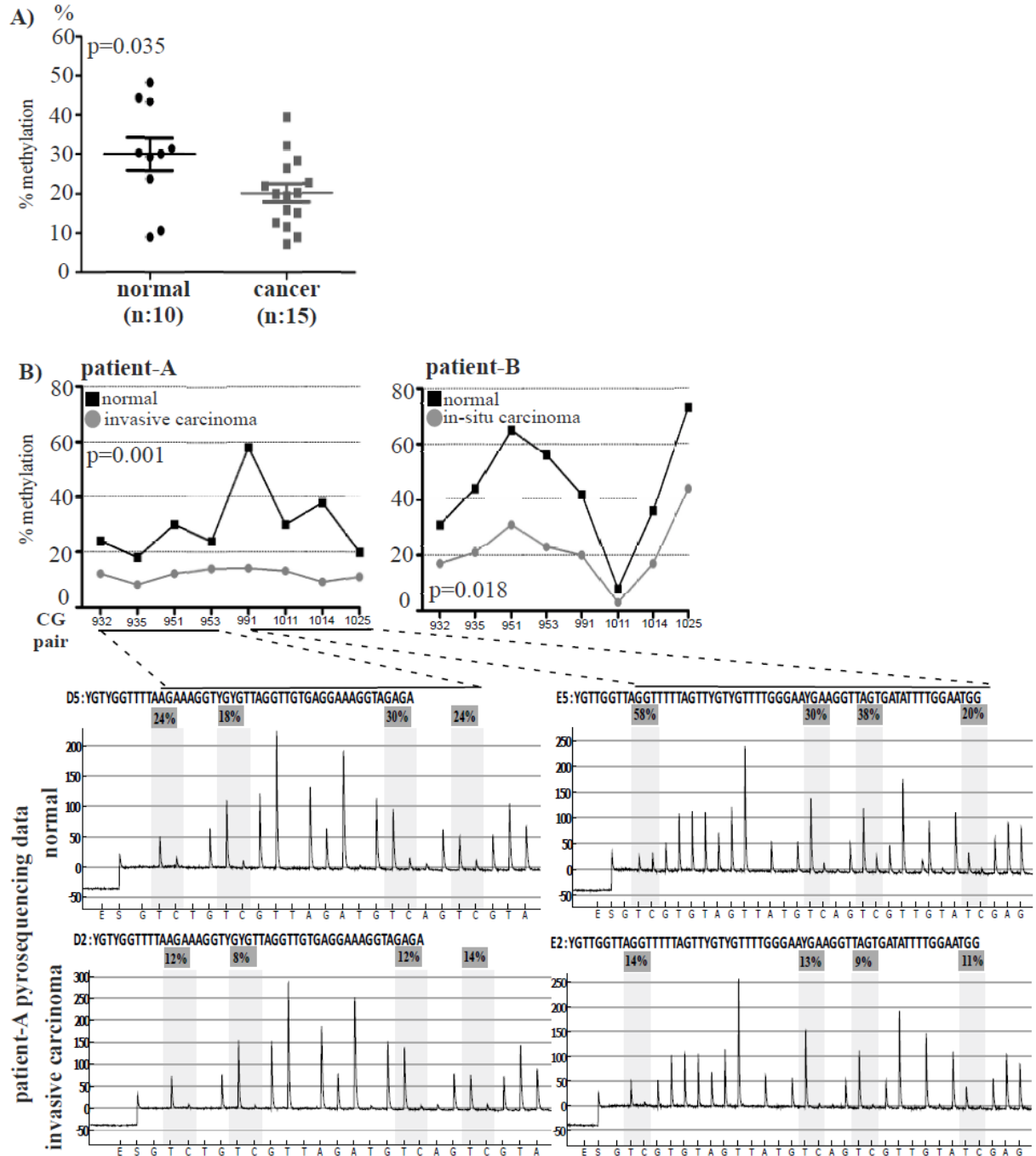
Figure 4.7) Methylation profile of *CERIG* CpG island in breast cancer patients. A) Normal epithelial cells from benign tissue and cancer cells from invasive and in-situ samples were collected by LCM method. Isolated DNA was treated with bisulfate and second subregion was sequenced by pyrosequencing. Normal cells have significantly higher rate of methylation (p< 0.05). **B)** Methylation profile between the +932 and +1025 was shown for the benign and invasive cells of one patient.

4.2.2 Activator and Repressor Histone Modifications in *CERIG* regulatory region

To investigate the role of histone modifications in *CERIG* expression, well documented histone modifying enzyme were identified in the first step. H3K4me3 modification was chosen as an activator marker and H3K27me3 modification was chosen as a repressive marker since both of them were characterized and documented well for their designated function (Wang et al., 2009; Ezhkova et al., 2011). Besides the breast cancer cell lines, Human Umbilical Vein Endothelial Cells (HUVEC) was used as a primary cell line since it has a minimal level of *CERIG* expression that is seven times less than the expression seen in MCF-7 cells. After crosslinking with 0.75% formaldehyde, cell lysates were sonicated and used in a chromatin immunoprecipitation assay (ChIP) followed by real-time PCR to measure the level of each histone mark in all cells tested. Five different primer pairs were used to investigate the various regions from upstream of promoter into first intron region in the downstream of promoter according to the +1 site. Within HUVEC, the nucleosomes around the promoter and downstream regions were markedly trimethylated at histone H3 Lys 27. The level of H3K27me3 was decreased slightly in MCF-7, and the lowest level of H3K27me3 was obtained in MDA-MB-231 cells, which has the highest level of *CERIG*, at four different locations around the promoter and promoter downstream (**Fig 4.8**). When the same regions were analyzed for the H3K4 modification, these regions were heavily enriched with trimethylation of lysine in MDA-MB-231 cells, whereas the level of H3K4me3 was reduced in MCF-7 cells and at a minimal level in HUVEC. Taken together, a correlation between the presence of activation or absence of repression markers and the expression level of *CERIG* in breast cancer cells was established (**Fig 4.8**).
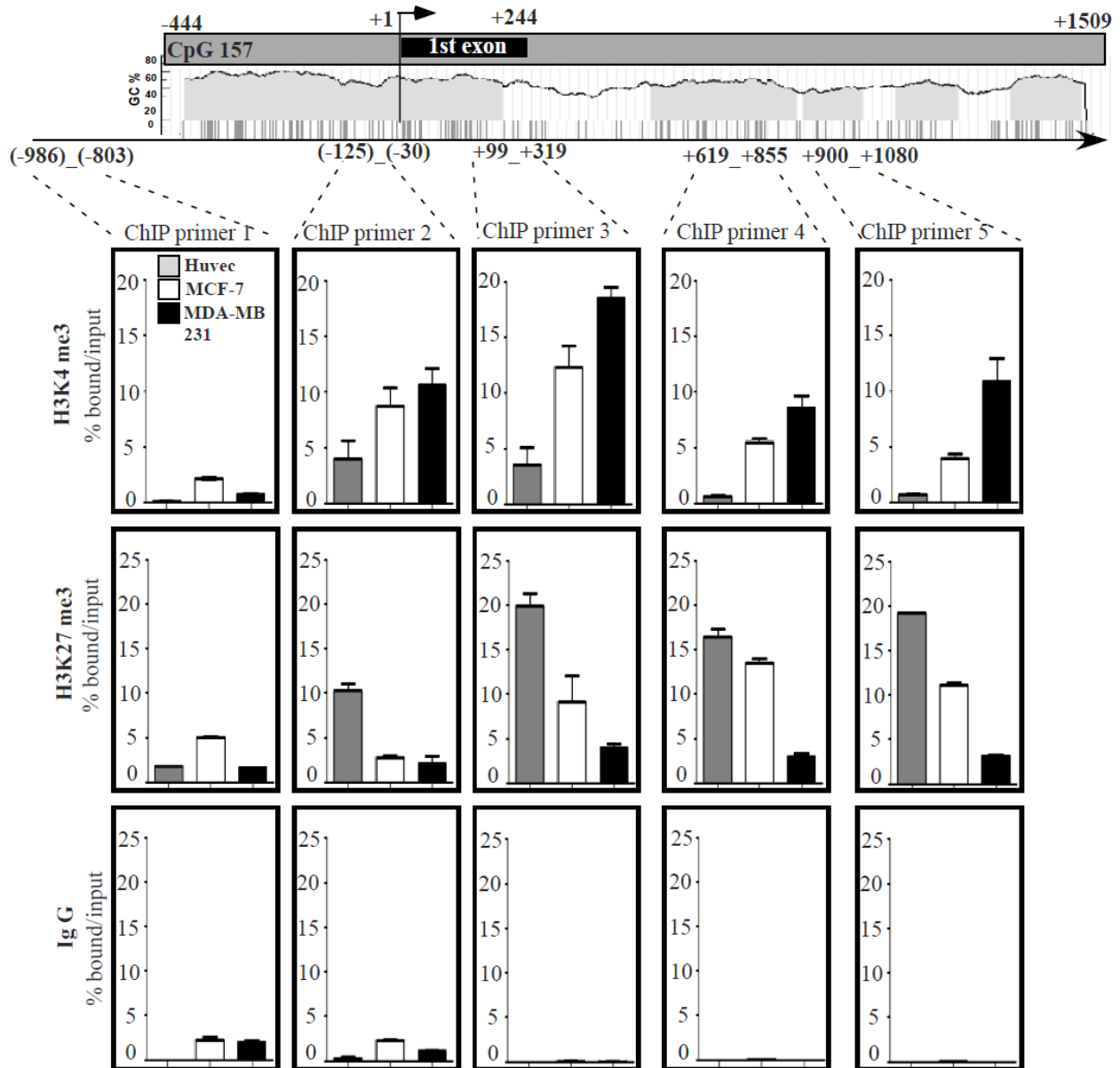
Figure 4.8) Chromatin modifications around the promoter region of *CERIG*. Trimethylation of activation marker (H3K4) and repressive marker (H3K27) were summarized on five different positions of *CERIG* regulatory elements, including distal promoter, promoter and intron.

4.2.3 Effects of Hypoxia on DNA Methylation and Histone Modifications

Hypoxia is known as the deprivation of adequate oxygen supply in the body. During the progression of solid tumors, some portions of the tumor, particularly in the tissue center, become hypoxic. By taking advantage of the low amount of oxygen in their surrounding environment, cancer cells became resistant to chemotherapy and radiotherapy (Denny, 2001). In addition, tumor hypoxia has been shown to induce expression of many oncogenes or growth-related genes. For example, induction of MT1-MMP in renal cell carcinoma (Petrella, Lohi, & Brinckerhoff, 2005) and activation of K-Ras in colon cancer has been shown in hypoxic cells (Zeng, Kikuchi, Pino, & Chung, 2010). To gain information about the response of *CERIG* to hypoxia, breast cancer cells were incubated in hypoxic conditions (1% O_2 and 5% CO_2) for four days. MCF-7 cells expressing low *CERIG* were shown to induce the expression of *CERIG* by six fold after the hypoxia treatment. On the other hand, *CERIG* expression was also induced in MDA-MB-231 cells, but the fold difference was only around two fold (**Fig 4.9 A**). In addition, MCF-7 cells were incubated with $CoCl_2$ which mimics the hypoxic environment. In agreement with the hypoxia results, $CoCl_2$ induced *CERIG* expression more than 5 fold in the MCF-7 cells (**Fig 4.9 A**). Moreover, Niki Calabrase, our colleague in our lab, transfected COS-1 and HeLa cells with the 1.4 kb promoter of *CERIG* followed by exposure to hypoxic and normoxic conditions for 24-48 hours. Approximately three fold increase in the reporter activity of 48 hours hypoxia treated cells further indicates the positive effects of oxygen tension on *CERIG* regulation (**Fig 4.9 B**).
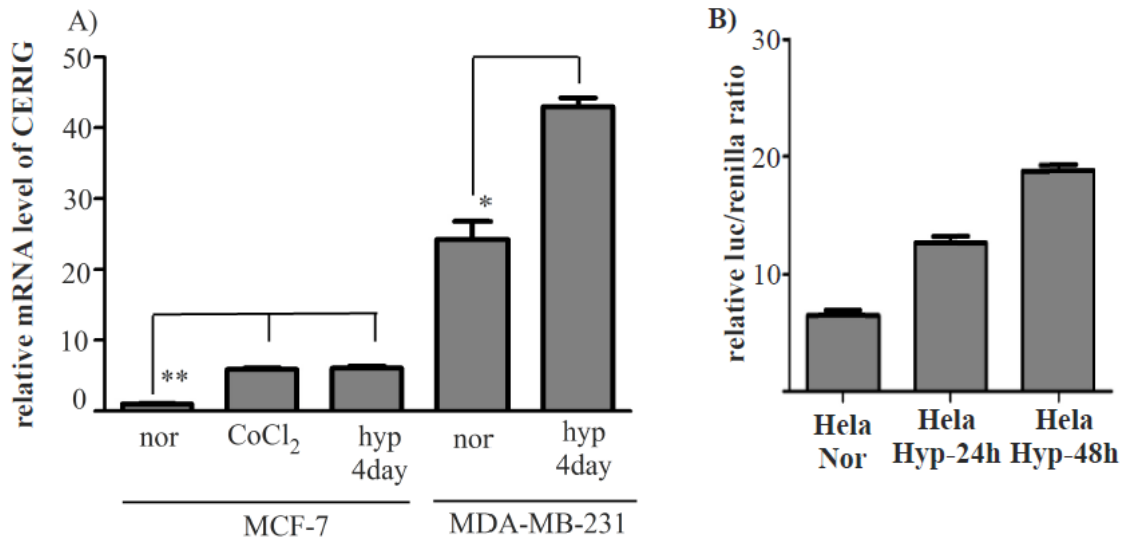
Fig 4.9) Effects of hypoxia on *CERIG* expression. A) Effect of hypoxia or hypoxia mimicking chemical, CoCl2, on the *CERIG* expression in MCF-7 and MDA-MB-231 cells B) Hypoxia induced the reporter gene activity in the HeLa cells transfected with the pGL3 +pro-1.4 kb *CERIG* promoter.

Alteration of epigenetic homeostasis under hypoxia was shown by several research groups and they indicated that these changes promote cancer progression (Johnson, Denko, & Barton, 2008; Jung et al., 2005). Therefore, the role of hypoxia on DNA methylation and for the level of methylation on H3K4 and on H3K27 in the promoter and downstream regions of *CERIG* promoter was investigated. First, the role of hypoxia on DNA methylation was assessed by using the four day hypoxia incubated MCF-7 cells. Since DNA methylation was observed on cytosine residues in the second-subregion of the CpG island **(Fig 4.5)**, only second region was analyzed to see the effects of hypoxia on DNA methylation. DNAs were isolated from normaxia and hypoxia treated MCF-7 cells and methylation status of 36 CG pairs in the second sub-region was analyzed by pyrosequencing. According to the average methylation value of the second subregion in cells cultured under normoxia and hypoxia, hypoxia treatment up to 10 days did not change the methylation level in CpG island of *CERIG* (**Fig 4.10 A**). Next, the effects of hypoxia

on chromatin modifications such as tri-methylation of H3K4 or H3K27 were analyzed. Since histone modifications respond faster to external stimuli, cells incubated in hypoxia chamber were collected at day1, day2 and day4. MCF-7 cells was used for ChIP assay since they responded to the hypoxia treatment to a greater extent than MDA-MB-231 cells (**Fig 4.9 A**). Interestingly, the level of repression marker (H3K27me3) started to reduce immediately after day 1 and the level of activation marker (H3K4me3) required more time to show a response for hypoxia. Collectively, these findings demonstrate that hypoxia might induce the expression level of *CERIG* by increasing the activation marker and decreasing the repression marker without changing the DNA methylation level within the CpG island.
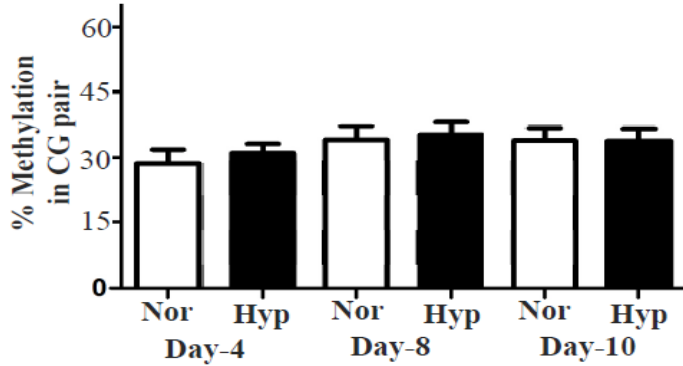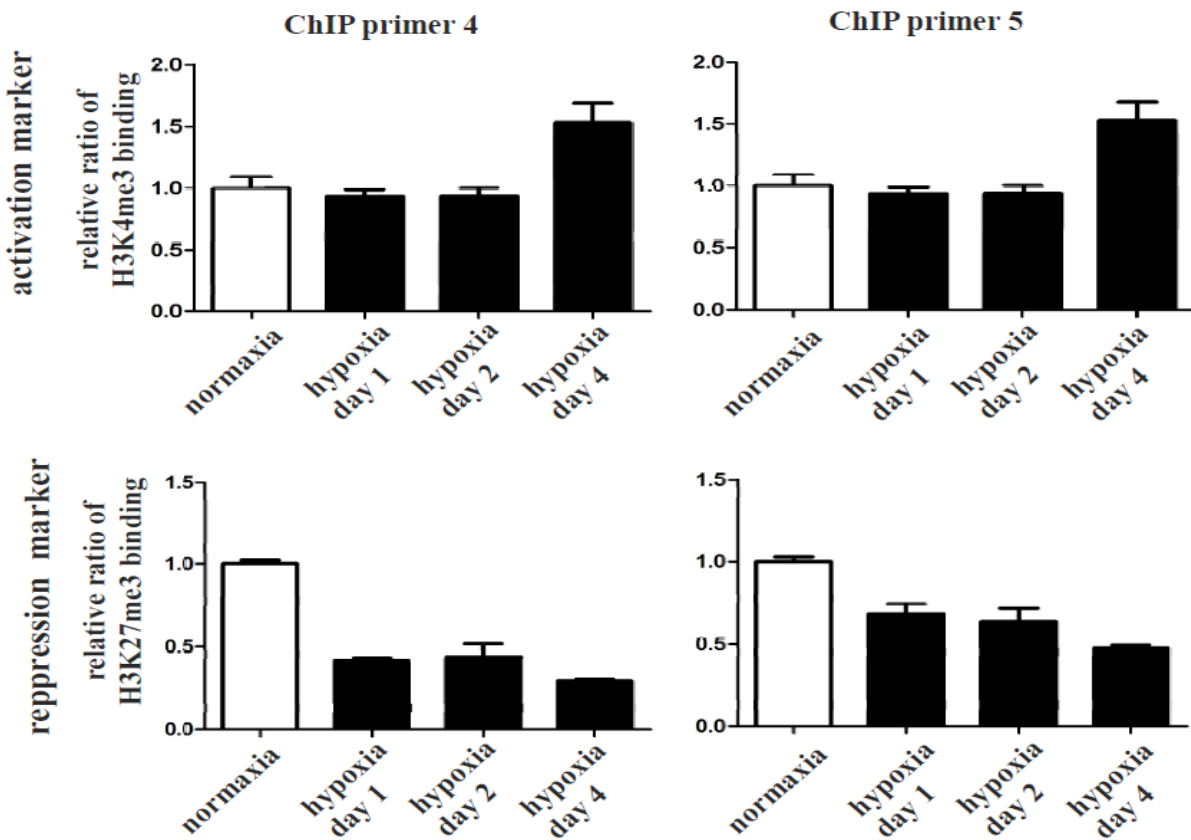
**Figure 4.10) Effects of hypoxia in epigenetic regulation mechanism of *CERIG*.**
A) Hypoxia did not change the Methylation profile of CpG island after 10 days in MCF-7 cells. **B)** Quick response of histone modification enzyme to the hypoxia. H3K27me3 level decrease after 24 hour treatment, but H3K4me3 needs several days to be affected by low oxygen amount.

4.2.4 Jarid1A as a potential H3K4me3 demethylase for *CERIG*

Each modification on the nucleosomes is performed by an individual enzyme, and sometimes more than one enzyme is responsible for the same type of modification. In **Figure 4.3**, histone-modifying enzymes are summarized and the one whose function was affected by hypoxia was labeled with pink. One of these enzymes under the control of hypoxia, named Jarid1A, also known as RBP2 (Retinoblastoma Binding Protein 2), was characterized for its demethylase activity. It has been reported that it specifically catalyzes the demethylation of di- or tri-methyl groups from the activator marker H3K4me3 (Klose et al., 2007). Since tri-methylation level of H3K4 was related with an increase in *CERIG* expression (**Fig 4.8** and **Fig 4.10 B**) and the hypoxia induced the *CERIG* expression (**Fig 4.9**), Jarid1A might be considered as a potential histone demethylase of H3K4me3 in *CERIG* promoter region. In addition to these observations, the ARID domain in Jarid1A was shown to interact with a specific sequence of genomic DNA, CCGCCC, by using NMR study (Tu et al., 2008). Two separate binding sites of Jarid1A were identified in the proximal promoter of *CERIG* after the investigation of "CCGCCC" fragment in the *CERIG* regulatory region. Collectively, all these data raises the possibility that Jarid1A might remove the methyl group(s) of H3K4me3 in the promoter region of *CERIG* and reduce the expression level. To support this possibility, mRNA level of Jarid1A was assessed in cells cultured under hypoxic condition versus normal condition. However, Jarid1A mRNA level did not change after 4 days treatment of hypoxia (**Fig 4.11 A**). In another study, hypoxia was shown to inhibit the function of Jarid1A without changing the Jarid1A mRNA or protein amount (Zhou et al., 2010). Actually, our results are in agreement with this previous publication which shows the inhibition of enzyme activity in the presence of low amount of oxygen that binds to the active cleft of enzyme rather than fluctuation of mRNA or protein (Zhou et al., 2010). In the next step,

Jarid1A was overexpressed in MDA-MB-231 cells which have high level of *CERIG* in our breast cancer cell lines to support a direct link between the level of Jarid1A and *CERIG* expression. Overexpression of Jarid1A reduced *CERIG* expression by half according to semi-quantitative and real-time PCR (**Fig 4.11B**).
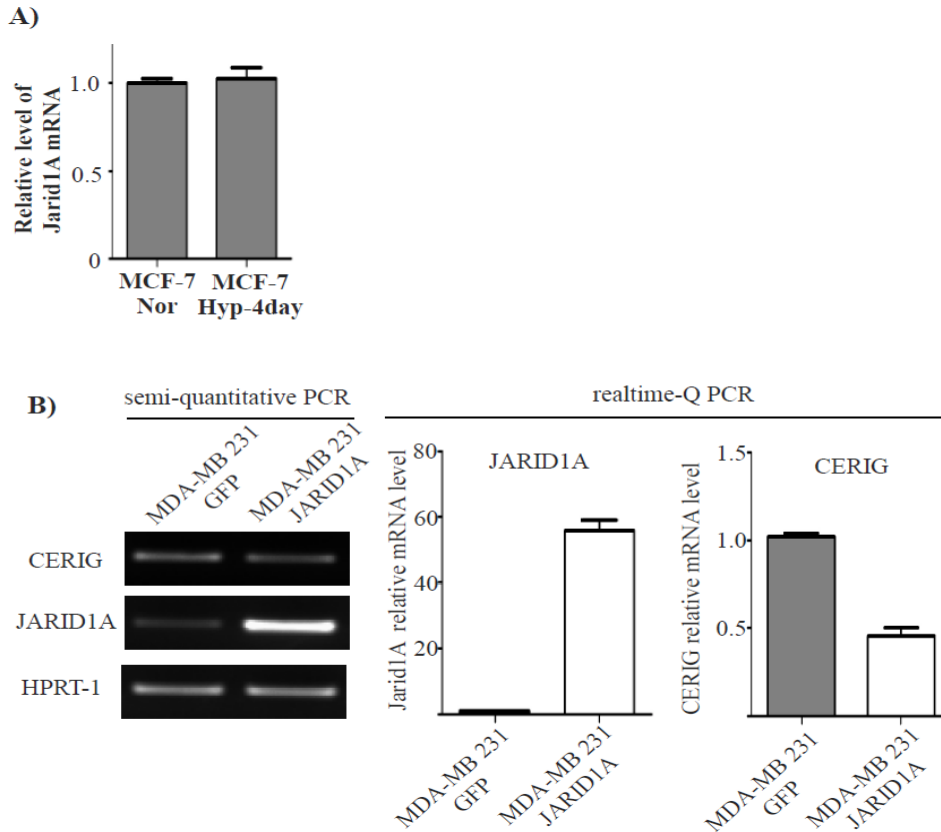

Figure 4.11) Role of Jarid1A in *CERIG* expression. A) Jarid1A level did not change in the presence of low oxygen amount, but it functionally inhibited. **B)** Overexpression of Jarid1A decreased the level of *CERIG* in MDA-MB-231 cells.

Next, the role of overexpressed Jarid1A expression on histone modifications, such as H3K4me3 and H3K27me3 was taken into consideration. After crosslinking of MDA-MB-231 cells overexpressing Jarid1A with formaldehyde, cell lysates were sonicated and followed by pulldown with appropriate antibody. Overexpression of Jarid1A, which was shown to decrease

CERIG expression, significantly decreased the level of H3K4me3 at four different positions within the promoter and first intron regions of *CERIG*, indicating a potential role for Jarid1A as a specific H3K4 demethylase activity for *CERIG*. Overexpression of Jarid1A did not change the methylation level of H3K27 significantly. IgG ChIP assay was performed as a negative control (**Fig 4.12**).
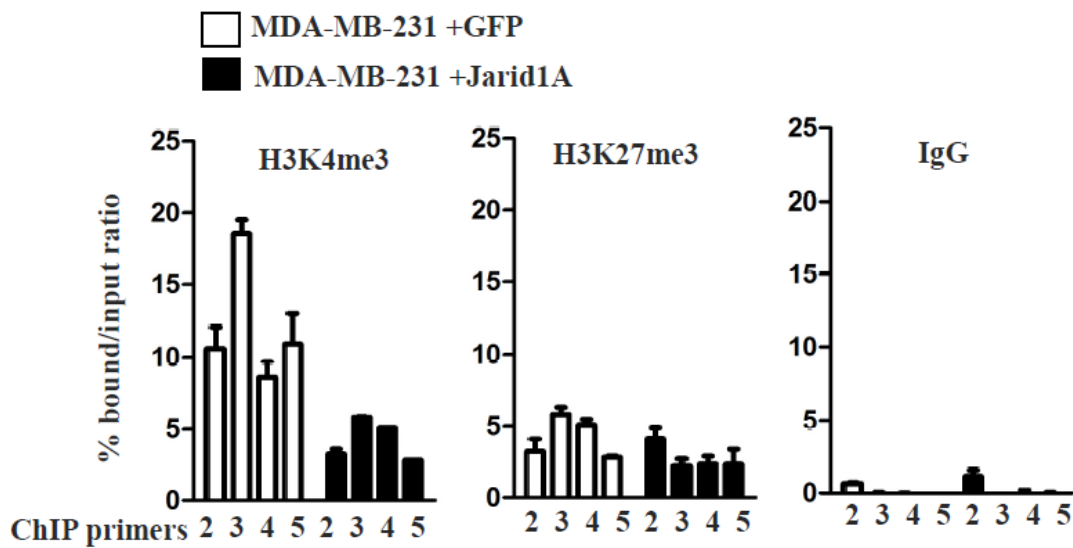

Figure 4.12) Role of Jarid1A overexpression on the histone modifications. The role of Jarid1A overexpression on the histone modifications It significantly decreased the amount of activation marker on the CpG island and promoter region, but it did not alter the level of repression marker. IgG was used as a negative control.

4.3 Discussion

In human genome, 40-50 % of genes have CpG islands within their promoter region with little or no DNA methylation (Fatemi et al., 2005). As expected, there is no DNA methylation in the CpG islands of constitutively expressed housekeeping genes. However, tissue specific genes have been reported to show diverse pattern of methylation in their CpG island in various tissues. Methylation profiles of promoters in tumor suppressor and proto-oncogenes were also subject to change in many different cancers (Counts & Goodman, 1995; Li & Chen, 2011). Taking these observations into consideration, Human Reference Epigenome Mapping Project (UCSC) was initially investigated for *CERIG*, and high level of methylation in the *CERIG* CpG island was observed for human stem cells, sperm cells and cerebellum tissues, indicating methylation as a potential mechanism for the tissue specific expression of *CERIG*. In our study, methylation level in the *CERIG* CpG island of HUVEC cell, which is a primary cell line with low level *CERIG* expression, was measured and reported with hypermethylation. This result was in agreement with the "Human Epigenome" data in which tissues that have low *CERIG* expression and hypermethylated CpG island. Since *CERIG* is highly expressed in the spinal cord and brain, methylation analysis of the CpG island in these tissues will provide vital information that will allow us to fully understand the role of methylation in *CERIG* expression.

In our study, the essential aim is to shed light on the role of DNA methylation in the regulation of *CERIG*. In addition to genetic regulation, DNA methylation has been shown to take a role in the regulation of *CERIG* expression. To begin with, two *in-vitro* experiments were performed. While *in-vitro* promoter methylation decreases the promoter activity of *CERIG*, demethylating reagent 5'azacytidine, which works in the opposite direction of aforementioned assay, increases *CERIG* expression in MCF-7 cells that has low level of *CERIG*. Results from

these studies indicate a potential role of DNA methylation for the regulation of *CERIG* and encourage us to further analyze the quantification of methylation in the CpG island. Of interest, the CpG island of *CERIG* starts in the promoter and extends into the first intron of *CERIG*. Two separate subregions have been identified with high GC content in the following areas; first one locates in -444/+280 region and second one locates in +525/+1059 region. Methylation level of *CERIG* CpG island was markedly reduced in human colon carcinoma samples according to the high throughput analysis. In addition, Birkenkamp *et al* ,2011 also showed less methylation in the CpG island of *CERIG* in colon carcinoma samples with respect to normal colon mucosa (Birkenkamp-Demtroder et al., 2011). However, they did not mention the positions and the methylation rate of cytosines in the *CERIG* CpG island. Therefore, quantification of DNA methylation in individual cytosine residues was focused first. Secondly, hypomethylation of *CERIG* in the second subregion of the CpG island in breast cancer cells and breast primary cancer tissues was identified in detail by using both methylation specific PCR (MSP) and pyrosequencing (Tost & Gut, 2007). In agreement with the observation that CG pairs in the promoter area are generally not methylated in tissue specific genes, no methylated cytosine residues were observed in the first region, which lies in the promoter. Therefore, it was assumed that *de-novo CERIG* expression in breast cancer cell lines is related with the hypomethylation of CG pairs in the second-subregion. Although DNA methylation events in cancer are primarily considered in the inactivation of tumor suppressor genes (Esteller, 2002), mounting evidence has demonstrated that demethylation occurs in cancer progression and results in upregulation of cancer-related genes (Hanada, Delia, Aiello, Stadtmauer, & Reed, 1993; Honda et al., 2004; Jun et al., 2009; Nishigaki et al., 2005; Smith et al., 2009; Tabu et al., 2008; Watt, Kumar, & Kees, 2000; Zhou et al., 2001). Another question that needs to be answered is whether or not

methylation in introns has similar effects as that of methylation in the promoter region regarding gene silencing? In some of the above mentioned studies, demethylation of specific CpG sites within downstream of transcription start site or introns has been reported to induce gene expression. (Felgner et al., 1992; Nishigaki et al., 2005; Roman-Gomez et al., 2005; Tabu et al., 2008). Hence, methylation in the second region of CpG island might be responsible for the repression of *CERIG* in normal tissues with respect to the cancerous tissue or reduction of *CERIG* level in less aggressive cancer cell lines compared to the aggressive one. However, whether methylation directly elicits gene inactivation or is an outcome of gene silencing remains to be determined (Clark & Melki, 2002). Nevertheless, detection of aberrant DNA methylation is considered a promising diagnostic tool in cancer (Harbeck et al., 2008). Our studies identifying a correlation between DNA methylation and human breast cancers provide a novel alternative approach for diagnosis of human breast cancer.

At the beginning of this project, upregulation of *CERIG* was observed after ConA treatment in human fibrosarcoma cells. However there is very limited information about the effects of ConA on DNA methylation in the literature. In one study, ConA was shown to induce the methylase activity because of its mitogenic effect, but methylation level on DNA conserved after ConA treatment (Szyf et al., 1985). Therefore, there might be an indirect interaction between the methylation status of *CERIG* and ConA. Further experiments needs to be done to understand the direct role of ConA for DNA methylation of *CERIG*. One possible scenario for the role of ConA on *CERIG* expression might be that induction of AP-1 DNA binding by ConA inhibits the association of DNMT to the CpG island and results in hypomethylation which increases the level of *CERIG* transcripts. Of note, AP-1 binding on promoter has been shown to inhibit association of DNMTs with CpG island (Kollmann, Heller, & Sexl, 2011).

It is generally accepted that two mechanisms account for transcriptional repression via DNA methylation. In the first mechanism, DNA methylation directly inhibits the binding of transcription factors, such as AP-2, c-Myc, E2F, and NF-κB to their binding sites within promoter sequences (Comb & Goodman, 1990; Watt & Molloy, 1988). The second mode of repression includes binding of proteins, e.g. m5CpG-binding (MeCP) and m5CpG-binding domain (MBD) proteins, specific for m5CpG dinucleotides of methylated DNA. These DNA-protein interactions recruit histone modifying enzymes, which change the structure of chromatin to silence gene expression (Nan, Cross, & Bird, 1998; Rupon, Wang, Gaensler, Lloyd, & Ginder, 2006). Some methyl binding proteins called MBD1, MBD2, MBD3 and MBD4 bind to methylated cytosine residues and bring other chromatin regulating enzymes (Lewis et al., 1992). For example, MBD1 recruits the methyltransferase SETDB1 which methylates the H3K9 as a repressive marker and brings HDACs (histone deacetly transferase) that deacetylates lysines on histone tails for silencing of gene expression (Hashimoto, Vertino, & Cheng, 2010). Recruitment of EZH2 that methylates repression marker (H3K27) and recruitment of Jarid1A that demethylates activation marker (H3K4) by MBDs might be the explanation of how *CERIG* expression was inhibited in the methylated CpG island. These mechanisms suggest that methylation within the promoter region can inhibit transcription initiation. In addition to these mechanisms, methylation of CpG islands in non-*cis*-element containing regions, e.g. extrons or introns, could present a significant obstacle to the progressive transcription complexes, thus inhibiting transcription elongation. Since there are no transcription factor binding sites in the second region of the *CERIG* CpG island, it is assumed that prevention of transcription elongation occurs in non-*CERIG* expressing cells resulting in little to no expression of *CERIG*. In support of this notion, increased H3K27 trimethylation on histone tails of nucleosomes in the second region

of the *CERIG* CpG island has been identified in human umbilical vein endothelial cells (HUVEC) and MCF-7 cells, expressing low levels of *CERIG*, as compared to low level of H3K27 trimethylation in MDA-MB-231 cells. This decrease in *CERIG* expression is possibly due to the fact that methylation of H3K27 potentially leads to more condensed chromatin creating an obstacle that prevents transcription elongation. In the absence of DNA methylation, histone methyltransferase of H3K4 might take a role to produce an accessible chromatin structure (Bird & Wolffe, 1999). Moreover, MDA-MB-231 cells have high H3K4 trimethylation in the second-subregion supporting the higher expression of *CERIG* in this cell type. Additional experiments are necessary to further address these possibilities.

Identification of the binding site for Jarid1A, whose function inhibited under hypoxia, in the *CERIG* promoter encourages us to investigate the role of hypoxia for *CERIG* regulation. While hypoxia did not produce enough stress to change the methylation profile of cytosine residues, it has the ability to change methyl group modifications of histone tails according to several studies (Bachman et al., 2003; Jones & Baylin, 2007). In agreement with the previous publications, there is no change in the level of DNA methylation even after 10 days of hypoxia treatment. However, there was a quick response to hypoxia in the histone modifications within the *CERIG* CpG island. After 4 days of hypoxia treatment, tri-methylation of lysine on the activation marker, H3K4, increased significantly. Therefore, inhibition of Jarid1A by hypoxia might be the reason of this increment. Around fifty percent reduction of the *CERIG* expression in MDA-MB-231 cells after the overexpression of Jarid1A support the hypothesis in which hypoxia regulates the CERIG expression by using its suppressive effect on Jarid1A function.

4.4 Materials and Methods

Identification of CpG island

For the region located between -444 and +1509, the observed ratio of CG pair is 4.55%, and then observed $_{CG}$/expected $_{CG}$ ratio becomes 4.55/6.25=0.729. The GC content in this region is 66.7 %. Since both value is above the critical value of the CpG island definition i.e 0.6 for observed CG/ expected CG and minimum 50 % for GC content in the minimum area of 200 bp. This region was also confirmed by UCSC genome browser and CpG island search program as a potential CpG island.

5'azacytidine treatment

5'-azacytidine (5'-Aza) was provided by Sigma. Stock solutions of 5-AzaC (5mM) were prepared in PBS and stored at -80 ^{0}C. The stock solutions were further diluted in PBS (1 µM) for cell culture experiments.

In vitro promoter methylation

An aliquot of 5 µg plasmid DNA was incubated with *Sss.I* methyltransferase (NEB) for 2 hours at 37 ^{o}C in a reaction mixture containing 50 mM Tris–HCl, pH 7.5, 300 mM *S*-adenosylmethionine (SAM), 10 mM EDTA and 5 mM dithiothreitol (DTT). The reaction was terminated by heating at 65 ^{o}C followed by phenol extraction and chloroform/isoamyl alcohol (24:1). The DNA was precipitated with ethanol. The efficiency of methylation was controlled by digestion of methylated pGL3 and pGL3 including CpG island with *Hpa*II incubation at 37 ^{o}C for one hour.

DNA methylation

DNA was isolated from cancer cell lines and tissues by using DNeasy Kit of Qiagen. For the FFPE treated tissue sections, extra steps were performed according to the company instructions. 1 µg of isolated DNA was bisulfate treated by using the EZ DNA methylation gold kit (Zymo Research). DNA was eluted in 20 µl and 2 µl of the elution was used in the MSP and BSP PCR reactions.

Methylation Specific PCR (MSP)

Methyl specific primers and unmethyl specific primers were designed by using the methprimer bioinformatic application (http://www.urogene.org/methprimer/rules). MSP and UMP primers were shown in **Table 4.1**.PCR was performed with following conditions; 95 $^{\circ}$C for 5 min, (94 $^{\circ}$C for 30 s, 60 $^{\circ}$C for 30s and 72 $^{\circ}$C for 30s) for 30 times, 72 $^{\circ}$C for 7 min.

Primer Name	Target Region	Primer Sequence
MSP1-for	1st sub-region of CpG island	TTAGCGTTTAAGTAGAGTTTAGCGC
MSP1-rev	1st sub-region of CpG island	CCGAAACGTAAACAAACGAA
UMP1-for	1st sub-region of CpG island	AGTGTTTAAGTAGAGTTTAGTGTGG
UMP1-for	1st sub-region of CpG island	AGTGTTTAAGTAGAGTTTAGTGTGG
MSP2-for	2nd sub-region of CpG island	GTTCGGGAGAGTGTTTCGTC
MSP2-rev	2nd sub-region of CpG island	GCGTCGCTAAATACCTAACG

UMP2-for	2nd sub-region of CpG island	GTTTGGGAGAGTGTTTTGTTGA
UMP2-rev	2nd sub-region of CpG island	AACACATCACTAAATACCTAACACT

Table 4.1 List of MSP and UMP primers

Bisulfite Sequencing PCR

Bisulfite sequencing primers (BSP) to clone the second subregion and several sequencing primers were designed by using Pyrosequencing™Assay Design Software. Primers were shown in **Table 4.2.** 2585 and 2586 primers were biotinylated at their 5' end to produce PCR fragment used in the pyrosequencing reactions. PCR was performed with following conditions; 95 $^{\circ}$C for 5 min, (94 $^{\circ}$C for 30 s, 56 $^{\circ}$C for 30s and 72 $^{\circ}$C for 30s) for 30 times, 72 $^{\circ}$C for 7 min.

Primer Number	Target Sequence	Primer Sequence
2585	2nd subregion	TTAGAGGGTATTTAGAGGGGGTAGA
2586	2nd subregion	CCTTAAAATCCATTCCAAAATATCA
2718	pyro-seq1	TTTTTTTTTTGAATGGT
2719	pyro-seq2	AGTATTATAAGGGTTAGGTT
2720	pyro-seq3	GATTGGGAGTTAGTTTA
2721	pyro-seq4	TTTTTTTTTTAGTTTTTTT
2722	pyro-seq5	AAAAAAACTAAAAAAAAA

2723	pyro-seq6	ACATTACAATTACCTCAACT
2724	pyro-seq7	GATGTTTAGATTGAGATTTT
2725	pyro-seq8	TGGTTTTTGGTAAAGGA
2726	pyro-seq9	GGTAGAGAGTTTTGGTAGAA

Table 4.2 List of BSP and pyrosequencing primers

Chromatin Immunoprecipitation assay (ChIP)

It was previously described in section 3.4

5. Summary

Understanding the steps of metastasis, which accounts the 90% of deaths in cancer, is very crucial for the treatment of patients. *CERIG*, or *KIAA1199*, was previously identified with its role in cancer cell migration and invasion that help the dissemination of cancer cells into adjacent tissue or secondary site in the body. *CERIG* or *KIAA1199* was identified as one of the induced gene after Concanavalin A (ConA) treatment in human fibrosarcoma cell line. Initial studies in our lab demonstrated the product of *CERIG* gene as an endoplasmic reticulum resident protein which plays a critical role in the migration and invasion ability of cancer cells. Since ConA treatment has been shown to change the trafficking of proteases and matrixmetallo-proteinases towards to the plasma membrane where they become functional for the degradation of extracellular matrix, *CERIG* was thought to carry a protease function initially. However, no protease activity of *CERIG* was identified in our loss-or gain-of function studies.

Understanding the genetic and epigenetic mechanisms may help us to treat diseases related with aberrant expression of *CERIG*. In the current project, regulation mechanisms of *CERIG* in breast cancer progression have been identified. **First**, overexpression of *CERIG* in different types of cancer was demonstrated by using oncomine cancer data sets. In another study, low survival probability of breast cancer patients with high *CERIG* expression was identified by using the GEO profiles of three data sets from NCBI. Next, overexpression profile of *CERIG* was recapitulated in breast cancer tissue compared to the benign tissue by using the LCM (laser capture microdissection) isolated cells. High *CERIG* expression in aggressive cancer cells with respect to the less-aggressive cells in prostate, breast and colon cancers confirmed the role of *CERIG* expression in aggressiveness. As gene regulation starts with transcription, explanation of how transcription starts and what type of molecules are taking a role in transcription give as a clue for the inhibitory mechanism of a gene. **Second**, interactions of AP-1 and NF-κB transcription factors in the proximal and distal part of *CERIG* promoters were identified for genetic regulation of *CERIG* by using promoters carrying mutations for reporter assay, electrophoretic mobility shift assay and chromatin immunoprecipitation assay. Since aberrant expression of these two factors has been well documented for the upregulation of oncogenes or growth-related genes in different types of cancer, identification of these factors for *CERIG* gene regulation confirmed the important role of *CERIG* in cancer progression. Finding an antagonist molecule for AP-1 and NF-κB binding or finding a molecule which attenuates their upstream activation signal becomes a potential tool for downregulation of *CERIG*. Interestingly, positioning of AP-1 on target DNA has also been linked to the ConA treatment in several studies. Therefore it is assumed that ConA induction of *CERIG* in cancer cells might depend on the AP-1-Con A axis that needs to be discovered. Moreover, negative regulatory element in the

distal part of *CERIG* promoter was identified and waited for further characterization. These genetic regulators and treatments are summarized in **Figure 5.1** with green color if it is an activator or red color if it is a repressor for *CERIG* expression. **Third**, CpG island, which lies in the promoter and downstream region of *CERIG* compared to the TSS, about 1.9 kb was identified for *CERIG* via CpG island prediction programs and human genome browsers. Its presence initiated the studies to find the potential role of epigenetics for *CERIG* regulation. DNA hypomethylation has been linked to the upregulation of *CERIG* by using both *in-vitro* and *in-vivo* studies. Detailed analysis of CpG island by MSP and pyrosequencing revealed the occurrence of DNA methylation in the second subdomain of CpG island which exists in the first intron of *CERIG*. Hypomethylation of cytosine residues in this region was confirmed in both high-*CERIG* breast cancer cells compared to the low-*CERIG* cells and breast cancer tissue compared to the benign tissue. Potential inhibitory mechanisms depend on the methylated cytosine residues are shown in **Figure 5.1** with the same color code. For example, methylation of cytosines by DNMTs produces a docking site for MBDs (methyl binding proteins) which recruit some potential chromatin modifiers to the CpG island of *CERIG*. **Fourth,** another epigenetics regulation mechanism, called histone modifications, was evaluated since evidence of links between DNA methylation and histone modifications is accumulating. Two well-documented histone modifications, H3K4 for activation and H3K27 for repression, were used for further investigation. High level tri-methylation of H3K4, in high-*CERIG* cells and high level tri-methylation of H3K27, in low-*CERIG* cells were identified in breast cancer cell lines. Since cancer cells grow faster and become more migratory in the low oxygen containing environment, effect of hypoxia on these histone modifications was tested in the aspect of *CERIG* expression. *CERIG* expression has been shown to be induced by hypoxia treatment via both promoter

reporter assay based on chemiluminescence reading and q-PCR which based on endogenous transcription level. Studies trying to understand how hypoxia induces *CERIG* transcript have indicated that while low oxygen amount (1%) increases the level of activation marker, it decreases the repression marker on histone tails without altering the level of DNA methylation. Moreover, Jarid1A, which demethylase methyl group from activation marker of H3K4me3, was reported to be less functional under hypoxic environment. Fifty percent reduction in the expression level of *CERIG* after the overexpression of Jarid1A further confirmed the potential role of Jarid1A in the epigenetic regulation of *CERIG*. Further understanding of these genetic and epigenetic mechanisms may help us to eliminate the function of *CERIG* for cell migration and invasion during cancer dissemination.
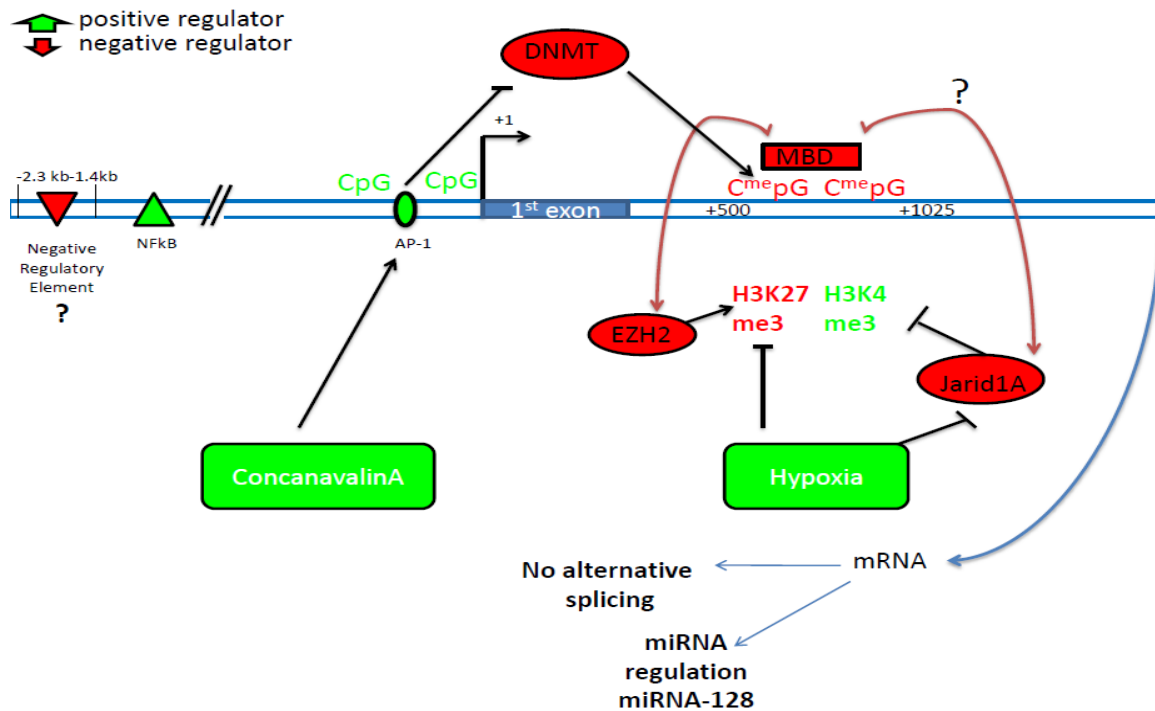

Figure 5.1 Summary of genetic and epigenetic regulation of *CERIG*. Positive factors for *CERIG* regulation are labeled with green color and negative regulators are labeled with red color.

6. Future Directions

This project was initiated after the identification of *CERIG* as a novel gene with its role in cell migration and invasion in cancer. *CERIG* was found upregulated in human fibrosarcoma cells after the treatment of ConA that is used as a mitogen in the activation of thymocytes. Since it has not been studied in detail for the activation of growth or protease related genes in the literature, the mechanism of how ConA induces CERIG expression is not known today. There might be a speculation for the role of ConA on *CERIG* induction. After the identification of the role of AP-1 in the *CERIG* regulation, ConA-AP-1 axis, which was shown before, will gain an interest for further investigation which might shed light on the role of ConA in growth related genes by using *CERIG* as a model gene.

In the current study, the promoter of *CERIG* was characterized for the first time and several regulatory regions were identified for its regulation. Initially, three negative glucocorticoid regulatory elements (nGRE) were identified between -1425 and -2341 bp *in-silico* studies, but further experiments might help us to find suppressor molecules for *CERIG* expression. Another study that might be done to find an inhibitor of *CERIG* depends on the screening of compound (chemical) library by using the promoter of *CERIG*. Since there are vast amount of protein and mRNA in the cells, high concentration of compounds are required to inhibit these molecules. However, since there are two copies of DNAs exist for each gene; inhibition of a gene might be easy and more productive at the transcription level rather than mRNA or protein level. For example, Vincent Yang identified novel-small molecule compounds for the inhibition of Kruppel-like factor 5 at the transcription level in colorectal cancer by using high-throughput screening (Bialkowska, Du, Fu, & Yang, 2009). In their study, they cloned the promoter of their target gene in pGL3 luciferase plasmid and screen the effects of compounds by

monitoring the alteration of luminescence signal intensity. By taking advantage of this simple approach, the compound libraries will be screened for *CERIG* promoter. Since 1.4 kb fragment in the *CERIG* promoter was identified with maximum activity of reporter gene expression (Luciferase), this construct will be used in the screening experiment. I also started the collaboration with Dr. Chunhong Yan who produces cancer cell line (F55) that allows us the homologous recombination of transgene at specific position inside the genome rather than random insertion. Since stable and high gene expression is necessary for high-throughput screening, *CERIG* promoter (1.4 kb)-Luc transgene will be inserted into FRT (Flip recombinase target) containing plasmid and the reporter gene activity will be tested in the presence of different compounds by using F55 cell (Nair et al., 2008; Yan & Boyd, 2006; Yan, Wang, Aggarwal, & Boyd, 2004).

In our current study, low O_2 (1%) environment, which was known to induce the expression profile of many oncogenes and growth-related genes, was reported to increase *CERIG* expression in less aggressive cancer cell lines, such as MCF-7 and SW480. Hypoxia induced the tri-methylation level of 4th lysine on H3 tail around *CERIG* promoter and CpG island. This additive effect might depend on the inhibition of the demethylase such as Jarid1A during hypoxia. Our preliminary data support the involvement of Jarid1A in *CERIG* regulation. Therefore, further investigation of the role of Jarid1A in the methylation/ demethylation status of H3K4 would be very important to understand the histone epigenetic regulation of *CERIG*. Gain- or loss of function assays will be performed to understand the role of Jarid1A in high-or low *CERIG* cells in different cancer types, such as breast and colon. The association of Jarid1A and CERIG might also be investigated in human tissues at mRNA and protein level.

Besides the DNA methylation and chromatin modification, miRNAs started to be considered as an epigenetic regulation mechanism. Because of increasing interest in the area of RNA interference, many researchers focused on the identification of miRNAs that are transcribed endogenously from human genome. So far, thousand different microRNAs have been identified in human genome (Bentwich et al., 2005) and some of their aberrant expression have been reported to take a role in various types of cancer. Now, there are many bioinformatic tools helping us the identification of putative miRNA sequence in the 3'UTR sequence of mRNA. One of the well known miRNA prediction program, TargetScanHuman (http://www.targetscan.org/) was used to find potential miRNA in *CERIG* 3' UTR (**Fig 6.1**).
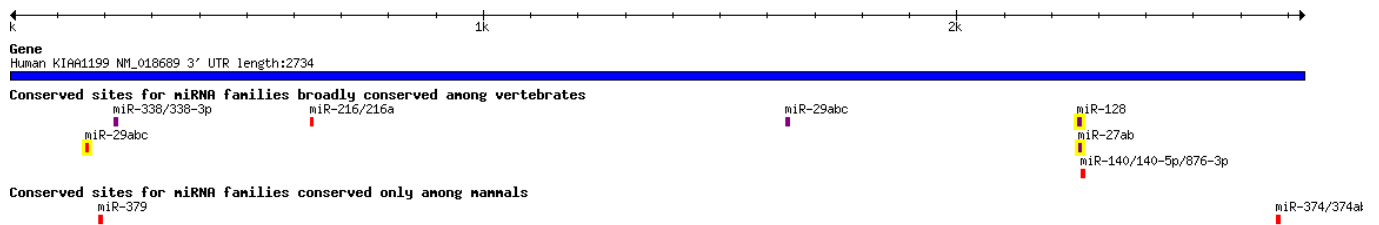

Figure 6.1 Putative miRNAs in the 3' UTR of *CERIG*/KIAA1199. miRNAs labeled with yellow frame were identified with higher chance. miR-128, miR-27 and miR-29 are identified with high probability.

Preliminary data in which miRNA array was evaluated to screen the miRNA profile of MCF-7 versus MDA-MB-231 cells demonstrated the upregulation of some microRNAs in MCF-7 cells compared to MDA-MB-231. miRNAs 128 was found both in the upregulated miRNA list in MCF-7 versus MDA-MB-231 cells and in the hypothetical miRNA list of *CERIG* by target scan. Therefore, future studies will be needed to determine potential miRNAs for *CERIG* regulation. miR-128 might be a good starting point for this line of investigation.

Reference:

Abe, S., Usami, S., & Nakamura, Y. 2003. Mutations in the gene encoding KIAA1199 protein, an inner-ear protein expressed in Deiters' cells and the fibrocytes, as the cause of nonsyndromic hearing loss. *J Hum Genet*, 48(11): 564-570.

Adib, T. R., Henderson, S., Perrett, C., Hewitt, D., Bourmpoulia, D., Ledermann, J., & Boshoff, C. 2004. Predicting biomarkers for ovarian cancer using gene-expression microarrays. *Br J Cancer*, 90(3): 686-692.

Agger, K., Cloos, P. A., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., Issaeva, I., Canaani, E., Salcini, A. E., & Helin, K. 2007. UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature*, 449(7163): 731-734.

Albrecht-Buehler, G. 1977. The phagokinetic tracks of 3T3 cells. *Cell*, 11(2): 395-404.

Ameyar, M., Wisniewska, M., & Weitzman, J. B. 2003. A role for AP-1 in apoptosis: the case for and against. *Biochimie*, 85(8): 747-752.

Bachman, K. E., Park, B. H., Rhee, I., Rajagopalan, H., Herman, J. G., Baylin, S. B., Kinzler, K. W., & Vogelstein, B. 2003. Histone modifications and silencing prior to DNA methylation of a tumor suppressor gene. *Cancer Cell*, 3(1): 89-95.

Balu-Maestro, C., Chapellier, C., Bleuse, A., Chanalet, I., Chauvel, C., & Largillier, R. 2002. Imaging in evaluation of response to neoadjuvant breast cancer treatment benefits of MRI. *Breast Cancer Res Treat*, 72(2): 145-152.

Bannister, A. J., & Kouzarides, T. 2011. Regulation of chromatin by histone modifications. *Cell Res*, 21(3): 381-395.

Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., Barzilai, A., Einat, P., Einav, U., Meiri, E., Sharon, E., Spector, Y., & Bentwich, Z. 2005. Identification of hundreds of conserved and nonconserved human microRNAs. *Nat Genet*, 37(7): 766-770.

Bialkowska, A. B., Du, Y., Fu, H., & Yang, V. W. 2009. Identification of novel small-molecule compounds that inhibit the proproliferative Kruppel-like factor 5 in colorectal cancer cells by high-throughput screening. *Mol Cancer Ther*, 8(3): 563-570.

Bird, A. P., & Wolffe, A. P. 1999. Methylation-induced repression--belts, braces, and chromatin. *Cell*, 99(5): 451-454.

Birkenkamp-Demtroder, K., Maghnouj, A., Mansilla, F., Thorsen, K., Andersen, C. L., Oster, B., Hahn, S., & Orntoft, T. F. 2011. Repression of KIAA1199 attenuates Wnt-signalling and decreases the proliferation of colon cancer cells. *Br J Cancer*, 105(4): 552-561.

Bonasio, R., Tu, S., & Reinberg, D. 2010. Molecular signals of epigenetic states. *Science*, 330(6004): 612-616.

Cao, J., Chiarelli, C., Kozarekar, P., & Adler, H. L. 2005. Membrane type 1-matrix metalloproteinase promotes human prostate cancer invasion and metastasis. *Thromb Haemost*, 93(4): 770-778.

Cao, J., Chiarelli, C., Richman, O., Zarrabi, K., Kozarekar, P., & Zucker, S. 2008. Membrane type 1 matrix metalloproteinase induces epithelial-to-mesenchymal transition in prostate cancer. *J.Biol.Chem.*, 283(10): 6232-6240.

Cao, J., Kozarekar, P., Pavlaki, M., Chiarelli, C., Bahou, W. F., & Zucker, S. 2004. Distinct roles for the catalytic and hemopexin domains of membrane type 1-matrix metalloproteinase in substrate degradation and cell migration. *J Biol Chem*, 279(14): 14129-14139.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C. A., Taylor, M. S., Engstrom, P. G., Frith, M. C., Forrest, A. R., Alkema, W. B., Tan, S. L., Plessy, C., Kodzius, R., Ravasi, T., Kasukawa, T., Fukuda, S., Kanamori-Katayama, M., Kitazume, Y., Kawaji, H., Kai, C., Nakamura, M., Konno, H., Nakano, K., Mottagui-Tabar, S., Arner, P., Chesi, A., Gustincich, S., Persichetti, F., Suzuki, H., Grimmond, S. M., Wells, C. A., Orlando, V., Wahlestedt, C., Liu, E. T., Harbers, M., Kawai, J., Bajic, V. B., Hume, D. A., & Hayashizaki, Y. 2006. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet*, 38(6): 626-635.

Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A., Frisch, M., Bayerlein, M., & Werner, T. 2005. MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics*, 21(13): 2933-2942.

Charron, M., DeCerbo, J. N., & Wright, W. W. 2003. A GC-box within the proximal promoter region of the rat cathepsin L gene activates transcription in Sertoli cells of sexually mature rats. *Biol Reprod*, 68(5): 1649-1656.

Chen, G., & Goeddel, D. V. 2002. TNF-R1 signaling: a beautiful pathway. *Science*, 296(5573): 1634-1635.

Christensen, J., Agger, K., Cloos, P. A., Pasini, D., Rose, S., Sennels, L., Rappsilber, J., Hansen, K. H., Salcini, A. E., & Helin, K. 2007. RBP2 belongs to a family of demethylases, specific for tri-and dimethylated lysine 4 on histone 3. *Cell*, 128(6): 1063-1076.

Clark, S. J., & Melki, J. 2002. DNA methylation and gene silencing in cancer: which is the guilty party? *Oncogene*, 21(35): 5380-5387.

Cohen, I., Poreba, E., Kamieniarz, K., & Schneider, R. 2011. Histone modifiers in cancer: friends or foes? *Genes Cancer*, 2(6): 631-647.

Comb, M., & Goodman, H. M. 1990. CpG methylation inhibits proenkephalin gene expression and binding of the transcription factor AP-2. *Nucleic Acids Res*, 18(13): 3975-3982.

Counts, J. L., & Goodman, J. I. 1995. Hypomethylation of DNA: a possible epigenetic mechanism involved in tumor promotion. *Prog Clin Biol Res*, 391: 81-101.

D'Errico, M., de Rinaldis, E., Blasi, M. F., Viti, V., Falchetti, M., Calcagnile, A., Sera, F., Saieva, C., Ottini, L., Palli, D., Palombo, F., Giuliani, A., & Dogliotti, E. 2009. Genome-wide expression profile of sporadic gastric cancers with microsatellite instability. *Eur J Cancer*, 45(3): 461-469.

Denny, W. A. 2001. Prodrug strategies in cancer therapy. *Eur J Med Chem*, 36(7-8): 577-595.

Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. D., & Siebert, P. D. 1996. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci U S A*, 93(12): 6025-6030.

Dwyer, J. M., & Johnson, C. 1981. The use of concanavalin A to study the immunoregulation of human T cells. *Clin Exp Immunol*, 46(2): 237-249.

Emmert-Buck, M. R., Bonner, R. F., Smith, P. D., Chuaqui, R. F., Zhuang, Z., Goldstein, S. R., Weiss, R. A., & Liotta, L. A. 1996. Laser capture microdissection. *Science*, 274(5289): 998-1001.

Escarcega, R. O., Fuentes-Alexandro, S., Garcia-Carrasco, M., Gatica, A., & Zamora, A. 2007. The transcription factor nuclear factor-kappa B and cancer. *Clin Oncol (R Coll Radiol)*, 19(2): 154-161.

Espina, V., Wulfkuhle, J. D., Calvert, V. S., VanMeter, A., Zhou, W., Coukos, G., Geho, D. H., Petricoin, E. F., 3rd, & Liotta, L. A. 2006. Laser-capture microdissection. *Nat Protoc*, 1(2): 586-603.

Esteller, M. 2002. CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. *Oncogene*, 21(35): 5427-5440.

Esteller, M. 2007. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet*, 8(4): 286-298.

Esteve, P. O., Chin, H. G., & Pradhan, S. 2005. Human maintenance DNA (cytosine-5)-methyltransferase and p53 modulate expression of p53-repressed promoters. *Proc Natl Acad Sci U S A*, 102(4): 1000-1005.

Ezhkova, E., Lien, W. H., Stokes, N., Pasolli, H. A., Silva, J. M., & Fuchs, E. 2011. EZH1 and EZH2 cogovern histone H3K27 trimethylation and are essential for hair follicle homeostasis and wound repair. *Genes Dev*, 25(5): 485-498.

Fassina, A., Cappellesso, R., Guzzardo, V., Dalla Via, L., Piccolo, S., Ventura, L., & Fassan, M. 2012. Epithelial-mesenchymal transition in malignant mesothelioma. *Mod Pathol*, 25(1): 86-99.

Fatemi, M., Pao, M. M., Jeong, S., Gal-Yam, E. N., Egger, G., Weisenberger, D. J., & Jones, P. A. 2005. Footprinting of mammalian promoters: use of a CpG DNA methyltransferase revealing nucleosome positions at a single molecule level. *Nucleic Acids Res*, 33(20): e176.

Feinberg, A. P., & Vogelstein, B. 1983. Hypomethylation of ras oncogenes in primary human cancers. *Biochem Biophys Res Commun*, 111(1): 47-54.

Felgner, J., Kreipe, H., Heidorn, K., Jaquet, K., Heuss, R., Zschunke, F., Radzun, H. J., & Parwaresch, M. R. 1992. Lineage-specific methylation of the c-fms gene in blood cells and macrophages. *Leukemia*, 6(5): 420-425.

Flatmark, K., Maelandsmo, G. M., Martinsen, M., Rasmussen, H., & Fodstad, O. 2004. Twelve colorectal cancer cell lines exhibit highly variable growth and metastatic capacities in an orthotopic model in nude mice. *Eur J Cancer*, 40(10): 1593-1598.

Flouriot, G., Pope, C., Kenealy, M. R., & Gannon, F. 1997. Improved efficiency for primer extension by using a long, highly-labeled primer generated from immobilized single-stranded DNA templates. *Nucleic Acids Res*, 25(8): 1658-1659.

Fraga, M. F., & Esteller, M. 2002. DNA methylation: a profile of methods and applications. *Biotechniques*, 33(3): 632, 634, 636-649.

Friedman, R. C., Farh, K. K., Burge, C. B., & Bartel, D. P. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res*, 19(1): 92-105.

Galamb, O., Spisak, S., Sipos, F., Toth, K., Solymosi, N., Wichmann, B., Krenacs, T., Valcz, G., Tulassay, Z., & Molnar, B. 2010. Reversal of gene expression changes in the colorectal normal-adenoma pathway by NS398 selective COX2 inhibitor. *Br J Cancer*, 102(4): 765-773.

Gardiner-Garden, M., & Frommer, M. 1987. CpG islands in vertebrate genomes. *J Mol Biol*, 196(2): 261-282.

Garg, A., & Aggarwal, B. B. 2002. Nuclear transcription factor-kappaB as a target for cancer drug development. *Leukemia*, 16(6): 1053-1068.

Giancotti, V. 2006. Breast cancer markers. *Cancer Lett*, 243(2): 145-159.

Ginos, M. A., Page, G. P., Michalowicz, B. S., Patel, K. J., Volker, S. E., Pambuccian, S. E., Ondrey, F. G., Adams, G. L., & Gaffney, P. M. 2004. Identification of a gene expression

signature associated with recurrent disease in squamous cell carcinoma of the head and neck. *Cancer Res*, 64(1): 55-63.

Grutzmann, R., Pilarsky, C., Ammerpohl, O., Luttges, J., Bohme, A., Sipos, B., Foerder, M., Alldinger, I., Jahnke, B., Schackert, H. K., Kalthoff, H., Kremer, B., Kloppel, G., & Saeger, H. D. 2004. Gene expression profiling of microdissected pancreatic ductal carcinomas using high-density DNA microarrays. *Neoplasia*, 6(5): 611-622.

Guo, J., Cheng, H., Zhao, S., & Yu, L. 2006. GG: a domain involved in phage LTF apparatus and implicated in human MEB and non-syndromic hearing loss diseases. *FEBS Lett*, 580(2): 581-584.

Hanada, M., Delia, D., Aiello, A., Stadtmauer, E., & Reed, J. C. 1993. bcl-2 gene hypomethylation and high-level expression in B-cell chronic lymphocytic leukemia. *Blood*, 82(6): 1820-1828.

Hanahan, D., & Weinberg, R. A. 2000. The hallmarks of cancer. *Cell*, 100(1): 57-70.

Harbeck, N., Nimmrich, I., Hartmann, A., Ross, J. S., Cufer, T., Grutzmann, R., Kristiansen, G., Paradiso, A., Hartmann, O., Margossian, A., Martens, J., Schwope, I., Lukas, A., Muller, V., Milde-Langosch, K., Nahrig, J., Foekens, J., Maier, S., Schmitt, M., & Lesche, R. 2008. Multicenter study using paraffin-embedded tumor tissue testing PITX2 DNA methylation as a marker for outcome prediction in tamoxifen-treated, node-negative breast cancer patients. *J Clin Oncol*, 26(31): 5036-5042.

Hasegawa, H., Senga, T., Ito, S., Iwamoto, T., & Hamaguchi, M. 2009. A role for AP-1 in matrix metalloproteinase production and invadopodia formation of v-Crk-transformed cells. *Exp Cell Res*, 315(8): 1384-1392.

Hashimoto, H., Vertino, P. M., & Cheng, X. 2010. Molecular coupling of DNA methylation and histone methylation. *Epigenomics*, 2(5): 657-669.

Hayslip, J., & Montero, A. 2006. Tumor suppressor gene methylation in follicular lymphoma: a comprehensive review. *Mol Cancer*, 5: 44.

He, Q. Y., Liu, X. H., Li, Q., Studholme, D. J., Li, X. W., & Liang, S. P. 2006. G8: a novel domain associated with polycystic kidney disease and non-syndromic hearing loss. *Bioinformatics*, 22(18): 2189-2191.

Hess, J., Angel, P., & Schorpp-Kistner, M. 2004. AP-1 subunits: quarrel and harmony among siblings. *J Cell Sci*, 117(Pt 25): 5965-5973.

Holliday, R., & Ho, T. 2002. DNA methylation and epigenetic inheritance. *Methods*, 27(2): 179-183.

Honda, T., Tamura, G., Waki, T., Kawata, S., Terashima, M., Nishizuka, S., & Motoyama, T. 2004. Demethylation of MAGE promoters during gastric cancer progression. *Br J Cancer*, 90(4): 838-843.

Hou, J., Aerts, J., den Hamer, B., van Ijcken, W., den Bakker, M., Riegman, P., van der Leest, C., van der Spek, P., Foekens, J. A., Hoogsteden, H. C., Grosveld, F., & Philipsen, S. 2010. Gene expression-based classification of non-small cell lung carcinomas and survival prediction. *PLoS One*, 5(4): e10312.

Huber, M. A., Azoitei, N., Baumann, B., Grunert, S., Sommer, A., Pehamberger, H., Kraut, N., Beug, H., & Wirth, T. 2004. NF-kappaB is essential for epithelial-mesenchymal transition and metastasis in a model of breast cancer progression. *J Clin Invest*, 114(4): 569-581.

Jackson, M. R., Nilsson, T., & Peterson, P. A. 1990. Identification of a consensus motif for retention of transmembrane proteins in the endoplasmic reticulum. *EMBO J*, 9(10): 3153-3162.

Jaenisch, R., & Bird, A. 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet*, 33 Suppl: 245-254.

Jemal, A., Siegel, R., Xu, J., & Ward, E. 2010. Cancer statistics, 2010. *CA Cancer J Clin*, 60(5): 277-300.

Johnson, A. B., Denko, N., & Barton, M. C. 2008. Hypoxia induces a novel signature of chromatin modifications and global repression of transcription. *Mutat Res*, 640(1-2): 174-179.

Jones, P. A., & Baylin, S. B. 2007. The epigenomics of cancer. *Cell*, 128(4): 683-692.

Jonk, L. J., Itoh, S., Heldin, C. H., ten Dijke, P., & Kruijer, W. 1998. Identification and functional characterization of a Smad binding element (SBE) in the JunB promoter that acts as a transforming growth factor-beta, activin, and bone morphogenetic protein-inducible enhancer. *J Biol Chem*, 273(33): 21145-21152.

Jun, H. J., Woolfenden, S., Coven, S., Lane, K., Bronson, R., Housman, D., & Charest, A. 2009. Epigenetic regulation of c-ROS receptor tyrosine kinase expression in malignant gliomas. *Cancer Res*, 69(6): 2180-2184.

Jung, J. E., Lee, H. G., Cho, I. H., Chung, D. H., Yoon, S. H., Yang, Y. M., Lee, J. W., Choi, S., Park, J. W., Ye, S. K., & Chung, M. H. 2005. STAT3 is a potential modulator of HIF-1-mediated VEGF expression in human renal carcinoma cells. *FASEB J*, 19(10): 1296-1298.

Kalluri, R., & Weinberg, R. A. 2009. The basics of epithelial-mesenchymal transition. *J Clin Invest*, 119(6): 1420-1428.

Kang, Q., & Chen, A. 2009. Curcumin inhibits srebp-2 expression in activated hepatic stellate cells in vitro by reducing the activity of specificity protein-1. *Endocrinology*, 150(12): 5384-5394.

Karlic, R., Chung, H. R., Lasserre, J., Vlahovicek, K., & Vingron, M. 2010. Histone modification levels are predictive for gene expression. *Proc Natl Acad Sci U S A*, 107(7): 2926-2931.

Kassi, E., & Moutsatsou, P. 2011. Glucocorticoid receptor signaling and prostate cancer. *Cancer Lett*, 302(1): 1-10.

Kawaji, H., Kasukawa, T., Fukuda, S., Katayama, S., Kai, C., Kawai, J., Carninci, P., & Hayashizaki, Y. 2006. CAGE Basic/Analysis Databases: the CAGE resource for comprehensive promoter analysis. *Nucleic Acids Res*, 34(Database issue): D632-636.

Kikuno, R., Nagase, T., Nakayama, M., Koga, H., Okazaki, N., Nakajima, D., & Ohara, O. 2004. HUGE: a database for human KIAA proteins, a 2004 update integrating HUGEppi and ROUGE. *Nucleic Acids Res*, 32(Database issue): D502-504.

Klose, R. J., Yan, Q., Tothova, Z., Yamane, K., Erdjument-Bromage, H., Tempst, P., Gilliland, D. G., Zhang, Y., & Kaelin, W. G., Jr. 2007. The retinoblastoma binding protein RBP2 is an H3K4 demethylase. *Cell*, 128(5): 889-900.

Kollmann, K., Heller, G., & Sexl, V. 2011. c-JUN prevents methylation of p16(INK4a) (and Cdk6): the villain turned bodyguard. *Oncotarget*, 2(5): 422-427.

Kornberg, R. D. 2007. The molecular basis of eukaryotic transcription. *Proc Natl Acad Sci U S A*, 104(32): 12955-12961.

Kouzarides, T. 2007. Chromatin modifications and their function. *Cell*, 128(4): 693-705.

Kovacs, A. D., Cebers, G., & Liljequist, S. 2000. Kainate receptor-mediated activation of the AP-1 transcription factor complex in cultured rat cerebellar granule cells. *Brain Res Bull*, 52(2): 127-133.

Kvanta, A., Kontny, E., Jondal, M., Okret, S., & Fredholm, B. B. 1992. Mitogen stimulation of T-cells increases c-Fos and c-Jun protein levels, AP-1 binding and AP-1 transcriptional activity. *Cell Signal*, 4(3): 275-286.

Lashkari, D. A., DeRisi, J. L., McCusker, J. H., Namath, A. F., Gentile, C., Hwang, S. Y., Brown, P. O., & Davis, R. W. 1997. Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc Natl Acad Sci U S A*, 94(24): 13057-13062.

Leaner, V. D., Chick, J. F., Donninger, H., Linniola, I., Mendoza, A., Khanna, C., & Birrer, M. J. 2009. Inhibition of AP-1 transcriptional activity blocks the migration, invasion, and experimental metastasis of murine osteosarcoma. *Am J Pathol*, 174(1): 265-275.

Lehembre, F., Yilmaz, M., Wicki, A., Schomber, T., Strittmatter, K., Ziegler, D., Kren, A., Went, P., Derksen, P. W., Berns, A., Jonkers, J., & Christofori, G. 2008. NCAM-induced focal adhesion assembly: a functional switch upon loss of E-cadherin. *EMBO J*, 27(19): 2603-2615.

Lewis, J. D., Meehan, R. R., Henzel, W. J., Maurer-Fogy, I., Jeppesen, P., Klein, F., & Bird, A. 1992. Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell*, 69(6): 905-914.

Li, B., Carey, M., & Workman, J. L. 2007. The role of chromatin during transcription. *Cell*, 128(4): 707-719.

Li, Q., & Chen, H. 2011. Epigenetic modifications of metastasis suppressor genes in colon cancer metastasis. *Epigenetics*, 6(7): 849-852.

Liu, Z., Brattain, M. G., & Appert, H. 1997. Differential display of reticulocalbin in the highly invasive cell line, MDA-MB-435, versus the poorly invasive cell line, MCF-7. *Biochem Biophys Res Commun*, 231(2): 283-289.

Lu, J., Getz, G., Miska, E. A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B. L., Mak, R. H., Ferrando, A. A., Downing, J. R., Jacks, T., Horvitz, H. R., & Golub, T. R. 2005. MicroRNA expression profiles classify human cancers. *Nature*, 435(7043): 834-838.

Maier, H. J., Schmidt-Strassburger, U., Huber, M. A., Wiedemann, E. M., Beug, H., & Wirth, T. 2010. NF-kappaB promotes epithelial-mesenchymal transition, migration and invasion of pancreatic carcinoma cells. *Cancer Lett*, 295(2): 214-228.

Manivel, P., Muthukumaran, J., Kannan, M., & Krishna, R. 2011. Insight into residues involved in the structure and function of the breast cancer associated protein human gamma synuclein. *J Mol Model*, 17(2): 251-263.

Marcucci, G., Baldus, C. D., Ruppert, A. S., Radmacher, M. D., Mrozek, K., Whitman, S. P., Kolitz, J. E., Edwards, C. G., Vardiman, J. W., Powell, B. L., Baer, M. R., Moore, J. O., Perrotti, D., Caligiuri, M. A., Carroll, A. J., Larson, R. A., de la Chapelle, A., & Bloomfield, C. D. 2005. Overexpression of the ETS-related gene, ERG, predicts a worse outcome in acute myeloid leukemia with normal karyotype: a Cancer and Leukemia Group B study. *J Clin Oncol*, 23(36): 9234-9242.

Martin, A., & Cano, A. 2010. Tumorigenesis: Twist1 links EMT to self-renewal. *Nat Cell Biol*, 12(10): 924-925.

Martin, C., & Zhang, Y. 2005. The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol*, 6(11): 838-849.

Matsuzaki, S., Tanaka, F., Mimori, K., Tahara, K., Inoue, H., & Mori, M. 2009. Clinicopathologic significance of KIAA1199 overexpression in human gastric cancer. *Ann Surg Oncol*, 16(7): 2042-2051.

Matteucci, E., Bendinelli, P., & Desiderio, M. A. 2009. Nuclear localization of active HGF receptor Met in aggressive MDA-MB231 breast carcinoma cells. *Carcinogenesis*, 30(6): 937-945.

May, M. J., & Ghosh, S. 1997. Rel/NF-kappa B and I kappa B proteins: an overview. *Semin Cancer Biol*, 8(2): 63-73.

Michishita, E., Garces, G., Barrett, J. C., & Horikawa, I. 2006. Upregulation of the KIAA1199 gene is associated with cellular mortality. *Cancer Lett*, 239(1): 71-77.

Miller, L. D., Smeds, J., George, J., Vega, V. B., Vergara, L., Ploner, A., Pawitan, Y., Hall, P., Klaar, S., Liu, E. T., & Bergh, J. 2005. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci U S A*, 102(38): 13550-13555.

Miyajima, A., Furihata, T., & Chiba, K. 2009. Functional analysis of GC Box and its CpG methylation in the regulation of CYP1A2 gene expression. *Drug Metab Pharmacokinet*, 24(3): 269-276.

Moazed, D. 2009. Small RNAs in transcriptional gene silencing and genome defence. *Nature*, 457(7228): 413-420.

Muckenfuss, H., Kaiser, J. K., Krebil, E., Battenberg, M., Schwer, C., Cichutek, K., Munk, C., & Flory, E. 2007. Sp1 and Sp3 regulate basal transcription of the human APOBEC3G gene. *Nucleic Acids Res*, 35(11): 3784-3796.

Mullenbrock, S., Shah, J., & Cooper, G. M. 2011. Global expression analysis identified a preferentially NGF-induced transcriptional program regulated by sustained MEK/ERK and AP-1 activation during PC12 differentiation. *J Biol Chem*.

Nabel, G. J., & Verma, I. M. 1993. Proposed NF-kappa B/I kappa B family nomenclature. *Genes Dev*, 7(11): 2063.

Nagi, C., Guttman, M., Jaffer, S., Qiao, R., Keren, R., Triana, A., Li, M., Godbold, J., Bleiweiss, I. J., & Hazan, R. B. 2005. N-cadherin expression in breast cancer: correlation with an aggressive histologic variant--invasive micropapillary carcinoma. *Breast Cancer Res Treat*, 94(3): 225-235.

Nair, R. R., Avila, H., Ma, X., Wang, Z., Lennartz, M., Darnay, B. G., Boyd, D. D., & Yan, C. 2008. A novel high-throughput screening system identifies a small molecule repressive for matrix metalloproteinase-9 expression. *Mol Pharmacol*, 73(3): 919-929.

Nan, X., Cross, S., & Bird, A. 1998. Gene silencing by methyl-CpG-binding proteins. *Novartis Found Symp*, 214: 6-16; discussion 16-21, 46-50.

Nishigaki, M., Aoyagi, K., Danjoh, I., Fukaya, M., Yanagihara, K., Sakamoto, H., Yoshida, T., & Sasaki, H. 2005. Discovery of aberrant expression of R-RAS by cancer-linked DNA hypomethylation in gastric cancer using microarrays. *Cancer Res*, 65(6): 2115-2124.

Pan, D., Kocherginsky, M., & Conzen, S. D. 2011. Activation of the glucocorticoid receptor is associated with poor prognosis in estrogen receptor-negative breast cancer. *Cancer Res*, 71(20): 6360-6370.

Pawitan, Y., Bjohle, J., Amler, L., Borg, A. L., Egyhazi, S., Hall, P., Han, X., Holmberg, L., Huang, F., Klaar, S., Liu, E. T., Miller, L., Nordgren, H., Ploner, A., Sandelin, K., Shaw, P. M., Smeds, J., Skoog, L., Wedren, S., & Bergh, J. 2005. Gene expression profiling

spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Res*, 7(6): R953-964.

Petrella, B. L., Lohi, J., & Brinckerhoff, C. E. 2005. Identification of membrane type-1 matrix metalloproteinase as a target of hypoxia-inducible factor-2 alpha in von Hippel-Lindau renal cell carcinoma. *Oncogene*, 24(6): 1043-1052.

Ponticos, M., Harvey, C., Ikeda, T., Abraham, D., & Bou-Gharios, G. 2009. JunB mediates enhancer/promoter activity of COL1A2 following TGF-beta induction. *Nucleic Acids Res*, 37(16): 5378-5389.

Portela, A., & Esteller, M. 2010. Epigenetic modifications and human disease. *Nat Biotechnol*, 28(10): 1057-1068.

Puhka, M., Vihinen, H., Joensuu, M., & Jokitalo, E. 2007. Endoplasmic reticulum remains continuous and undergoes sheet-to-tubule transformation during cell division in mammalian cells. *J Cell Biol*, 179(5): 895-909.

Radvanyi, L., Singh-Sandhu, D., Gallichan, S., Lovitt, C., Pedyczak, A., Mallo, G., Gish, K., Kwok, K., Hanna, W., Zubovits, J., Armes, J., Venter, D., Hakimi, J., Shortreed, J., Donovan, M., Parrington, M., Dunn, P., Oomen, R., Tartaglia, J., & Berinstein, N. L. 2005. The gene associated with trichorhinophalangeal syndrome in humans is overexpressed in breast cancer. *Proc Natl Acad Sci U S A*, 102(31): 11005-11010.

Reed, J. C., Alpers, J. D., Scherle, P. A., Hoover, R. G., Nowell, P. C., & Prystowsky, M. B. 1987. Proto-oncogene expression in cloned T lymphocytes: mitogens and growth factors induce different patterns of expression. *Oncogene*, 1(2): 223-228.

Robertson, K. D. 2005. DNA methylation and human disease. *Nat Rev Genet*, 6(8): 597-610.

Roman-Gomez, J., Jimenez-Velasco, A., Agirre, X., Cervantes, F., Sanchez, J., Garate, L., Barrios, M., Castillejo, J. A., Navarro, G., Colomer, D., Prosper, F., Heiniger, A., & Torres, A. 2005. Promoter hypomethylation of the LINE-1 retrotransposable elements activates sense/antisense transcription and marks the progression of chronic myeloid leukemia. *Oncogene*, 24(48): 7213-7223.

Rupon, J. W., Wang, S. Z., Gaensler, K., Lloyd, J., & Ginder, G. D. 2006. Methyl binding domain protein 2 mediates gamma-globin gene silencing in adult human betaYAC transgenic mice. *Proc Natl Acad Sci U S A*, 103(17): 6617-6622.

Sabates-Bellver, J., Van der Flier, L. G., de Palo, M., Cattaneo, E., Maake, C., Rehrauer, H., Laczko, E., Kurowski, M. A., Bujnicki, J. M., Menigatti, M., Luz, J., Ranalli, T. V., Gomes, V., Pastorelli, A., Faggiani, R., Anti, M., Jiricny, J., Clevers, H., & Marra, G. 2007. Transcriptome profile of human colorectal adenomas. *Mol Cancer Res*, 5(12): 1263-1275.

Sakai, D. D., Helms, S., Carlstedt-Duke, J., Gustafsson, J. A., Rottman, F. M., & Yamamoto, K. R. 1988. Hormone-mediated repression: a negative glucocorticoid response element from the bovine prolactin gene. *Genes Dev*, 2(9): 1144-1154.

Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., & Hume, D. A. 2007. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat Rev Genet*, 8(6): 424-436.

Sariego, J. 2010. Breast cancer in the young patient. *Am Surg*, 76(12): 1397-1400.

Saxonov, S., Berg, P., & Brutlag, D. L. 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A*, 103(5): 1412-1417.

Schmidt, M., Petry, I. B., Bohm, D., Lebrecht, A., von Torne, C., Gebhard, S., Gerhold-Ay, A., Cotarelo, C., Battista, M., Schormann, W., Freis, E., Selinski, S., Ickstadt, K., Rahnenfuhrer, J., Sebastian, M., Schuler, M., Koelbl, H., Gehrmann, M., & Hengstler, J. G. 2011. Ep-CAM RNA expression predicts metastasis-free survival in three cohorts of untreated node-negative breast cancer. *Breast Cancer Res Treat*, 125(3): 637-646.

Schweizer, A., Matter, K., Ketcham, C. M., & Hauri, H. P. 1991. The isolated ER-Golgi intermediate compartment exhibits properties that are different from ER and cis-Golgi. *J Cell Biol*, 113(1): 45-54.

Shimoyama, Y., Hirohashi, S., Hirano, S., Noguchi, M., Shimosato, Y., Takeichi, M., & Abe, O. 1989. Cadherin cell-adhesion molecules in human epithelial tissues and carcinomas. *Cancer Res*, 49(8): 2128-2133.

Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T., Fukuda, S., Sasaki, D., Podhajska, A., Harbers, M., Kawai, J., Carninci, P., & Hayashizaki, Y. 2003. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci U S A*, 100(26): 15776-15781.

Sikora, E., Grassilli, E., Bellesia, E., Troiano, L., & Franceschi, C. 1993. Studies of the relationship between cell proliferation and cell death. III. AP-1 DNA-binding activity during concanavalin A-induced proliferation or dexamethasone-induced apoptosis of rat thymocytes. *Biochem Biophys Res Commun*, 192(2): 386-391.

Skrzypczak, M., Goryca, K., Rubel, T., Paziewska, A., Mikula, M., Jarosz, D., Pachlewski, J., Oledzki, J., & Ostrowski, J. 2010. Modeling oncogenic signaling in colon tumors by multidirectional analyses of microarray data directed for maximization of analytical reliability. *PLoS One*, 5(10).

Smith, I. M., Glazer, C. A., Mithani, S. K., Ochs, M. F., Sun, W., Bhan, S., Vostrov, A., Abdullaev, Z., Lobanenkov, V., Gray, A., Liu, C., Chang, S. S., Ostrow, K. L., Westra, W. H., Begum, S., Dhara, M., & Califano, J. 2009. Coordinated activation of candidate proto-oncogenes and cancer testes antigens via promoter demethylation in head and neck cancer and lung cancer. *PLoS One*, 4(3): e4961.

Sporn, M. B. 1996. The war on cancer. *Lancet*, 347(9012): 1377-1381.

Su, A. I., Cooke, M. P., Ching, K. A., Hakak, Y., Walker, J. R., Wiltshire, T., Orth, A. P., Vega, R. G., Sapinoso, L. M., Moqrich, A., Patapoutian, A., Hampton, G. M., Schultz, P. G., & Hogenesch, J. B. 2002. Large-scale analysis of the human and mouse transcriptomes. *Proc Natl Acad Sci U S A*, 99(7): 4465-4470.

Sunahori, K., Juang, Y. T., & Tsokos, G. C. 2009. Methylation status of CpG islands flanking a cAMP response element motif on the protein phosphatase 2Ac alpha promoter determines CREB binding and activity. *J Immunol*, 182(3): 1500-1508.

Szyf, M., Kaplan, F., Mann, V., Giloh, H., Kedar, E., & Razin, A. 1985. Cell cycle-dependent regulation of eukaryotic DNA methylase level. *J Biol Chem*, 260(15): 8653-8656.

Tabu, K., Sasai, K., Kimura, T., Wang, L., Aoyanagi, E., Kohsaka, S., Tanino, M., Nishihara, H., & Tanaka, S. 2008. Promoter hypomethylation regulates CD133 expression in human gliomas. *Cell Res*, 18(10): 1037-1046.

Takai, D., & Jones, P. A. 2002. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A*, 99(6): 3740-3745.

Tompa, M., Li, N., Bailey, T. L., Church, G. M., De Moor, B., Eskin, E., Favorov, A. V., Frith, M. C., Fu, Y., Kent, W. J., Makeev, V. J., Mironov, A. A., Noble, W. S., Pavesi, G.,

Pesole, G., Regnier, M., Simonis, N., Sinha, S., Thijs, G., van Helden, J., Vandenbogaert, M., Weng, Z., Workman, C., Ye, C., & Zhu, Z. 2005. Assessing computational tools for the discovery of transcription factor binding sites. *Nat Biotechnol*, 23(1): 137-144.

Tost, J., & Gut, I. G. 2007. DNA methylation analysis by pyrosequencing. *Nat Protoc*, 2(9): 2265-2275.

Tu, S., Teng, Y. C., Yuan, C., Wu, Y. T., Chan, M. Y., Cheng, A. N., Lin, P. H., Juan, L. J., & Tsai, M. D. 2008. The ARID domain of the H3K4 demethylase RBP2 binds to a DNA CCGCCC motif. *Nat Struct Mol Biol*, 15(4): 419-421.

Usami, S., Takumi, Y., Suzuki, N., Oguchi, T., Oshima, A., Suzuki, H., Kitoh, R., Abe, S., Sasaki, A., & Matsubara, A. 2008. The localization of proteins encoded by CRYM, KIAA1199, UBA52, COL9A3, and COL9A1, genes highly expressed in the cochlea. *Neuroscience*, 154(1): 22-28.

Vaiopoulos, A. G., Papachroni, K. K., & Papavassiliou, A. G. 2010. Colon carcinogenesis: Learning from NF-kappaB and AP-1. *Int J Biochem Cell Biol*, 42(7): 1061-1065.

Vallabhapurapu, S., & Karin, M. 2009. Regulation and function of NF-kappaB transcription factors in the immune system. *Annu Rev Immunol*, 27: 693-733.

van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R., & Friend, S. H. 2002. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 415(6871): 530-536.

van de Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., Marton, M. J., Parrish, M., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., van der Velde, T., Bartelink, H., Rodenhuis, S., Rutgers, E. T., Friend, S. H., & Bernards, R. 2002. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*, 347(25): 1999-2009.

Wadman, I. A., Osada, H., Grutz, G. G., Agulnick, A. D., Westphal, H., Forster, A., & Rabbitts, T. H. 1997. The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *EMBO J*, 16(11): 3145-3157.

Wang, P., Lin, C., Smith, E. R., Guo, H., Sanderson, B. W., Wu, M., Gogol, M., Alexander, T., Seidel, C., Wiedemann, L. M., Ge, K., Krumlauf, R., & Shilatifard, A. 2009. Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. *Mol Cell Biol*, 29(22): 6074-6085.

Wang, Y., Klijn, J. G., Zhang, Y., Sieuwerts, A. M., Look, M. P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M. E., Yu, J., Jatkoe, T., Berns, E. M., Atkins, D., & Foekens, J. A. 2005. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet*, 365(9460): 671-679.

Watt, F., & Molloy, P. L. 1988. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev*, 2(9): 1136-1143.

Watt, P. M., Kumar, R., & Kees, U. R. 2000. Promoter demethylation accompanies reactivation of the HOX11 proto-oncogene in leukemia. *Genes Chromosomes Cancer*, 29(4): 371-377.

Wiklund, E. D., Kjems, J., & Clark, S. J. 2010. Epigenetic architecture and miRNA: reciprocal regulators. *Epigenomics*, 2(6): 823-840.

Yan, C., & Boyd, D. D. 2006. Histone H3 acetylation and H3 K4 methylation define distinct chromatin regions permissive for transgene expression. *Mol Cell Biol*, 26(17): 6357-6371.

Yan, C., Wang, H., Aggarwal, B., & Boyd, D. D. 2004. A novel homologous recombination system to study 92 kDa type IV collagenase transcription demonstrates that the NF-kappaB motif drives the transition from a repressed to an activated state of gene expression. *FASEB J*, 18(3): 540-541.

Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-Even, A., Horn-Saban, S., Safran, M., Domany, E., Lancet, D., & Shmueli, O. 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics*, 21(5): 650-659.

Yang, C., Bolotin, E., Jiang, T., Sladek, F. M., & Martinez, E. 2007. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene*, 389(1): 52-65.

Yang, J., Ledaki, I., Turley, H., Gatter, K. C., Montero, J. C., Li, J. L., & Harris, A. L. 2009. Role of hypoxia-inducible factors in epigenetic regulation via histone demethylases. *Ann N Y Acad Sci*, 1177: 185-197.

Yang, M. H., Hsu, D. S., Wang, H. W., Wang, H. J., Lan, H. Y., Yang, W. H., Huang, C. H., Kao, S. Y., Tzeng, C. H., Tai, S. K., Chang, S. Y., Lee, O. K., & Wu, K. J. 2010. Bmi1 is essential in Twist1-induced epithelial-mesenchymal transition. *Nat Cell Biol*, 12(10): 982-992.

Yu, M., Sato, H., Seiki, M., & Thompson, E. W. 1995. Complex regulation of membrane-type matrix metalloproteinase expression and matrix metalloproteinase-2 activation by concanavalin A in MDA-MB-231 human breast cancer cells. *Cancer Res*, 55(15): 3272-3277.

Zecchini, S., Bombardelli, L., Decio, A., Bianchi, M., Mazzarol, G., Sanguineti, F., Aletti, G., Maddaluno, L., Berezin, V., Bock, E., Casadio, C., Viale, G., Colombo, N., Giavazzi, R., & Cavallaro, U. 2011. The adhesion molecule NCAM promotes ovarian cancer progression via FGFR signalling. *EMBO Mol Med*, 3(8): 480-494.

Zeng, M., Kikuchi, H., Pino, M. S., & Chung, D. C. 2010. Hypoxia activates the K-ras proto-oncogene to stimulate angiogenesis and inhibit apoptosis in colon cancer cells. *PLoS One*, 5(6): e10966.

Zhou, H., Chen, W. D., Qin, X., Lee, K., Liu, L., Markowitz, S. D., & Gerson, S. L. 2001. MMTV promoter hypomethylation is linked to spontaneous and MNU associated c-neu expression and mammary carcinogenesis in MMTV c-neu transgenic mice. *Oncogene*, 20(42): 6009-6017.

Zhou, X., Sun, H., Chen, H., Zavadil, J., Kluz, T., Arita, A., & Costa, M. 2010. Hypoxia induces trimethylated H3 lysine 4 by inhibition of JARID1A demethylase. *Cancer Res*, 70(10): 4214-4221.

Zhou, Y., Yau, C., Gray, J. W., Chew, K., Dairkee, S. H., Moore, D. H., Eppenberger, U., Eppenberger-Castori, S., & Benz, C. C. 2007. Enhanced NF kappa B and AP-1 transcriptional activity associated with antiestrogen resistant breast cancer. *BMC Cancer*, 7: 59.

Zucker, S., Drews, M., Conner, C., Foda, H. D., DeClerck, Y. A., Langley, K. E., Bahou, W. F., Docherty, A. J., & Cao, J. 1998. Tissue inhibitor of metalloproteinase-2 (TIMP-2) binds

to the catalytic domain of the cell surface receptor, membrane type 1-matrix metalloproteinase 1 (MT1-MMP). *J Biol Chem*, 273(2): 1216-1222.

Zucker, S., Hymowitz, M., Conner, C. E., DiYanni, E. A., & Cao, J. 2002. Rapid trafficking of membrane type 1-matrix metalloproteinase to the cell surface regulates progelatinase a activation. *Lab Invest*, 82(12): 1673-1684.