Investigations into expression and regulation of Cystatin A and its implication in breast cancer

A thesis

submitted for partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

By

DIXCY JABA SHEEBA J M

Department of Biosciences and Bioengineering

Indian Institute of Technology Guwahati

Guwahati, Assam-781039, India

May 2020

DEPARTMENT OF BIOSCIENCES AND BIOENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
GUWAHATI, ASSAM-781039

DECLARATION

The thesis titled **"Investigations into expression and regulation of Cystatin A and its implication in breast cancer,"** is a presentation of my original research work carried out in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, India, under the supervision of Dr. Anil Mukund Limaye.

Sincere efforts have been made to acknowledge contributions from other investigators that helped in conceptualizing and execution of the research work. Those who have provided suggestions and technical help have been duly acknowledged. All the research articles and resources used have been cited in the reference section.

21st May 2020

Dixcy Jaba Sheeba J M
Roll No. 146106007
Department of Biosciences and Bioengineering
Indian Institute of Technology Guwahati

DEPARTMENT OF BIOSCIENCES AND BIOENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
GUWAHATI, ASSAM-781039

CERTIFICATE

This is to certify that the work described in the thesis titled, **"Investigations into expression and regulation of Cystatin A and its implication in breast cancer,"** submitted by Dixcy Jaba Sheeba J M (Roll No. 146106007) to Indian Institute of Technology Guwahati, India, for the award of the degree of Doctor of Philosophy is an authentic record of the research work carried out under my supervision in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, India.

This thesis or any part thereof has not been submitted elsewhere for the award of any other degree or diploma.

21st May 2020

Dr. Anil Mukund Limaye
Thesis Supervisor

Acknowledgements

"He hath made everything beautiful in His time"

First and foremost, I thank Almighty for walking through with me in all the steps of my life and for enabling me to accomplish this.

With deepest respect and esteem regards, I owe my profound gratitude to my research Supervisor, Dr. Anil Mukund Limaye, for providing me an opportunity to place my first step towards my dream of accomplishing a remarkable thing in the field of cancer biology. I consider it an honor to work under his invaluable guidance. Without his persistent guidance and encouragement, this thesis would not have seen the light of the day. Our scientific discussions always energized me to pursue further in tough times. Both by his words and action, he taught us, "*To aim high and to stay grounded,*" which I would like to carry forward in my career. Thank you, Sir.

I would like to express my sincere thanks to the Doctoral Committee members, Prof. Sanjukta Patra, Dr. Sachin Kumar, Prof. Debasis Manna, for their constructive suggestions and comments. I acknowledge infrastructural support from Department of Biosciences and Bioengineering and DBT funded Bioinformatics Facility (BIF) at IIT Guwahati. I thank the past and present HODs of Department of Biosciences and Bioengineering, Prof. Arun Goyal, Prof. Veeranki Venkata Dasu, Prof. Kannan Pakshirajan, and Prof. Latha Rangan. I'd like to acknowledge the financial assistance from Department of Biotechnology, Government of India, and scholarship from IIT Guwahati. My sincere thanks to Dr. Senthil Kumar, Dr. Sachin Kumar, Dr. Manish Kumar, Dr. Rajkumar Thummer, Dr. Kusum Kumari Singh, for providing access to their lab facilities.

I would like to thank Dr. Prathibha Ranganathan for her suggestions in the stable cell line generation. I truly appreciate the contribution of the TCGA team for their incredible effort in establishing a database of immeasurable worth.

I would like to express my deepest gratitude to my senior, Dr. Mohan, for his constant support, care, and for helping me to make the right decisions in crucial moments of my research. I especially thank him for teaching me the cell culture, molecular techniques and *in silico* work in my initial years of Ph.D. I am particularly grateful for the assistance provided by Manohar and Girija, which was a milestone in the completion of this thesis. I would like to thank Ajay for his suggestions and guidance in western blotting and ChIP experiments. It is a great pleasure to thank our Molecular Endocrinology Lab members, Ajay, Snigdha, Gaurav, Uttariya, Sonia, Ankita, Musfica, Juana, Akshita for their constructive comments, discussions, for all the good moments we had together and for the constant encouragement during tough times. I would like to acknowledge Shalini for giving up the opportunity of working under the sole guidance of our Sir in MTP, due to which I got this

privilege. I extend my sincere thanks to all the past Molecular Endocrinology lab members, especially Marine, Dr. Gauri, Sukhdev, Swati, Paridhi, Sarbajeet, Mrinal, Ramya for providing a friendly research environment.

My special thanks to Ma'am. Prajaktha for her concern, encouragement, and for giving us memorable lab gatherings.

I owe my deepest gratitude to Sis. Christy and Sir. Daniel, for their unconditional love, prayers, and for taking care of me as their daughter. I would like to pay my special regards to Sir. Haokip and Ma'am. Haokip for spiritual guidance, prayer support, and for the extraordinary family bonding.

I thank my friends Priya, Malathi, Neha, Sajitha, Ashutosh, Ganesh, Vimal, Vaishu, Gaja, Subbi, Ruchira for their cordial support, and for being a part of my good and tough times.I cannot find words to express my heartfelt gratitude towards my EU family, Sharmila, Darilang, Them, Aurelia, Churchill, Philip, Onisha, Jumrick, Wanrisa, Mercy, Iban, Tympang, Yagom, Archana, Ann, Anjali and all sisters and brothers of EU family for their prayer support and giving me blessed times of fellowship. My sincere thanks to Pastors and members of Kanchipuram, Chennai, and Guwahati Church for their prayer support.

I wish to express my deepest gratitude to Gayathri for believing in me even when I ceased to believe in myself.

I am indebted to Amma, Appa, Dicson, Dhasammaal Paati, Devadoss Thatha, Joyce paati, Chithi, Vini, Anni for being the reason for me reaching this far in my life. This thesis would not have been possible without their fervent prayers and unwavering support. I'm richly blessed to have them in my life.

Finally, yet importantly, in this most valuable book, I would like to place on record my gratitude to Amma and Appa for the freedom they have given me in decision making because of which I have seen a day of writing acknowledgment section of a Ph.D. thesis.

Table of contents

List of figures

List of tables

List of abbreviations

ACPI	Acid cysteine protease inhibitor
ADAM	A Disintegrin and Metalloproteinase
ADAMTS	A Disintegrin and Metalloproteinase with Thrombospondin motifs
AF	Activation function
AJCC	American Joint Committee on Cancer
ANOVA	Analysis of variance
AP	Activator protein
ATF	Activating transcription factor
BAM	Binary version of sequence alignment map
BRCA	Breast Invasive Carcinoma
BSA	Bovine serum albumin
cAMP	Cyclic adenosine monophosphate
cDNA	Complementary DNA
ChIP	Chromatin immunoprecipitation
ChIP-Seq	Chromatin immunoprecipitation sequencing
CI	Confidence interval
CRE	cyclic AMP response element
CRES	Cystatin-related epididymal spermatogenic
CRP	Cystatin-related protein
csFBS	Charcoal stripped FBS
CSTA	Cystatin A
CycA	Cyclophilin A
D	Dexamethasone
DBD	DNA-binding domain
DCIS	Ductal carcinoma *in situ*
DMFS	Distant metastasis-free survival
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DNMT	DNA methyltransferase
DPBS	Dulbecco's phosphate-buffered saline
DTT	Dithiothreitol
E1	Estrone
E2	17β-estradiol
E3	Estriol
ECL	Enhanced chemiluminescence
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
ELISA	Enzyme-linked immunosorbent assay
EMC	Expression-methylation correlation
EMT	Epithelial-mesenchymal transition
ENCODE	Encyclopedia of DNA Elements

ER	Estrogen receptor
ERE	Estrogen response elements
ERK	Extracellular signal-regulated kinase
FBS	Fetal bovine serum
Ful	Fulvestrant
gDNA	Genomic DNA
GPER	G-protein coupled estrogen receptor
GPR30	G-protein coupled receptor 30
GSK 3β	Glycogen synthase kinase 3β
HB-EGF	Heparin-binding EGF-like growth factor
HCl	Hydrochloride
HDAC	Histone deacetylase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HER2	Human epidermal growth factor receptor 2
HRG	Histidine-rich glycoprotein
HRP	Horseradish peroxidase
HSD	Honest significant difference
Hsp	Heat shock protein
IARC	International Agency for Research on Cancer
IDC	Invasive ductal carcinoma
IGF-1	Insulin-like growth factor-1
IgG	Immunoglobulin G
IHC	Immunohistochemistry
IUPHAR	International Union of Basic and Clinical Pharmacology
JNK	c-Jun N-terminal kinase
KCl	Potassium chloride
kDa	Kilo Dalton
KLH	Keyhole Limpet Haemocyanin
LBD	Ligand-binding domain
MACS	Model-based analysis of ChIP-Seq
MAPK	Mitogen-activated protein kinases
MCBPs	Methyl-CpG binding proteins
MEF2C	Myocyte-specific enhancer factor 2C
MEK1	MAPK/ERK kinase 1
MEKK	MEK kinase
MET	Mesenchymal-epithelial transition
MKK	MAP kinase kinase
mL	Milliliter
MMP	Matrix metalloproteinase
mRNA	Messenger ribonucleic acid
mTOR	Mammalian target of rapamycin
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide
NaCl	Sodium chloride
NHG	Nottingham histologic grade
OS	Overall survival
P4	Progesterone
PAGE	Polyacrylamide gel electrophoresis
PAI	Plasminogen activator inhibitor

PBS	Phosphate-buffered saline
PBST	PBS containing 0.05% Tween 20
PCR	Polymerase chain reaction
PI3K	Phosphoinositide 3-kinase
pmol	Picomole
PPT	Propyl pyrazole triol
PR	Progesterone receptor
qRT-PCR	Quantitative reverse transcription PCR
RFS	Relapse-free survival
RNA	Ribonucleic acid
RNase	Ribonuclease
RPKM	Reads per kilobase million
RPMI-1640	Roswell Park Memorial Institute-1640 medium
RT-PCR	Reverse transcription PCR
RUNX	Runt-related transcription factor
SAM	Sequence alignment map
SCID	Severe combined immunodeficient
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SERDs	Selective estrogen receptors down-regulators
SERMs	Selective estrogen receptor modulators
siRNA	Short interfering ribonucleic acid
Sp1	Specificity protein 1
SRA	Sequence Read Archival
Sulfo-SMCC	Sulfosuccinimidyl 4-(N-maleimidomethyl) cyclohexane-1-carboxylate
Tam	4-hydroxytamoxifen
TBST	Tris-buffered saline containing 0.05% Tween 20
TCF-8	Transcription factor-8
TCGA	The Cancer Genome Atlas
TDLU	Terminal ductal lobular units
TGFβ	Transforming growth factor β
TIMPs	Tissue inhibitors of metalloproteinases
T_a	Annealing temperature
T_m	Melting temperature
TMB	3,3',5,5'-tetramethylbenzidine
TNBC	Triple-negative breast cancer
TNM	Tumor-node-metastasis
TPA	12- O-tetradecanoylphorbol-13-acetate
tPA	Tissue plasminogen activator
TRE	TPA responsive elements
UCSC	University of California, Santa Cruz
UICC	International Union Against Cancer
uPA	Urokinase-type plasminogen activator
WHO	World Health Organization
ZO-1	Zonula occludens-1
µg	Microgram
µL	Microliter
µm	Micrometer

μM	Micromolar
5-aza	5-azacytidine

CHAPTER 1

Introduction

1.1. Introduction

Metastasis is a major hurdle in successful treatment of solid tumors, including those of the breast. Tumor cells orchestrate a battery of proteolytic enzymes to remodel the matrix, escape from the tumor environment, and metastasize[1,2]. Apart from the matrix metalloproteinases (MMPs), there is increasing evidence on the association of overt or mislocalized expression of cysteine cathepsins with tumorigenesis and metastasis[3,4]. The presence of cysteine in their active site distinguishes them from other lysosomal serine- or aspartic-cathepsins[5,6]. Cysteine cathepsins hydrolyze proteins within and outside the cell[7]. Cathepsins are involved in various cellular processes, such as protein turnover, antigen presentation, apoptosis, and matrix remodeling[6-9].

In physiological conditions, activities of cathepsins are controlled by endogenous small protein inhibitors called cystatins. Cathepsins and cystatins are aberrantly expressed in a wide variety of tumors[10]. Altered expression of cysteine cathepsins or cystatins can tilt the homeostatic balance to favor extracellular matrix (ECM) remodeling, thereby promoting disease progression, tumor invasion and metastasis[11-13].

Cystatin A (CSTA, also known as Stefin A), is a member of cystatin superfamily. CSTA inhibits the activity of cathepsins B, H and L[14]. These cathepsins have been implicated in tumor development, progression and overall survival[10,15,16]. It is abundantly expressed in the

myoepithelial cells and plays a critical role in their ability to resist the malignant transformation of the luminal epithelium[17,18]. Inverse correlation between levels of cathepsin B and CSTA is reported in various tumors indicating the potential tumor suppressor role of CSTA[19,20].

Estrogens are the steroid hormones that regulate the growth, differentiation, and function of target tissues of the reproductive, central nervous, skeletal, cardiovascular and immune systems[21-26]. Estrogen exerts its action on target tissues by regulating the transcriptional activity of various genes. Vendrell and co-workers performed cDNA mini-arrays analysis to detect novel estrogen-regulated genes in MCF-7 derived cell line, in which CSTA was reported as one of the estrogen suppressed genes[27]. However, the precise mechanism of estrogen-mediated regulation of CSTA in breast cancer cells is not understood. Due to the paucity of literature on CSTA regulation in breast cancer, this purported tumor suppressor requires further attention. Investigations presented in this thesis were carried out to fill the lacunae in the understanding of CSTA in terms of hormonal regulation and its role in breast cancer.

1.2. Aim and scope of the work

About two-thirds of breast tumors are estrogen receptor (ER)- or progesterone receptor (PR)-positive at the time of diagnosis[28]. Prognosis and therapeutic decisions are based on the status of ER, PR and human epidermal growth factor receptor 2 (HER2) expression[29]. Being the second most common endocrine-related cancer, a better understanding of molecular players in tumor suppression or promotion is required for the identification of novel prognostic or diagnostic markers in breast cancer. Attempts have been made earlier to correlate CSTA expression with the known histopathological markers, and the results from several clinical studies have led to contradictory inferences about the prognostic value of CSTA[30-32]. This necessitates an independent study on its prognostic potential by taking into consideration the various molecular subtypes of breast tumors.

Deregulated estrogen–ER signaling is linked with breast cancer development and progression[21,22]. Targeting ER signaling through endocrine therapy has been a preferred therapeutic option for estrogen-dependent tumors for decades[33]. Given the critical role of CSTA in regulating the cysteine proteases in the metastatic cascade, identifying the precise mechanisms by which cystatin itself is regulated will further the understanding of molecular mechanisms involved in metastasis. Hence, in the context of breast cancer, it is desirable to obtain mechanistic insights into the estrogen regulation of CSTA expression.

Loss of CSTA expression in breast tumors evidently shifts the balance in favor of cysteine cathepsins, thereby promoting ECM remodeling, tumor invasion, and metastasis[34,35]. However, the underlying mechanism behind the loss of CSTA expression in breast tumors is not known. Epigenetic silencing of tumor suppressor genes is a common phenomenon associated with tumor initiation and progression[36,37]. Interestingly, members of the cystatin superfamily are reported to be epigenetically silenced by DNA methylation in breast, pancreatic, brain, and lung cancer[14]. Recently, Stone and co-workers have found that DNA hypermethylation of estrogen-responsive enhancers is associated with reduced ESR1 binding[38]. The present study examined the impact of DNA methylation on CSTA expression and regulation.

Inhibition of cathepsin B was initially thought to be the sole function of CSTA[39]. Subsequently, other functions such as cell-cell adhesion and apoptosis were attributed to CSTA[40,41]. Moreover, *in vivo* and *in vitro* studies revealed that CSTA affects tumor growth, angiogenesis, invasion, and metastasis in lung and esophageal cancer[42,43]. However, its precise role in breast cancer is not yet elucidated. This study addresses the functional role of CSTA in terms of proliferation, invasion and migration of breast cancer cells.

1.3. Objectives

The work embodied in the thesis is based on the following objectives,

1. To analyze the expression of CSTA in primary breast tumors.
2. To investigate the mechanism of estrogen-mediated regulation of CSTA expression in breast cancer cells.
3. To analyze the role of methylation in the expression and regulation of CSTA.
4. To assess the role of CSTA in proliferation, migration and invasion of breast cancer cells.

CHAPTER 2

Review of literature

2.1. Breast cancer

Breast cancer is the most common cancer among women worldwide, accounting for 24.2% of all female cancers. It is the leading cause of death due to cancer among women, followed by lung cancer[44] (Figure 2.1). International Agency for Research on Cancer (IARC) estimated 2.1 million new breast cancer cases for the year 2018. This signifies that one in every four women diagnosed with cancer would be of breast cancer[45]. In the Indian scenario, breast cancer accounts for 27.7% of total cancer cases detected among women[44]. The age-adjusted rate of breast cancer incidence in India is 25.8 per 100,000 women, with a mortality of 12.7 per 100,000 women[46]. Breast cancer is the most common cancer in urban women[47]. It is the second most prevalent cancer in the case of rural women, next to cervical cancer[47]. In the western population, the majority of breast cancer patients are postmenopausal women, and the peak age of the incidence is between 60 and 70 years[47-49]. Whereas, in India, breast cancer patients are found a decade younger than breast cancer patients of western countries. The peak age of incidence in India is between 45 and 50 years[48] suggesting that breast cancer occurs in premenopausal age among Indian women. Particularly in Assam, among women, breast cancer accounts for 16% of total cancer cases, followed by cancer of the gall bladder (13.6%) and cervix (12.8%)[50]. In the Kamrup district of Assam, from 2003 to 2014, the breast cancer incidence has been significantly increased with an age-adjusted incidence rate of 27.1

per 100,000 women[51]. Notably, 54.6% of total breast cancer cases in the Kamrup district belong to the 30-49 age group[52]. Though uncommon, breast cancer also affects males contributing 1% of total breast cancer cases[53]. Thus, breast cancer is a major global health concern (Figure 2.2).

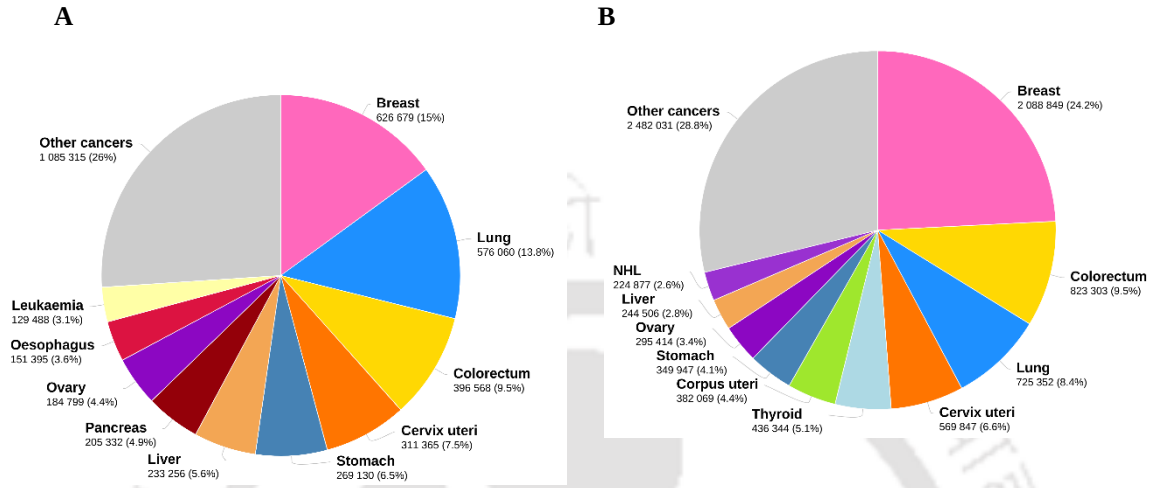

Figure 2.1. Cancer incidence and mortality among women worldwide. Pie charts representing the estimated number of cancer cases among women of all ages (total: 8,622,539) (A), and estimated number of cancer-related deaths among women of all ages (total: 4,169,387) (B). Area of each segment in pie charts indicates the proportion of cancer cases (A) or deaths (B) for each category. Data obtained from GLOBOCAN 2018 (IARC, WHO).

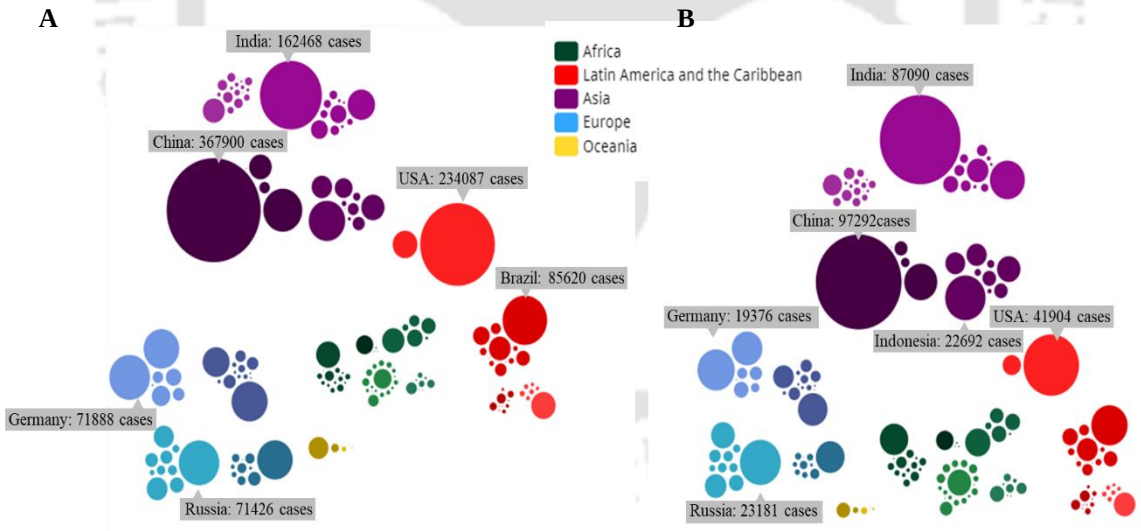

Figure 2.2. Breast cancer incidence and mortality among women worldwide. Circular packing charts representing the estimated number of breast cancer cases among women of all ages (Left panel) and estimated number of breast cancer-related deaths among women of all ages (Right panel) worldwide. Each circle represents a country and the countries are represented in the different shades of five major colors based on the five continents. Circles representing the top six countries in breast cancer incidence and mortality are labeled with the number indicating total breast cancer cases and deaths, respectively. The area of circles indicates the proportion of breast cancer cases (A) or deaths (B) in each country. Data obtained from GLOBOCAN 2018 (IARC, WHO).

2.2. Classification

Breast cancer represents a group of heterogeneous diseases that originates from the terminal ductal lobular units (TDLU) of the breast with varying pathological features. It is characterized by the neoplastic transformation and uncontrolled proliferation of cells within the breast tissue.

Breast cancer is broadly classified, either based on histological and/or cytological appearance as lobular and ductal carcinoma, or based on invasiveness as invasive and *in situ* carcinoma. However, in 2003, the World Health Organization (WHO) defined invasive carcinomas, precursor lesions, and benign epithelial proliferations into 25 major subtypes and 32 minor subtypes[54]. Though the histopathological classification has more prognostic value, molecular markers are useful in therapeutic decisions and prediction of response to targeted therapies. Over the past 19 years, expression of various molecular markers such as ER, PR and HER2 were extensively studied in breast tumors[55-57]. Based on their expression, breast tumors are classified into five major molecular subtypes, namely luminal A, luminal B, HER2+, basal-like, and normal-like groups[57,58]. Luminal A and luminal B are ER and/or PR-positive. HER2+ are ER and/or PR-negative with poor prognosis[57,59]. The primary characteristic that distinguishes luminal B from luminal A is the high proliferation index (Ki-67 staining) and poor prognosis. Basal-like subtype lacks the expression of ER, PR, or HER2. In 2011 the expert panel of St. Gallen International Breast Cancer Conference redefined the intrinsic molecular subtypes of breast tumors. Clinicopathologic definition of molecular subtypes of breast tumors is summarized in Table 2.1[60].

Table 2.1. Clinicopathologic definition of molecular subtypes of breast tumors.

Molecular subtype	Clinicopathologic definition
Luminal A	ER and/or PR positive, HER2 negative, Ki-67^{low}
Luminal B (HER2-negative)	ER and/or PR positive, HER2 negative, Ki-67^{high}
Luminal B (HER2-positive)	ER and/or PR positive, HER2 over-expressed or amplified, Ki-67 $^{high/low}$
HER2-positive	ER and PR negative, HER2 over-expressed or amplified
Basal-like	ER and PR negative, HER2 negative

Cancer staging and grading systems are also used in the classification of the tumors to evaluate the clinical spread of malignancies and to establish suitable therapies. In 1940s, Pierre Denoix developed tumor-node-metastasis (TNM) staging system for the classification of malignant tumors. The first clinical-stage classification of cancers of breast and larynx was published in 1958[61,62]. Later, American Joint Committee on Cancer (AJCC), in collaboration

with International Union Against Cancer (UICC), defined staging criteria for various anatomical sites[63]. The internationally accepted TNM system describes the extent of cancer by defining- i) tumor size at the primary site (T), ii) the extent of lymph node metastases (N), and iii) presence or absence of distant metastases (M) of the tumors[62]. T, N, and M are determined separately and are used to classify the tumors into four stages (I-IV). T is represented as T1-T4 based on increasing size and spread of the primary tumor. N and M are represented as 0 or 1 (0-absence of tumor, 1-presence of tumor)[62]. In the past 60 years, the TNM system has evolved gradually due to the advances in diagnosis and treatment. Moreover, non-anatomic prognostic factors are also being explored to be incorporated in the TNM system to provide the most specific treatment to the patients[62].

The histological grading system is used to classify malignancies based on the degree of differentiation. Low-grade tumors (Grade 1) are well-differentiated, less aggressive with favorable prognosis. Intermediate grade tumors (Grade 2) are moderately differentiated, and high-grade tumors (Grade 3) are poorly differentiated, fast-growing with poor prognosis[64]. Various scoring systems were developed for grading breast cancer. Among those, Nottingham Histologic grade (NHG), established by Elston and Ellis, is a widely accepted system for grading breast tumors. NHG scores tumors from 1 to 3 for each of the three morphological features: tubule formation, nuclear pleomorphism, and mitotic activity. The totality of three scores makes the Nottingham Score. Grade 1 is assigned for tumors with a score of 3 to 5. Grade 2 for the tumors with a score of 6 to 7 and Grade 3 for tumors with a score of 8-9[65].

2.3. Risk factors

Major risk factors for breast cancer include age, genetic make-up, race, obesity, alcohol consumption, and smoking. Other factors such as nulliparity, early puberty, late menopause, and late pregnancy, were also found to be associated with the incidence of breast cancer[66]. For decades, several clinical, epidemiological and experimental studies have provided substantial evidence that prolonged exposure to endogenous estrogens influences the risk of breast cancer[22,67-71].

2.4. Treatment

The main options for breast cancer treatment are surgery, radiotherapy, chemotherapy, endocrine therapy, and targeted therapy. In the past half-century, significant advancement has

occurred in breast cancer therapeutics. Radical mastectomy, a surgical method of treating breast cancer, was substituted by modified radical mastectomy and eventually by breast-conserving surgery[72]. Radiotherapy involves the usage of high-energy X-rays or gamma rays to eradicate subclinical disease after surgical removal[73]. In targeted therapy, only cancer cells are targeted without affecting normal cells. This is achieved by targeting specific overexpressing markers such as HER2. In 1998, advancement in the treatment of HER-2/neu overexpressing tumors occurred when the clinical trial with monoclonal anti-HER2 antibody (trastuzumab) improved the survival of HER2-positive patients[74]. Since targeted or endocrine therapy is ineffective in triple-negative breast cancer (TNBC) patients, cytotoxic chemotherapy is the only effective option[75]. Chemotherapy regimens comprising of anthracyclines and taxanes reduce mortality by about one-third[76].

Endocrine therapy has its origin back in 1895 when oophorectomy performed by Beatson in premenopausal breast cancer patients resulted in the complete remission of breast tumors, and extended survival[77]. This foundation led to the discovery of ovarian hormone, estrogen in 1923 by Edgar Allen and Edward Doisy[78]. A series of studies in the subsequent decades on estrogen signaling unveiled the potential of endocrine therapy as a treatment option[27,79-86]. Endocrine therapy has gained more attention over the course of time, as 60-75% of breast tumors are positive for ER or PR, making it a preferred treatment option for non-TNBC cases.

Endocrine therapy essentially involves the manipulation of estrogen signaling pathway using strategies like i) blocking estrogen synthesis by using aromatase inhibitor (AI) such as anastrozole or letrozole[87], ii) blocking the binding of estrogen to ER with selective estrogen receptor modulators (SERM) such as tamoxifen or raloxifene[88,89], iii) reducing the level of functional ER with selective estrogen receptor down-regulators (SERD) such as fulvestrant[90](Figure 2.3).

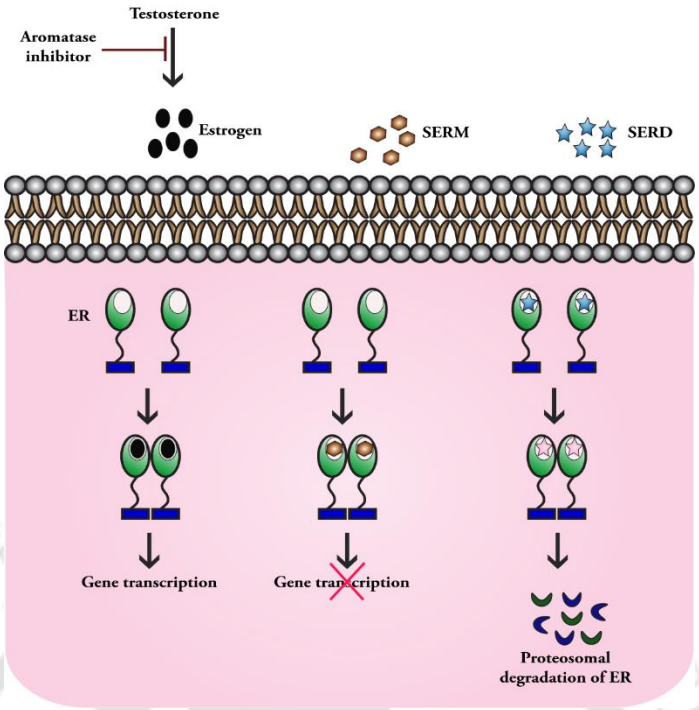

Figure 2.3. Effect of aromatase inhibitors, SERM and SERD on ER signaling pathway. Aromatase inhibitors prevent the synthesis of estrogen. SERM blocks the binding of estrogen to ER. SERD accelerates ER to proteasomal degradation (*recreated from Patel and co-workers, 2018*[91]).

2.5. Estrogen

Estrogens are the steroid hormones primarily involved in the development of primary and secondary sex characteristics in women. Besides, they also play a vital role in maintaining physiological functions of central nervous, cardiovascular, immune, and musculoskeletal systems[21-26]. The three major forms of physiological estrogens in humans are estrone (E1), estradiol (E2, 17β-estradiol), and estriol (E3). E2, the most predominant and bioactive form of estrogen in premenopausal women, is primarily synthesized by the ovary, corpus luteum, and placenta. In postmenopausal women, E1 is the predominant form and is synthesized in extragonadal tissues. E3 is the least potent form of estrogen, and it is synthesized by the placenta during pregnancy[82]. Estrogen synthesized by the ovary is released in the bloodstream and acts as an endocrine factor. On the other hand, extragonadal estrogens act locally in paracrine or intracrine manner at the site of synthesis. Localized estrogen synthesis is vital for tissue-specific function[82]. Unfortunately, estrogens induce carcinogenesis by three mechanisms, i) receptor-mediated stimulation of proliferation, ii) induction of aneuploidy, and iii) increasing mutation rates via cytochrome P450-mediated metabolic activation[83].

2.6. Estrogen receptors

The physiological and pathological functions of estrogens are mediated by estrogen receptors (ERs). In the 1950s, Elwood V. Jensen made a significant progress in the field of hormone action by reporting the evidence for the existence of a receptor for estrogen[86,92]. In 1986, Green and co-workers cloned ER cDNA for the first time from MCF-7 breast cancer cells[93]. Till 1996, ER was considered as the sole signaling receptor of estrogens. However, Kuiper and co-workers cloned another ER in 1996[94]. Thereafter, the ER, which was discovered earlier was referred as ERα, and latter one as ERβ. Interestingly, in 1967, Szego and Davis reported that acute exposure to E2 increases cAMP production in the ovariectomized rat[95]. Later in 1977, Szego and Pietras reported the binding of estrogen to the plasma membrane of endometrial and liver cells[96]. In the 1990s, several groups cloned a gene whose protein product has homology to G-protein coupled receptor superfamily. It was considered as an orphan receptor since its ligand was not known that time[97-100]. In the early 2000s, this orphan G-protein coupled receptor, GPR30, was suggested to mediate the non-genomic response in estrogen signaling[101]. Filardo and co-workers demonstrated that E2 activates extracellular signal-regulated kinase 1 (ERK1) and ERK2 in the cells, which expresses GPR30 but not ERs[102]. Further studies resulted in the establishment of GPR30 as a bona fide ER. Later in 2007, GPR30 was officially named as G-protein coupled estrogen receptor (GPER) by International Union of Basic and Clinical Pharmacology (IUPHAR)[103]. The binding affinity of E2 towards GPER (K_d -3.3 nM) is lower than the binding affinity of E2 towards ERs (K_d - 0.13 to 0.6 nM)[94,104,105].

ERs are members of the nuclear receptor superfamily. ERα and ERβ share common structural architecture but differ in the transcriptional targets, tissue distribution, and functional effects. ERs have five distinct domains (Figure 2.4). N-terminal A/B domain contains ligand-independent activation function 1 (AF1), which is involved in the transcriptional activity by interacting with co-activators. C domain is the DNA-binding domain (DBD), which enables sequence-specific binding of ER to the cognate response elements. The D domain encompasses hinge region, which is involved in nuclear localization. The C-terminal E/F domain serves as a ligand-binding domain (LBD) and also contains activation function 2 (AF2)[106].

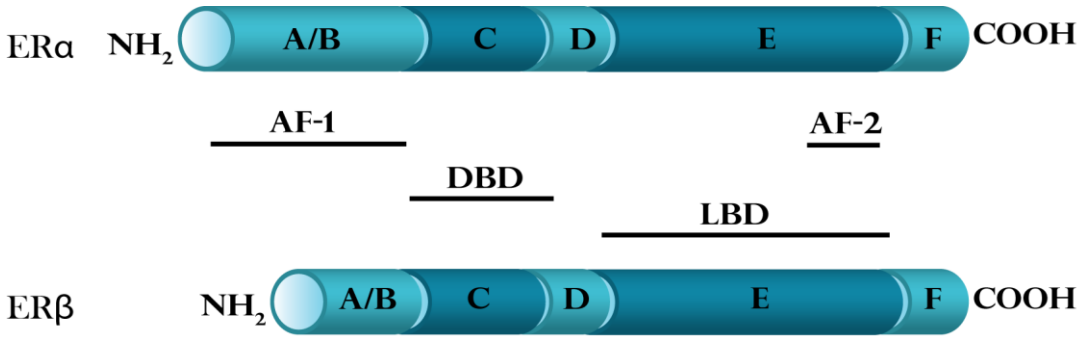

Figure 2.4. Schematic representation of ERα and ERβ structural and functional domains. ERs contain five domains: two transcriptional AF domains (AF-1; A/B domain and AF-2; F domain), DBD (C domain), hinge domain (D domain), LBD (E domain) (*recreated from Matthews and co-workers, 2003*[107]).

Despite having a similar mechanism of action, ERβ has an inhibitory effect on ERα-mediated transcriptional activation. It is suggested that these antagonizing effects are due to the difference in the transactivation functions[107]. ER-mediated transcriptional activation requires both the AFs. C-terminal AF-2 is active in both the ERs, but the AF-1 domain of ERβ is 30% identical to ERα[108] and possesses weaker activity[109]. ERα and ERβ exhibit differential responses to the estrogen and anti-estrogens. Ligands such as tamoxifen and raloxifene act as partial agonists/antagonists for ERα but as pure antagonists for ERβ[110]. The transactivation properties of both the receptors were studied in the context of AP-1 (activator protein-1) response element. Upon treatment with E2, ERα stimulated transcription of AP-1 containing reporter gene, while ERβ inhibited the transcription of the AP-1 containing reporter gene[111]. The differential response of ERα and ERβ to the same ligands is attributed to the difference in the AF-1 domain[107,112]. Moreover, co-regulators also determine the specificity and transcriptional activity of the ERs by selectively interacting with ERα and ERβ. For example, SRC-3 enhances ERα-mediated transcription, but does not affect the ERβ-mediated transcription[113].

Five isoforms of ERβ (ERβ1, ERβ2, ERβ3, ERβ4 and ERβ5) are known[114]. Most of the studies have focused on ERβ1 and the functional role of other isoforms is not fully understood. ERβ1 is a full-length isoform with intact helix 11 and helix 12. ERβ2 has shortened C-terminus resulting in disoriented helix 12. ERβ4 and ERβ5 isoforms completely lack helix 12[115]. ERβ2, ERβ4, and ERβ5 do not form homodimers. However, they can form a heterodimer with ERβ1 and can enhance its transactivation[115]. Chang and co-workers introduced ERβ1 in ERα-positive MCF-7 and T47D cells and assessed the impact of ERβ on ERα regulated genes. Interestingly, gene expression analysis revealed that unliganded ERβ1 modulates the expression of genes that are generally regulated by estrogen via ERα[116]. ERβ2 act as dominant-negative inhibitor of ERα by directing proteasomal degradation of

ERβ2/ERα heterodimers[114]. Binding affinity of ERβ2 for E2 (8 nM) is 8-fold lower as compared with binding affinity of ERβ1 for E2 (1 nM)[117]. Interestingly, by *in vitro* experiments, Poola and co-workers demonstrated that both ERβ4 and ERβ5 do not bind to E2 but can bind to estrogen response elements (ERE)[118].

Besides the classical estrogen receptors, truncated variants of ERα66 (namely ERα46, ERα36) are also known. In 2006, Wang and co-workers cloned alternatively spliced variant of classical ER and termed as ERα36. ERα36 lacks both AF-1 and -2 domain but retains the DNA- and ligand-binding domain. ERα36 is predominantly present in the plasma membrane and transduce estrogen-dependent activation of MAPK/ERK pathway[119,120]. Several groups have demonstrated that ERα66-negative breast cancer cell lines (MDA-MB-231 and MDA-MB-436) and ERα66-negative breast tumors (around 40%) express ERα36[119,121,122]. In ERα66-negative MDA-MB-231 and MDA-MB-436 cells, E2 stimulated c-Myc and cyclin D1 expression, induced cell proliferation and *in vivo* tumor growth. These effects were abrogated by ERα36 knockdown, suggesting that rapid estrogen signaling occurring via ERα36 can contribute to the malignant growth of ERα66-negative tumors[123].

ERα46 lacks the N-terminal 173 amino acids coded by the first exon of the ERα gene, thereby devoid of AF-1. ERα46 expression is reduced in endocrine-resistant breast cancer cells, and forced expression of ERα46 partially restored responsiveness to tamoxifen[124]. Overexpression of ERα46 in MCF-7 cells inhibits the ERα66-mediated estrogenic induction of c-fos and cyclin D1. This study suggested that ERα46 antagonizes the proliferative effect of ERα66 in MCF-7 cells partially by inhibiting ERα66 AF-1 activity[125].

2.7. Estrogen signaling

2.7.1. Genomic pathway

Degradation of unliganded ERs is prevented by its association with heat shock protein 90 (Hsp90)[126]. In the classical model of ER action, binding of the ligand to ERs stimulates Hsp90 dissociation and subsequent phosphorylation of ER. Phosphorylated ER forms homo- or hetero-dimers and undergoes nuclear translocation. E2-ER complex, then directly binds to specific DNA sequences called ERE and modulates gene transcription by interacting with transcriptional machinery and co-regulator proteins[127]. Co-regulators comprise of co-activators, co-repressors, co-integrators, histone deacetylases and acetyltransferases, and other transcription factors[128]. Upon binding of an agonist to ER, helix 12 repositions and aligns over LBD and covers the ligand-bound pocket. This conformation forms the specific

binding site at AF-2 for consensus LXXLL motif of co-activators. In contrast, binding of an ER antagonist sterically hinders helix 12 positioning, thereby preventing the formation of co-activator binding site and facilitating co-repressor binding[128,129]. Besides the classical mechanism, ER can also influence the transcriptional activity of genes, which do not contain functional EREs. Ligand bound-ER binds to other transcription factors such as c-Jun/ activating transcription factor-2 (ATF-2), Jun/Fos, and specificity protein 1 (Sp1) and forms complexes. These complexes, in turn, binds to AP-1, CRE (cyclic AMP response element), or Sp1 sites respectively and modulate gene transcription. This mechanism can also be activated by ER lacking DBD as it does not involve direct binding of ER to DNA. This pathway is the non-classical pathway[130-132]. The schematic representation of estrogen signaling pathways is given in Figure 2.5.

In addition to ligand-mediated activation of ER, its activity can be modulated in a ligand-independent manner by means of phosphorylation of serine and tyrosine residues in AF-1. Binding of polypeptide growth factors such as insulin-like growth factor-1 (IGF-1) or epidermal growth factor (EGF) to their respective receptors triggers the activation of various MAPK and Akt pathways. These kinases, in turn, phosphorylate serine residues (ERα: Ser104, Ser106, Ser118, Ser167, ERβ: Ser106, Ser124) of AF-1 domain. AF1 phosphorylation stimulates the transcriptional activity of ER by recruiting various co-regulators[133-135]. This cross-talk with growth factor signaling contributes to the development of endocrine resistance[134]. These three pathways are broadly grouped as the genomic pathway of estrogen action.

2.7.2. Non-genomic pathway

Estrogen can elicit rapid biological responses that occur in the time frame of seconds to minutes. This transcription-independent pathway is called non-genomic pathway. The non-genomic effects of estrogen are mediated by a sub-population of the classical ERs, which are located at the plasma membrane and GPER. GPER-mediated estrogen action includes cAMP production, intracellular mobilization of calcium, activation of MAPK signaling, membrane tyrosine kinase receptor activation and phosphatidylinositol 3,4,5-trisphosphate synthesis[101,136,137]. Moreover, activation of GPER with estrogen facilitates secretion of MMP, which in turn promotes the shedding of heparin-binding EGF-like growth factor (HB-EGF) from the surface. HB-EGF on binding to epidermal growth factor receptor (EGFR) promotes receptor dimerization and transactivation of EGFR/MAPK pathway. This pathway

explains the molecular basis of EGF-like effects of estrogen and ER-independent growth of tumors[136-139].

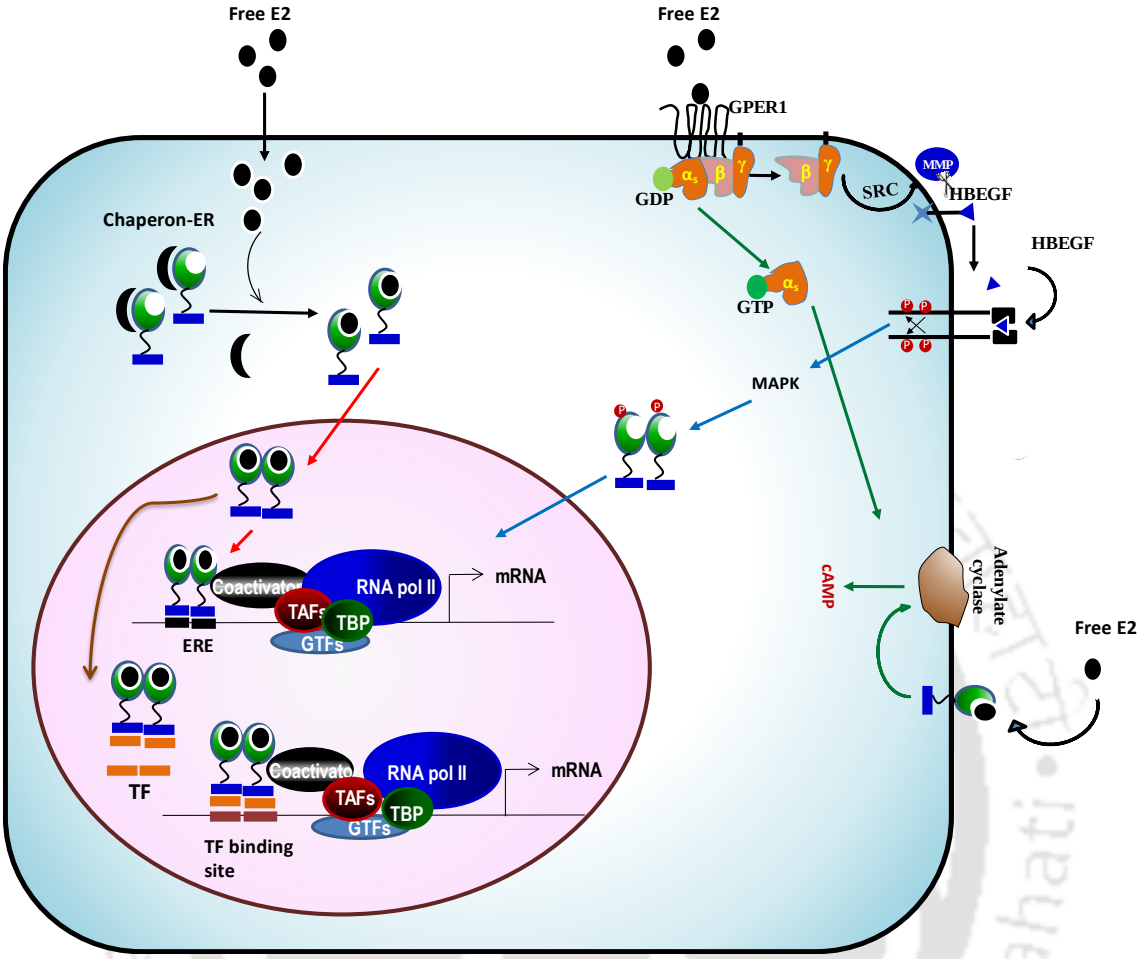

Figure 2.5. Schematic representation of genomic and non-genomic signaling pathways of estrogen. Red arrows and brown arrows indicate ligand-dependent direct- and indirect-genomic pathways, respectively. Blue arrows indicate ligand-independent genomic pathway. Green arrows indicate non-genomic pathway.

2.8. Endocrine resistance

Endocrine resistance is a major problem in breast cancer treatment. Advancement in breast cancer treatment was brought about by the approval of tamoxifen in 1973 for treating ER-positive breast tumors. Tamoxifen was used alone or in combination with cytotoxic agents for adjuvant and neoadjuvant therapy[140,141]. Despite the efficacy in regressing the ER-positive tumors, its success is limited by the acquisition of resistance in a large number of patients. Clinically, resistance is characterized by relapse of tumor after the completion of adjuvant therapy in adjuvant setting or progression of disease in neoadjuvant setting. Pathologically, an increase in tumor grade and proliferation marks the resistance to endocrine therapy. Studies have postulated several mechanisms for occurrence of endocrine therapy. Mutation or loss of ER expression occurs in 15-20% resistant tumors. Loss of ER is attributed

to the aberrant methylation of CpG island or histone deacetylation in the ER locus. *In vitro* studies have reported the frequent occurrence of deletion or point mutation in ER gene of tamoxifen-resistant MCF-7 and T47D breast cancer cells. Moreover, crosstalk with other growth factor signaling pathways such as EGFR/HER2 signaling, PI3K-Akt-mTOR pathway majorly contributes to endocrine resistance[142]. Therefore, targeting these signaling molecules may promote endocrine response with delayed resistance. Interestingly, *in vitro* study showed that the development of tamoxifen resistance in MCF-7 cells is accompanied by transition to more aggressive phenotype[143].

2.9. Metastasis

Metastasis is the ultimate cause of breast cancer-associated mortality and remains a steadfast challenge. It is a complex process characterized by dissemination of individual or group of cells from the primary tumor to a distant organ. The sequential events in metastasis are epithelial to mesenchymal transition, disruption of the basement membrane, dissociation from bulk tumor, localized invasion, intravasation into blood and lymph vessels, circulation through bloodstream to a distant organ, extravasation, and establishment of secondary tumor[144] (Figure 2.6). Several reports suggested that tumor microenvironment comprising of surrounding ECM, and stromal cells such as fibroblast, endothelial cells, and immune cells influence tumorigenesis and progression of breast tumor[145-147]. Interestingly, tumor cells can also modulate the surrounding microenvironment to promote their growth and survival via a positive feedback loop[148]. ECM is a dynamic network with biochemically distinct components such as proteins, proteoglycans, polysaccharides, and glycoproteins[149]. ECM is constantly remodeled in various stages of embryonic and postnatal development involving altered synthesis or degradation of its components. ECM homeostasis is achieved by controlled expression and activities of ECM degrading enzymes, which include MMPs, A Disintegrin and Metalloproteinases (ADAMs), A Disintegrin and Metalloproteinase with Thrombo-spondin motifs (ADAMTSs), cathepsins, heparanases, matriptases, serine/threonine proteases, hyaluronidases, urokinase-type plasminogen activators (uPA) and tissue plasminogen activators (tPA), and their endogenous inhibitors such as cystatins, tissue inhibitors of metalloproteinases (TIMPs), and plasminogen activator inhibitors (PAI)[148]. Disruption in the delicate balance between these proteases and their inhibitors has a profound effect on tumor invasion and metastasis.

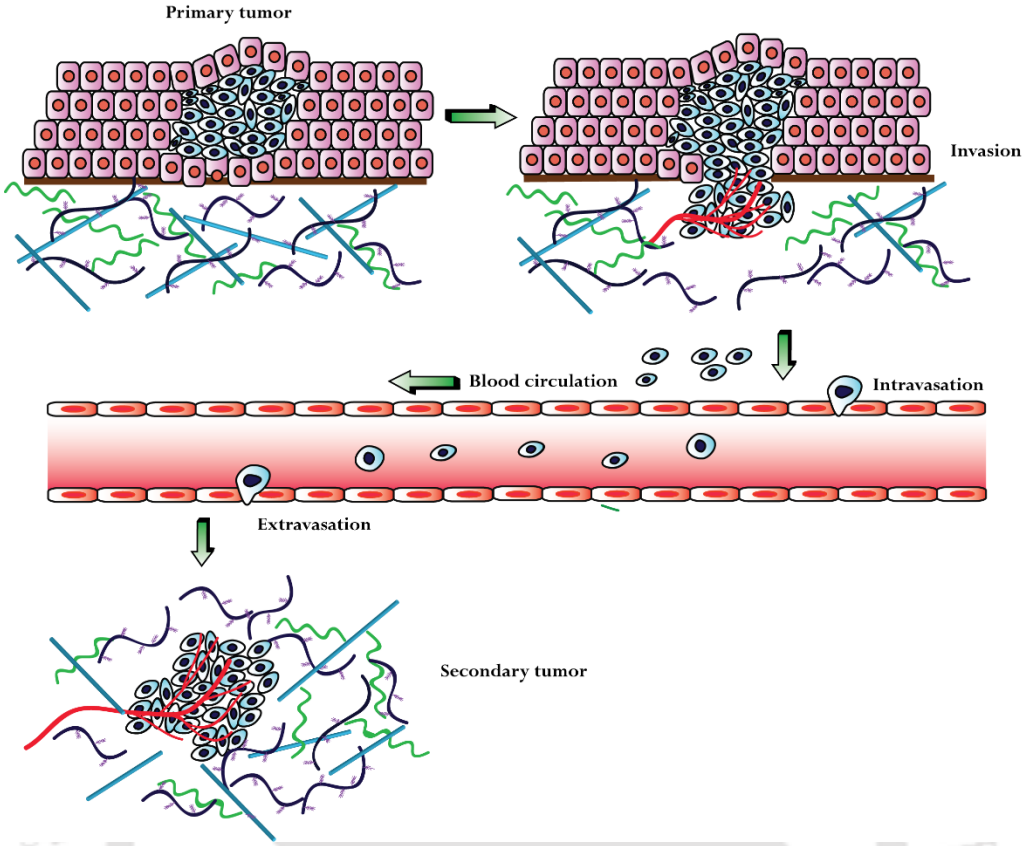

Figure 2.6. The process of tumor metastasis. Sequential events in metastasis are- i) proliferation of tumor cells in the primary site, ii) local invasion of detached cells, iii) intravasation into blood capillaries, iv) arrest and extravasation to the secondary site, and v) formation of a secondary tumor (*recreated from Gout and co-workers, 2008*[150])

2.10. Cathepsins

The discovery of lysosomes in 1955 by Christian de Duve paved the way for the discovery and characterization of numerous hydrolases in the past 65 years[151]. Hydrolases, including proteases, nucleases, lipases, and amylases, are involved in the effectual degradation of biological macromolecules in the lysosomes. Amidst the 50 lysosomal hydrolases, proteases of five catalytic types (aspartic, serine, cysteine, threonine and metallo) have gained importance over the years due to their remarkable role in cell death, protein turnover, proliferation, migration and invasion[152].

Cathepsins are involved in various physiological processes such as propeptide and prohormone processing, apoptosis, bone remodeling, and reproduction[153]. Cathepsins, based on the nucleophilic amino acid at the active site, are categorized as serine cathepsins (cathepsins A and G), aspartic cathepsins (cathepsins D and E) and cysteine cathepsins (cathepsins B, C, F, H, K, L, O, S, V, X and W). Except for cathepsin C, which is a 200 kDa tetrameric protein, cysteine cathepsins are 25-30 kDa monomeric proteins. Most of the

cysteine cathepsins are endopeptidases (cathepsins F, K, L, O, S, V and W), while cathepsin C (aminopeptidase) and cathepsin X (carboxypeptidase) are exopeptidases. Though cathepsins B and H are exopeptidases, they can also perform endopeptidase activity. Tissue-specific distribution of cathepsins K, S, V and W was observed while other cathepsins are ubiquitously expressed[8].

Full-length cysteine cathepsins consist of signal peptide, propeptide and catalytic domain. The catalytic domain represents the active enzyme. The signal peptide is responsible for translocation to endoplasmic reticulum during translation. The propeptide serves as a reversible inhibitor to prevent premature activation of the catalytic domain[154]. Cysteine cathepsins, which are primarily located in endolysosomal compartments, were initially thought to be active only in an acidic environment. However, the observed extracellular activity of cysteine cathepsins indicates that pH is not the sole determinant of their activity[6]. In pathological conditions, hypoxia leads to acidification of peri- and extracellular matrix, which facilitates the degradation of ECM protein by cysteine cathepsins[155]. Since ECM is composed of elastins, collagens, and proteoglycans, cathepsins with collagenolytic and elastolytic activities are majorly responsible for ECM remodeling.

Proteolytic activity of lysosomal cysteine proteases is regulated by pH, zymogen activation and endogenous protein inhibitors. Cathepsins are naturally synthesized as inactive precursors and activated by removal of propeptide by other proteases or by autocatalytic removal at acidic pH[156]. Once activated, because of its high concentration in lysosomes, cathepsins can be disruptive. Hence, its extra lysosomal activities are controlled by protein inhibitors, cystatins, serpins, and thyropins[154]. Cystatins are the endogenous competitive, reversible inhibitors that prevent the binding of substrate to the active site of cathepsins[39,157].

2.11. Cystatins

Back in the early 60s, Hayashi and his group reported that cultured rabbit skin explants release thiol protease and its inhibitor upon stimulation with antigen[158,159]. Later cysteine protease inhibitor was isolated from chicken egg white by Fossum and Whitaker in 1968[160]. Keilovit and Tomfigek demonstrated that this egg white protein inhibits bovine cathepsin B and cathepsin C[161]. Barrett then named this egg white protein as "cystatin"[162]. Barrett suggested that proteins of the higher organisms with the following characteristics: i) inhibition of papain and cathepsin B, ii) molecular weight close to 13 kDa, iii) pI of 4.5 to 6.5 and iv) the stability to heat and alkali, to be considered under a single homologous family of

cystatins[162]. Subsequently, several groups isolated number of proteins with similar sequence and characteristics of chicken cystatin from plants, tissues and body fluids of humans and animals. These proteins were clustered under cystatin superfamily[157,163-166]. In the first International Symposium on Cysteine Proteinases and their Inhibitors (1985), known cystatins were categorized into three families under cystatin superfamily based on location, presence of disulfide bond, size, and complexity of polypeptide. Family 1 is called the stefin family. It comprises of intracellularly expressed, unglycosylated inhibitors, cystatins A and B of ~11 kDa. Family 2, is called the cystatin family, includes extracellularly expressed cystatins C, D, F, E/M, S, SA, and SN, which are in the range of 13-14 kDa. Family 3 is called the kininogen family. It includes glycosylated inhibitors with the molecular mass in the range of 88-114 kDa. Family 1 cystatin lacks disulfide bonds and signal sequences, while Family 2 contains a signal sequence at the carboxy terminus and disulfide bonds[167-169]. Later, proteins that do not possess cysteine protease inhibitory activity were also included in cystatin superfamily due to the presence of cystatin-like domain. They are HRG (histidine-rich glycoprotein), fetuins, CRP (cystatin-related protein) and CRES (cystatin-related epididymal spermatogenic)[169].

Members of cystatin superfamily have a common fold of a central helix core, wrapped by five anti-parallel β-strands. Cystatins inhibits cysteine cathepsins by forming tight reversible complexes. They bind to the active site of protease via "tripartite wedge". This wedge-shaped structure is formed by three conserved motifs (N-terminal region and two loops between β-strands) that contain mostly hydrophobic residues. Since the wedge-shaped structure is complementary to the active-site groove of cathepsins, its binding to the active site blocks the access of substrates to the active site of the cathepsins[170] (Figure 2.7).

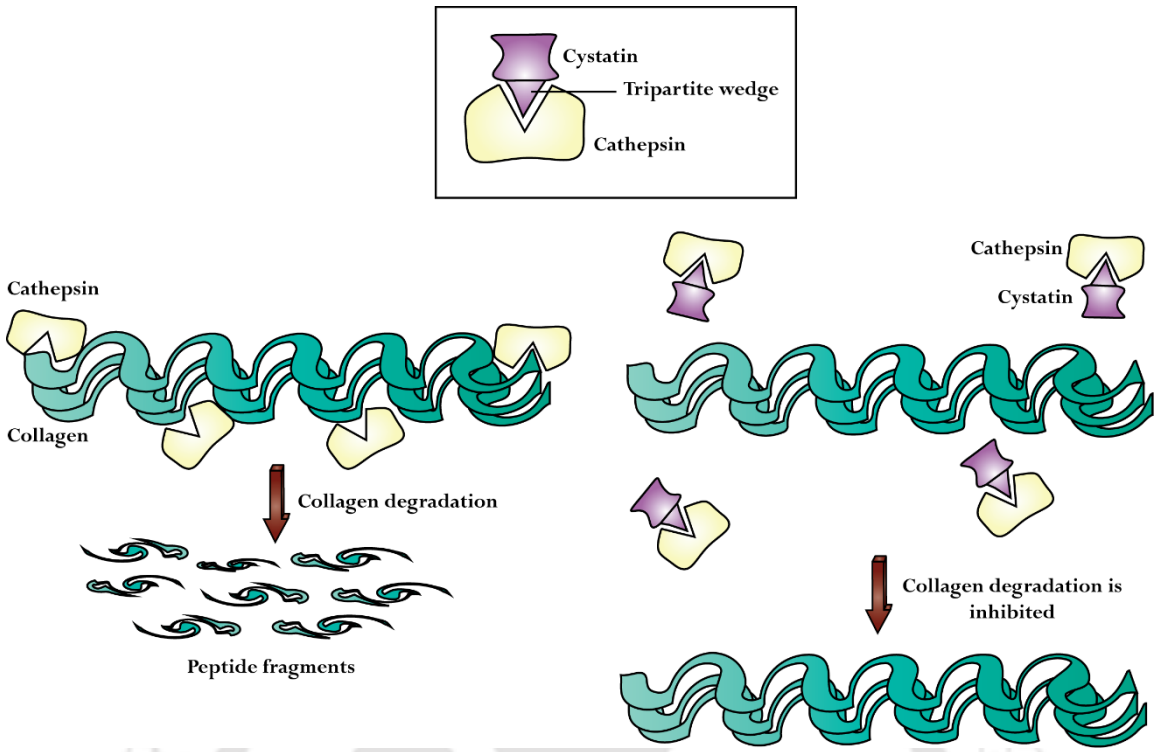

Figure 2.7. Schematic representation of inhibition of cathepsins by cystatins. Cystatins forms 'tripartite wedge' and binds to the active-site groove of cathepsins, thereby blocking the access of collagen to the cathepsins. This prevents cathepsin-mediated degradation of collagen.

2.12. Cathepsins and cystatins in cancer

Besides the normal physiological functions, several reports have established the dynamic role of cysteine cathepsins in various pathological conditions. Around 40 years ago, Poole and co-workers reported the involvement of cathepsin (cathepsin B) in the malignant progression of breast tumors[171]. Thereafter, aberrant expression, activity, and mislocalization of various cysteine cathepsins have been reported in many cancers, including melanoma, breast, colorectal, thyroid, gastric, brain, bladder, prostate and lung cancers[172,173]. Altered expression of cysteine cathepsins tilts the homeostatic balance to favor ECM remodeling, thereby promoting tumor progression, invasion, and metastasis[11-13]. Therefore, changes in the expression of cathepsins have diagnostic or prognostic value[174]. Though primarily intracellular, cysteine cathepsins are detected on the surface of tumor cells or in extracellular spaces[4,175]. Surface or extracellular cathepsins facilitate ECM remodeling and invasion of tumor cells into blood vessels and surrounding tissue[176].

Most of the cysteine cathepsins are upregulated in malignant tumors of various sites of origin (Table 2.2). Increased expression of cathepsins is often associated with increased motility and invasion of tumor. Cathepsin B induces the fibroblast-mediated invasion in breast cancer[177]. Cathepsin B is predominantly located in invasive edges of the malignant

cells[178]. Antibody-mediated blocking of cathepsin H resulted in inhibition of invasion of glioblastoma cell lines[179]. Coculture of cathepsin K-negative breast cancer cells with cathepsin K-expressing fibroblasts promoted invasion, and this effect was abrogated with cathepsin K inhibitors[180]. Downregulation of cathepsin L retarded the tumor migration and invasion by inhibiting transforming growth factor-β (TGF-β)-mediated epithelial-mesenchymal transition (EMT)[181]. Mutation in cathepsins B or S reduced tumor growth and angiogenesis in mouse model[155]. Cathepsin X-mediated inactivation of profilin 1, a tumor suppressor, induced migration and invasiveness of prostate cancer cells[182]. The involvement of cathepsins in dissemination of tumor cells, degradation of ECM components, angiogenesis, cancer autophagy, and the EMT is extensively studied[176].

Table 2.2. List of differentially expressed cathepsins in cancers.

Cathepsin	Cancer type	Expression in cancer
B	Gastric, lung, colon, ovarian, cervical, bladder, breast, thyroid, melanoma, glioblastoma, hepatocellular carcinoma	Up
B	Prostate	Down
F	Cervical	Up
H	Prostate, colorectal, breast, melanoma, head and neck carcinoma, glioma	Up
H	Melanoma	Down
L	Breast, lung, gastric, colon, ovarian, pancreatic, head and neck carcinoma, melanoma, glioma	Up
L	Prostate	Down
K	Gastric, breast, lung, prostate, renal, squamous cell carcinoma, basal cell carcinoma, melanoma	Up
K	Lung	Down
S	Gastric, pancreatic islet cell, astrocytoma, hepatocellular carcinoma, glioblastoma, melanoma	Up
X	Prostate, gastric, breast, lung, colorectal, melanoma	Up

(Adapted from Jedeszko and Sloane 2004, Tan et al., 2013)

Interestingly, one of the deleterious aspects of cysteine cathepsins is its involvement in drug resistance. Studies showed that drug accumulation in the lysosome is directly proportional to the extent of tolerance of tumor cells to therapeutic agents. Hydrophobic weak base drugs upon internalization may get incorporated into lysosomes. Lysosomal sequestration of the drugs hinders them from reaching the intracellular target effectively[183]. Knocking down of cathepsin L potentiated paclitaxel-induced apoptosis in ovarian cancer cells[184]. Acquisition of cisplatin or paclitaxel resistance was associated with the upregulation of cathepsin L in A549 lung cancer cells and silencing of cathepsin L restored chemosensitivity[185]. Most importantly, inhibition of cathepsin L enhanced the stability and availability of drug targets such as ER, androgen receptor, histone deacetylase 1 (HDAC1),

and topoisomerase-Iiα, resulting in cellular sensitivity to tamoxifen, flutamide, trichostatin A, imatinib, and doxorubicin[186]. Due to these pathological roles of cathepsins, endogenous regulation of cathepsins activity is crucial.

In normal cells, an intricate balance is maintained between expression levels of cathepsins and cystatins. Dysregulated expression of cathepsins, when not balanced by cystatins, alters cathepsin: cystatin ratio, thereby contributing to malignant progression of tumors[187]. Studies suggest that inverse correlation exists between the level of cysteine proteases and cystatins in tumor microenvironment[188]. Moreover, as the tumor proceeds towards metastatic stage, the level of cystatins in extracellular spaces and cytosol are drastically reduced in most of the cancers[169]. However, both positive and negative roles played by cystatins A, B, C and F in tumorigenesis and progression are reported[189]. Therefore, cystatins are considered as potential biomarkers for diagnosis or prognosis. Prognostic significance of cystatins is summarized in Table 2.3.

Table 2.3. Clinical significance of cystatin levels in the prognosis of patients.

Type of cancer	Sample	Biomarker	Expression at favorable prognosis
Bladder	Urine	Cystatin B	Low
Multiple myeloma	Serum	Cystatin C	Low
Non-Hodgkin B-cell lymphoma	Serum	Cystatin C	Low
Glioma	Tissue	Cystatin A	Low
		Cystatin C	High
Meningioma	Tissue	Cystatin C	High
Breast	Tissue	Cystatin A	High
		Cystatin B	High
		Cathepsin B/ Cystatin B	Low
		Cystatin M	High
Colon and rectum	Serum	Cystatin B	Low
		Cystatin C	Low
	Tissue	Cystatin F	Low
Esophagus	Serum	Cystatin C/Cathepsin B	High
	Tissue	Cystatin SN	High
Head and neck	Tissue	Cystatin A	High
		Cystatin B	High
		Cystatin C	High
Lung	Serum	Cystatin C/ Cathepsin B	Low
	Tissue	Cystatin A	High
		Cystatin B	High
Prostate	Serum	Cathepsin B/ Cystatin A	Low
Kidney	Tissue	Cystatin C	Low

(Adapted from Breznik et al., 2019)

Lah and co-workers reported that cystatins extract from sarcoma was less effective against cathepsin B than cystatins extract from non-tumorous liver. Purified CSTA from

sarcoma showed less ability to inhibit papain, and cathepsin B, H and L[190]. Cystatin C/cathepsin B ratio is reduced in neoplastic breast tissues compared to the normal breast tissues[13]. Low cystatin C expression is associated with reduced overall survival in prostate cancer patients. Inhibition of cystatin C expression increased the invasiveness of prostate cancer (PC3) cells[191]. Cystatin M was consistently expressed in normal human breast epithelial cells, whereas its expression was decreased by 86% in invasive ductal carcinoma (IDC) cells from stage I to IV patients. Complete loss of expression of cystatin M was observed in two-third of stage IV patients. Severe combined immunodeficient (SCID) mice implanted with cystatin M expressing breast cancer cells exhibited delayed primary tumor growth and less metastatic burden. Cystatin M expression in the highly invasive MDA-MB-231 cells inhibited its proliferation, migration, and matrigel invasion[192].

During tumor development and progression, cystatins regulate several mechanisms and signaling pathways. In gastric cancer, cystatin B reduces cell proliferation and metastasis by downmodulating PI3K-Akt-mTOR signaling pathway[193]. Further, the silencing of integrin-linked kinase in gastric carcinoma cells increased cystatin B and reduced invasion via the Akt pathway[194]. Cystatin B protects the cells from both cathepsin-mediated and cathepsin-independent apoptosis[195]. Cystatin C induced apoptosis in melanoma cells[196]. Cystatin E/M increased apoptosis in T-box protein (TBX-2)-expressing breast cancer cells[197]. Several studies have demonstrated cystatin C as both a positive and negative regulator of tumor growth. Negative regulation is due to its inhibitory activity against cysteine cathepsins, while tumor promoter effect is likely to be via proteolysis-independent mechanisms[198]. The expression of cystatin E/M impairs cell growth by inhibiting cathepsin L[199]. Cystatin SN knockdown activates glycogen synthase kinase 3β (GSK3β). This results in increased glycogen accumulation and cell senescence[200].

2.13. Cystatin A (CSTA)

In 1976, the acid cysteine protease inhibitor (ACPI) was first isolated from rat skin[201]. Later, Brzin and co-workers isolated a similar protein from human polymorphonuclear granulocytes and named it as "stefin" after the J. Stefan Institute in Ljubljana[164]. Stefin A, also called cystatin A, belongs to the family-1 of cystatin superfamily and naturally inhibits cysteine cathepsins B, H, L, S, and papain[202-204]. Cystatin A, encoded by CSTA gene located on chromosome 3q21.1 (Figure 2.8), is a single polypeptide chain of 98 amino acids with molecular mass of 11 kDa. It is an acidic protein with isoelectric point ranging from 4.5-5[157,164,168,205]. Cystatin A (CSTA) is detected in the normal squamous epithelium[206], follicular

dendritic cells in lymphoid tissues[207], thymic cells[208], liver cells[209], granulocytes[210], and basal epithelial cells in the prostate[211]. In the epidermis, CSTA is located throughout the supra-basal layers as diffused cytoplasmic distribution, and synthesized strongly in the granular layer[40].

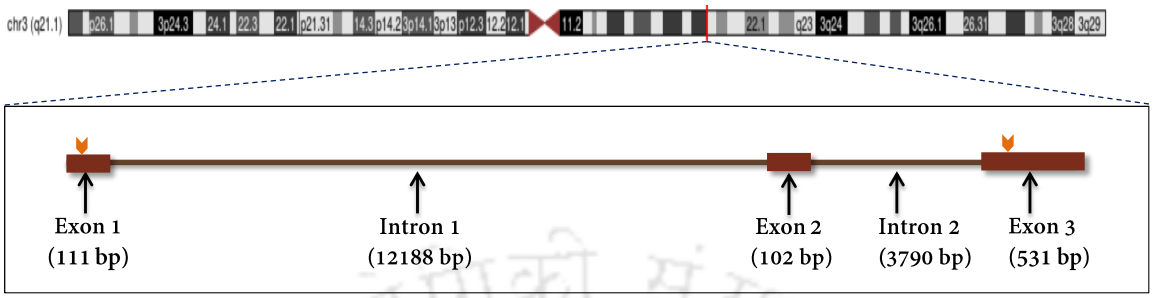

Figure 2.8. Localization and structure of CSTA gene. A snapshot from UCSC genome browser representing CSTA locus (red vertical line in upper panel) in chromosome 3. The structure of CSTA gene is graphically represented in the lower panel. CSTA gene comprises three exons and two introns. The size of the exon and intron is mentioned in base pairs (bp). Orange arrowheads indicate the start codon and stop codon.

Over the years, several studies have suggested various physiological functions for CSTA. As a significant contribution, Blaydon and co-workers reported that homozygous nonsense mutation in CSTA gene results in exfoliative ichthyosis, a skin disorder characterized by peeling of skin of the palms and soles. This unraveled the role of CSTA in desmosome-mediated cell-cell adhesion in the lower levels of the epidermis[40]. Moreover, Scott and co-workers reported that Csta1 and Csta3 are strongly expressed in neonatal mouse skin than adult skin, and the expression decreases with age, suggesting its role in the development of the epidermis[212]. CSTA acts as a biochemical skin barrier against mite allergens by inhibiting mite cysteine proteases Der f 1 and Der p 1[213]. Polymorphism in the CSTA gene and unstable CSTA mRNA has been associated with atopic dermatitis, a chronic inflammatory skin disease often associated with a defective epidermal barrier[214]. However, CSTA expression was found to be upregulated in the plaques of psoriasis vulgaris, a skin disease characterized by hyperproliferation of skin cells, eventually resulting in red, scaly plaques[215]. CSTA is suggested as an important player in keratinocyte proliferation and terminal differentiation[216].

In addition to its role in adhesion, differentiation, and epidermal protection, CSTA was reported to have a functional role in apoptosis by inhibiting cathepsin B. Lysosomal destabilization and release of cathepsins are the characteristic features of apoptosis initiation[217]. Forced expression of CSTA in hepatoma cell line prevented bile salt-induced cathepsin B activation and apoptosis[41]. Peri and co-workers have demonstrated that in Human Epithelial cells (HEp-2), CSTA can inhibit herpes simplex virus-induced apoptosis[218]. Moreover, ultraviolet B-induced procaspase-3 cleavage and activation were abrogated by

CSTA expression in human keratinocytes. However, no effect was detected in caspase-8 and caspase-9 activities in CSTA-transfected ultraviolet B irradiated cells[219]. Kuopio and co-workers reported that immunostaining of breast neoplasms showed strong positivity for CSTA in apoptotic bodies[31]. Recently, Ma and co-workers reported that overexpression of CSTA in squamous cell carcinoma cells enhances gemcitabine-induced apoptosis. This is contradictory to previously reported inhibitory role of CSTA in apoptosis[42]. However, all the studies reinforce the hypothesis that CSTA is associated with apoptosis.

Besides these physiological roles, most importantly, cystatins protect the host tissue from catastrophic proteolysis by accidentally released endogenous cathepsins or exogenous cathepsins of bacterial and viral origin[218,220-224].

2.14. CSTA in cancer

The relationship between CSTA expression with prognosis and cancer progression is confounding. CSTA expression is downregulated in various cancer, including brain, breast, esophageal, lung, prostate, and head and neck carcinoma[19,42,43,211,225,226] . In contrast, other studies have reported higher levels of CSTA in breast, colorectal, lung and, head and neck carcinoma[32,174,211,227,228]. Higher CSTA expression correlated with a better outcome in head and neck, and breast cancer[17,227]. CSTA expression was reduced with an increase in tumor grade, invasiveness and metastatic potential unveiling the tumor suppressor potential of CSTA. On the other hand, higher expression correlated with shorter or poor survival of glioma patients[229]. The confusion about CSTA as a prognostic marker was resolved to some extent when levels of cathepsins were also considered along with CSTA. In prostate tumors of the same Gleason score, tumors with high cathepsin B/CSTA ratio were associated with lymph node metastasis compared to tumors with low cathepsin B/CSTA ratio[230]. However, this needs further investigation in other types of cancers.

Studies on ectopic expression or knockdown of CSTA *in vitro* and *in vivo* gives further understanding of the functional effects of CSTA expression and the underlying mechanism. Li and co-workers have demonstrated that overexpression of CSTA in cell line model of esophageal squamous cell carcinoma reduced tumor growth, angiogenesis, and invasion[43]. CSTA suppresses tumor growth by modulating MAPK pathway and inhibits EMT in lung cancer cells. Stable transfection of CSTA in CSTA-negative cells promoted mesenchymal-epithelial transition (MET)[42].

However, in breast cancer, studies attempting to understand the precise role of CSTA are limited. In an immunohistochemical study on infiltrative breast carcinomas, Kuopio and co-workers found CSTA positivity in large tumors with higher mitotic activity. They also showed that CSTA positivity was associated with an increased risk of death[31]. Lah and co-workers found lower CSTA mRNA and protein levels in breast carcinoma compared to their matched control in the majority of the 50 matched pairs under their study[226]. Levicar and co-workers reported 1.9-fold higher levels of CSTA in cytosols of primary invasive breast tumors compared to normal breast parenchyma. However, they did not find a significant association between CSTA and breast cancer prognostic factors[32]. Parker and co-workers analyzed CSTA expression in 142 primary breast tumors and found that CSTA expression not only correlated with disease-free survival but also decreased the risk of distant metastasis[34]. Notwithstanding the conflicting results, these studies suggest a role for CSTA in breast cancer progression. More recent studies on *in vivo* breast tumor metastasis models have provided valuable insights into the mechanistic role of CSTA in breast tumor progression. Parker and co-workers showed that CSTA expression and propensity to metastasize were inversely related in 4T1-derived cell lines, which exhibit varying degrees of metastasis in the syngeneic murine model. Furthermore, forced expression of CSTA in 4T1.2, a highly metastatic 4T1 line, reduced spontaneous metastasis to the bone[34]. Using the same model, Withana and co-workers showed that knockdown or selective inhibition of cathepsin B could also reduce bone and lung metastasis[35]. Moreover, CSTA is considered a myoepithelial cell marker due to its abundant expression in myoepithelial cells. Myoepithelial cells are natural tumor suppressors with constitutive expression of protease inhibitors. Targeted loss of CSTA from an immortalized myoepithelial cell line prevented myoepithelial-induced suppression of invasion[17]. Immunostaining of CSTA showed more expression in normal breast ducts and low-grade ductal carcinoma *in situ* (DCIS) compare to high-grade DCIS denoting the potential of CSTA to serve as a marker for distinguishing DCIS lesions with low risk of relapse[17].

Taken together, these studies suggest that in breast cancer, CSTA inhibits tumor invasion and metastasis by inhibiting cathepsin B. Moreover, its tumor suppressor role is manifested in the tumor microenvironment.

2.15. Regulation of CSTA expression

Understanding CSTA regulation will help in developing novel treatment strategies based on cystatin function. The gene structure of CSTA comprises three exons and two introns. The promoter activity of the 5' flanking region of CSTA was demonstrated by

Takahashi and co-workers[231]. Activation of protein kinase C with TPA (12-O-Tetradecanoyl-phorbol-13-acetate) stimulated CSTA promoter activity upon binding of c-Jun, JunD, and c-Fos to the TPA responsive elements (TRE) in its 5' upstream region. Later the same group reported that CSTA expression is regulated positively by Ras/MEKK1/MKK7/JNK pathway and negatively by Ras/Raf/MEK1/ERK pathway. TRE in 5' upstream is critical for both the positive and negative regulation[232]. The stratified epidermis consists of following layers assembled upon basement membrane: basal, spinous and granular layers, and the cornified envelope. CSTA is reported to be expressed in the suprabasal layers of the epidermis[40]. Notably, though the same signaling molecule (Ras) triggers both the pathways due to the differential localization of active downstream targets, expression of CSTA varies between basal and suprabasal layers of the epidermis[232]. Immunohistochemical analysis showed the presence of phosphorylated ERK in nuclei of basal layer cells, suggesting that suppression of CSTA expression in the basal layer is probably due to activation of negatively influencing Raf-1/MEK1/ERK pathway. On the other hand, phosphorylated JNK was detected in the nuclei of spinous and granular layers suggesting that activation of MEKK1/MKK7/JNK pathway stimulates the expression of CSTA in suprabasal layers of the epidermis[232].

Calcium has the potential to stimulate terminal differentiation of keratinocytes. Stimulation of NHK cells with calcium downregulated Raf-1/MEK1/ERK pathway, which in turn induced the CSTA expression[232]. cAMP elevating agent, forskolin also stimulated CSTA mRNA expression in human keratinocytes[231]. However, the mechanism and significance of this induction remain unknown. Transfection of normal human keratinocyte cells with AP-2γ expressing vectors induced CSTA expression. Further analysis showed that AP-2γ binds to the AP-2 binding site (-75 to -84) of CSTA and induces its expression[233]. Gupta and co-workers reported desmosomal cadherin-mediated regulation of CSTA. Ectopic expression of desmoglein 2 (Dsg2) in mice induced CSTA expression. Moreover, knockdown of Dsg2 reduced cell-cell adhesion. Interestingly, the silencing of CSTA further reduced cell-cell adhesion[234]. Identification of Dsg2 binding sites in CSTA gene may give more insights on cadherin-mediated regulation of CSTA. Another study reported that c-Jun expression was frequently downregulated in esophageal cancer, and its expression was associated with CSTA. *In vitro* chromatin immunoprecipitation assay showed that c-Jun could directly bind to the promoter region of CSTA and regulate its expression[235].

Vendrell and co-workers performed cDNA mini-arrays analysis to detect novel estrogen-regulated genes in MCF-7 derived cell line, in which CSTA was reported as one of the estrogen suppressed genes[27]. However, the mechanism behind the E2-mediated regulation of CSTA is not yet elucidated.

Besides these mechanisms, understanding the epigenetic component of the regulation of CSTA may further unravel its prognostic and therapeutic value.

2.16. Epigenetic regulation

Chromatic structure determines the organization of genetic information in the cells. Packaging of chromatin influences the accessibility of DNA, thereby determining the status of gene expression[236]. Epigenetic events are the changes that result in the modulation of the gene expression without any alteration in the DNA sequence. In general, epigenetic events include DNA methylation, histone modifications, nucleosome positioning, and non-coding RNAs[237]. During developmental stages, these events provide cellular diversity by regulating the access of genetic information by cellular machinery[236]. Failure in the maintenance of epigenetic marks results in aberrant activation or inhibition of various genes resulting in pathological conditions, including cancer[236].

2.16.1. DNA methylation

DNA methylation involves the addition of a methyl group to the cytosine ring of CpG dinucleotide. This covalent modification of DNA is catalyzed by enzyme DNA methyltransferase (DNMT) with S-adenosyl-methionine as a methyl donor. This reaction is catalyzed by the DNMT family, including DNMT1, DNMT3A and DNMT3B[237]. DNMT3A and DNMT3B are *de novo* methyltransferases and majorly acts during embryo formation by establishing the pattern of methylation. DNMT1 are maintenance methyltransferases involved in the maintenance of methylation patterns in the genome by methylating the newly synthesized DNA strand after replication[238]. Mammalian genome doesn't have even distribution of CpG dinucleotides. In some regions, CpGs are densely located, which are called CpG island. CpG islands occupy approximately 60% of total human gene promoters[239].

DNA methylation suppresses gene transcription by two mechanisms. One way is by direct interference in the transcription factor binding to their response element due to the projection of methyl groups to the major groove of DNA[240]. On the other hand, methylation of cytosine facilitates the recruitment of methyl-CpG binding proteins (MCBPs) followed by

HDAC. This results in deacetylation and condensation of chromatin, thereby represses the gene expression[241]. The schematic representation of DNA methylation-mediated silencing of gene expression is given in Figure 2.9.

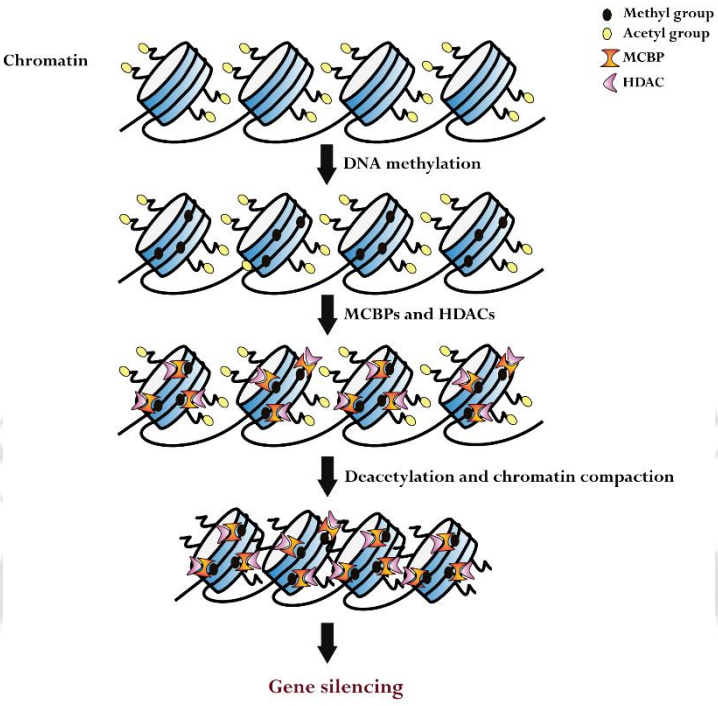

Figure 2.9. Mechanism of DNA methylation-mediated silencing of gene expression. Methylated CpGs are recognized by MCBPs, followed by the recruitment of HDACs. HDACs deacetylates the histone resulting in condensation of chromatin, which in turn inhibits transcription (*recreated from Marques-Magalhães and co-workers, 2018*[242]).

2.16.2. Epigenetic regulation of cystatins in cancer

During tumorigenesis, extensive reprogramming of epigenetic machinery occurs. DNA methylation was the first epigenetic aberration reported in cancer[243]. Hypermethylation of tumor suppressor genes is a frequent event in the origin of various tumors. On the other hand, DNA hypomethylation can contribute to tumorigenesis by three mechanisms: acquisition of chromosomal instability, loss of genome imprinting, reactivation of transposable elements. Due to the reversible nature of epigenetic modification, epigenetic therapy is emerging as a promising option in cancer therapeutics[244].

Members of cystatin superfamily are epigenetically silenced through DNA methylation-dependent mechanisms in various cancers, including breast, brain, and lung cancer[14]. In breast cancer cells, hypermethylation of cystatin M encoding CST6 gene was associated with the loss of cystatin M expression[245]. Kim and co-workers reported CST6 as one of the methylation-sensitive genes in malignant glioma[246]. Ma and co-workers demonstrated the re-expression of cystatin B in seven lung cancer cell lines upon global demethylation with 5-aza-2-

deoxycytidine[247]. However, the exact locus of hypermethylation in the CSTB gene was not identified. Thus, DNA methylation may play a crucial role in the regulation of CSTA and thereby its role in the genesis and progression of tumors.

CHAPTER 3

Materials and methods

3.1. Plasticware, chemicals, and reagents

Cell culture plasticware was purchased from Eppendorf (Hamburg, Germany). Cell culture media, fetal bovine serum (FBS), PowerUp SYBR Green PCR Master Mix, High-Capacity cDNA Reverse Transcription Kit, ERα siRNA (short interfering RNA) (Cat No. 4392420), scrambled siRNA (Cat No. AM4611), Lipofectamine RNAimax, Lipofectamine 3000, TRIzol reagent and AmpliTaq Gold PCR master mix were from Invitrogen (CA, USA). Dulbecco's phosphate buffered saline (DPBS), trypsin, penicillin, streptomycin, and charcoal-stripped FBS (csFBS), mitomycin C, puromycin were from HiMedia (Mumbai, India). ProteoGuard™ EDTA-Free Protease Inhibitor Cocktail and pMD20 vector were purchased from Clontech (CA, USA). Monoclonal β-actin (AM4302) antibody and EpiJET Bisulfite Conversion Kit were purchased from Thermo Scientific (PA, USA). Polyclonal peptide affinity-purified CSTA antibody was raised and affinity-purified by Abgenex (Bhubaneshwar, India). Monoclonal CSTA antibody (ab10442) was from Abcam (Cambridge, UK) and polyclonal ERα antibody (sc-543) was purchased from Santa Cruz Biotechnology (CA, USA). PR antibody (8757S), monoclonal ERα antibody (8644S) and EMT Antibody Sampler Kit (9782T) were from Cell Signaling Technology (Massachusetts, USA). Polyclonal histone H3 antibody (BB-AB0055) and normal rabbit immunoglobin G (IgG) antibody (BB-AB0001) were purchased from BioBharati LifeScience Pvt. Ltd (Kolkata, India). pBABE-puro was a

gift from Hartmut Land & Jay Morgenstern & Bob Weinberg (Addgene plasmid # 1764)[248]. Routine laboratory buffers, solvents and salts were from Merck (Mumbai, India) or SRL (Mumbai, India). Details of various ligands and inhibitors used in the present study are given in Table 3.1.

Table 3.1. Ligands and inhibitors used in this study.

Drugs	Catalogue No.	CAS No.	Company	Abbreviation
17β-estradiol	E8875	50-28-2	Sigma	E2
4-hydroxytamoxifen	H7904	68047-06-3	Sigma	Tam
Propyl pyrazole triol	H6036	263717-53-9	Sigma	PPT
5-azacytidine	100821	320-67-2	MP Biomedicals	5-aza
Testosterone propionate	86541	57-85-2	Sigma	T
Progesterone	P8783	57-83-0	Sigma	P4
Fulvestrant	I4409	129453-61-8	Sigma	Ful
Dexamethasone	D4902	50-02-2	Sigma	D

3.2. Cell culture and treatments

3.2.1. Cell lines and cell culture

MCF-7, MDA-MB-231, MDA-MB-453, T47D and ZR-75-1 cells were procured from the National Centre for Cell Science, Pune, India. MCF-7 cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) with phenol red. T47D, ZR-75-1, MDA-MB-231, and MDA-MB-453 were grown in Roswell Park Memorial Institute-1640 medium (RPMI-1640) with phenol red. The media for routine culture were supplemented with 10% (v/v) heat-inactivated FBS, 100 units/mL penicillin and 100 µg/mL streptomycin (M1 medium) in a humidified 37 °C incubator with 5% CO_2. Since, phenol red is reported to have estrogenic activity[249], for experiments involving estrogen treatment, phenol red-free DMEM/F12 or RPMI-1640 media supplemented with heat-inactivated csFBS, 100 units/mL penicillin and 100 µg/mL streptomycin (M2 medium) were used.

3.2.2. Sub-culturing and seeding

When cells were 90% confluent, the monolayer was washed with DPBS, treated with trypsin-EDTA and incubated until the cells detached from the surface. The cells were resuspended in 1 mL of M1 medium to inhibit trypsin. 200 µL of cell suspension was then reseeded into fresh cultures. For cell counting, 10 µL of cell suspension was mixed with 10 µL of trypan blue. This mixture was loaded in a hemocytometer and live cells that excluded the dye were counted. Cells were seeded in culture dishes or plates in varying densities

according to the surface area, doubling time and treatment duration, as mentioned in Table 3.2.

Table 3.2. Cell seeding density in this study.

Cell line	35 mm dish/ 6 well plate	100 mm dish
MCF-7	2 x 10^5 cells/well	1 x 10^6 cells/well
MDA-MB-231	4 x 10^4 cells/well	1 x 10^5 cells/well
MDA-MB-453	2 x 10^5 cells/well	1 x 10^6 cells/well
T47D	2 x 10^5 cells/well	4 x 10^5 cells/well
ZR-75-1	2 x 10^5 cells/well	1 x 10^6 cells/well

3.2.3. Treatment protocols

Dose-response experiments

MCF-7 cells were cultured till 70-80% confluence in M1 medium. M1 medium was replaced with M2 medium and incubated for 3 h. Spent M2 medium was removed and replaced with fresh M2 medium containing the indicated concentration of E2 or ethanol (vehicle control) and incubated at 37 °C for 72 h. pS2 was used as an indicator of E2 action in the present work as it is a well-documented estrogen-induced gene[250].

Time-course experiments

MCF-7 cells were cultured till 60-70% confluence in M1 medium. M1 medium was replaced with M2 medium for 3 h. Spent M2 medium was replaced with fresh M2 medium containing 10 nM E2 and incubated for 24, 48 or 72 h. Cells receiving M2 medium containing ethanol (vehicle) for 72 h served as controls. Alternatively, MCF-7 cells were also treated with 10 nM E2 for various duration ranging from 6 h to 96 h in which each group had individual vehicle-treated controls. In both the experiments, fresh M2 medium with 10 nM E2 was replenished every 48 h.

Effect of various hormones

MCF-7 cells were cultured till 70-80% confluence in M1 medium. M1 medium was replaced with M2 medium and incubated for 24 h. The spent M2 medium was replaced with fresh M2 medium containing various hormones and ligands such as E2, PPT, testosterone propionate, progesterone, dexamethasone and incubated for 24 h. Cells receiving M2 medium containing ethanol (vehicle) served as controls.

Effect of tamoxifen

MCF-7 cells were cultured in M1 medium and when the cells were 70% confluent, M1 medium was replaced with M2 medium. After 24 h, the cells were treated with M2 medium containing 10 nM E2, 1 μM tamoxifen or both for 24 h. Cells receiving M2 medium containing ethanol (vehicle) served as control.

Effect of fulvestrant

MCF-7 cells were cultured in M1 medium and when the cells were 60% confluent, M1 medium was replaced with M2 medium. After 24 h, the cells were treated with M2 medium containing 100 nM fulvestrant or ethanol for 3 or 24 h. Then, the medium was replaced with M2 medium containing 10 nM E2 or ethanol and incubated for 24 h.

Global demethylation using 5-aza

MDA-MB-231 or T47D cells were seeded in 100 mm dishes in M1 medium. After 24 h, the cells were treated with M1 medium containing 10 μM 5-aza for 5 days. Fresh M1 medium with 5-aza was replenished every 48 h. Cells treated with DMSO (vehicle) served as control.

E2 treatment on cells subjected to global demethylation

MDA-MB-231 or T47D cells were treated with 5-aza as described above, followed by incubation in M2 medium for 4 h. Thereafter, the cells were treated with 10 nM E2 or ethanol (vehicle) in M2 medium for 24 h.

Treatment of MCF-7 cells with E2 for chromatin immunoprecipitation (ChIP) assay

MCF-7 cells were seeded in 100 mm dishes in M1 medium. When the cells were 70% confluent, the spent medium was replaced with M2 medium and incubated for 24 h. Thereafter, the cells were treated with fresh M2 medium containing 10 nM E2 or ethanol (vehicle). After 24 h, the cells were harvested for ChIP assay.

Effect of ERα knockdown on estrogen modulation of CSTA expression

MCF-7 cells were cultured in M1 medium. When the cells were 60% confluent, cells were transfected with ERα siRNA or scrambled siRNA for 24 h, as described in section 3.3. This was followed by recovery in M2 medium. After 24 h, M2 medium containing 10 nM E2 or ethanol (vehicle) was added and incubated for 24 h.

3.3. siRNA transfection

MCF-7 cells were seeded in 6 well plates (4 x 10^5 cells per well) and incubated for 24 h. Cells were transfected with siRNA using Lipofectamine RNAiMAX for 24 h according to the manufacturer's instructions. siRNA and Lipofectamine RNAiMAX were individually diluted in Opti-MEM and an equal volume of both the components were mixed and incubated at room temperature for 10 minutes. 250 µL of the mixture was added dropwise to each well and swirled gently. Cells were then incubated for 24 h. Each well of the 6-well plate received 25 pmol of siRNA. In each well, siRNA to reagent ratio of 1:3 (v/v) was maintained.

3.4. Gene expression analysis

3.4.1. Primers

Primers were designed manually by considering the following criteria: length of 18-22 bases, 40-60% of GC content and melting temperature (T_m) of 55-65 °C. T_m was calculated using the formula $4 \times (G + C) + 2 \times (A + T)$. Primers used for routine RT-PCR and qRT-PCR were designed to hybridize different exons that are separated by a large intron in order to avoid amplification from genomic DNA. Details of the primers used in the present study are given in Appendix I.

3.4.2. Total RNA isolation and cDNA synthesis

Total RNA was extracted from cells using TRIzol (Invitrogen, CA, USA) or similar reagent prepared in house. The quality of RNA was checked by non-denaturing agarose gel electrophoresis. Total RNA was quantified using Biospectrometer (Eppendorf, Germany). 2 µg of total RNA obtained from the cells was reverse transcribed using the High-Capacity cDNA Reverse Transcription kit in a total reaction volume of 20 µL, according to manufacturer's instructions.

3.4.3. Routine RT-PCR

cDNA equivalent to 20 or 40 ng of total RNA was used as templates in PCR reactions with gene-specific primers. The PCR products were analyzed on 2% agarose gels and the images of ethidium bromide-stained bands were captured using Gel Doc™ EZ system or ChemiDoc XRS+ system (Bio-Rad laboratories, CA, USA). The images of the bands were processed and quantified with ImageJ[251]. The background-subtracted, and inverted images are shown in the Figures. The integrated densities of bands for the genes of interest were normalized against those obtained for cyclophilin A (CycA), which served as an internal

control. The normalized band intensities obtained for controls were assigned the value of 1 and those obtained for treatments were expressed relative to control. No-template control (water control) was included to detect the presence of contaminating DNA. The experiments were performed in 3 or 4 biological replicates. Each replicate comprising of one dish each for control and treated cells.

3.4.4. qRT-PCR

cDNA equivalent to 20 ng of RNA was amplified with gene-specific primers (Appendix I) in 7500 real-time PCR system (Applied Biosystems, CA, US) or Agilent AriaMx Real-time PCR System (Agilent, CA, US). Reactions were set up using AmpliTaq Gold PCR master mix (supplemented with 0.6X SYBR Green) or Powerup SYBR Green PCR master mix. ROX dye was used as passive reference. No-template control (water control) was included to detect the presence of contaminating DNA. Cyc A served as an internal control. Each sample was analyzed in triplicates. The expression levels of each gene in test samples relative to control were analyzed by the ΔΔCt method[252]. The experiments were performed in 3 or 4 biological replicates. Each replicate comprising of one dish each for control and treated cells.

3.5. Generation of polyclonal CSTA antibody

CSTA polyclonal antibody generation and peptide affinity purification were performed by Abgenex Pvt. Ltd. Bhubaneswar. Peptide used for antibody generation was designed using the software, AbDesigner. Peptide with high antigenicity was selected and chemically synthesized. The sequence of the peptide is - GQNEDLVLTGYQVDKNKDD. To facilitate Keyhole Limpet Haemocyanin (KLH, Pierce Cat#77600) conjugation, extra cysteine residue was added at the N-terminus. KLH was activated by sulfo-SMCC (Sulfosuccinimidyl 4-(N-maleimidomethyl) cyclohexane-1-carboxylate) (Pierce cat#22322) and purified by gel filtration chromatography. Then, it was mixed with peptide for conjugation. After collecting pre-immune sera, two New Zealand White Rabbits (A and B) were immunized with KLH-conjugated peptide (200 µg antigen/rabbit in Complete Freund's Adjuvant). This was followed by five boosters (100 µg antigen/ rabbit in Incomplete Freund's Adjuvant). After primary immunization, the first batch of immune sera was collected. The second and third batch of immune sera were collected after 6th and 7th boosters, respectively.

Hyper-immune serum was affinity-purified using CSTA peptide. CSTA peptide was covalently linked to Sulfo-Link coupling gel by passing 5 mL of peptide solution through 6

mL of sulfo-link coupling resin slurry, which was pre-equilibrated with coupling buffer. Diluted hyper-immune serum was passed through the column and incubated for 4 h at room temperature. Unbound components were washed and anti-CSTA antibodies were eluted using low pH elution buffer (50 mM Glycine-HCl, pH 2.7) and neutralized with 100 µL neutralization buffer (1 M Tris-HCl, pH 8.0, 1.5 M NaCl, 1 mM EDTA, 0.5% sodium azide). The eluted antibody was purified by dialyzing with 1 L phosphate-buffered saline (PBS). Sodium azide was then added to the antibody to the final concentration of 0.05%.

Indirect enzyme-linked immunosorbent assay (ELISA) was performed with immune sera and affinity-purified antibody to check the reactivity. CSTA peptide (200 ng/well) was coated in a 96-well plate for 2 h at room temperature followed by overnight incubation at 4 ºC. Next day, plates were incubated at room temperature for 2 h, followed by washing with PBS containing 0.05% Tween 20 (PBST). Peptide-coated wells were blocked with 5% skimmed milk in PBST for 1 h, followed by washing with PBST. Pre-immune serum, hyper-immune serum or peptide-affinity-purified antibody were diluted (1:5000) with 1% skimmed milk. 100 µL was added to each well and incubated for 2 h at room temperature. The wells were further washed and incubated with 100 µL of horseradish peroxidase (HRP) conjugated secondary antibody (1:5000) for 1 h at room temperature. After washing, 100 µL of 1X TMB (3,3',5,5'-tetramethylbenzidine)/H_2O_2 solution was added to the wells and incubated for 3-5 min in the dark. Then, the absorbance was measured at 450 nm.

3.6. Western blotting

Total protein was isolated from cells with Laemmli sample buffer, Triton-X lysis buffer (20 mM HEPES-KOH (pH 7.4), 2 mM EDTA, 250 mM NaCl, 0.1% Triton X-100, 1 mM DTT, 1X Protease Inhibitor Cocktail) or from organic phase of TRIzol lysates, as per manufacturer's instructions. Total protein was then quantified by TCA method[253] or Lowry's method[254]. Protein samples (30 µg) were resolved by 8 or 12% PAGE, transferred to 0.45 µm or 0.22 µm nitrocellulose membrane, and blocked with 1% gelatin in tris-buffered saline containing 0.05% Tween 20 (TBST) for 2 h. Blots were probed with CSTA, ERα, PR, β-actin, histone H3, or EMT marker antibodies overnight in TBST containing 0.1% gelatin. The blots were washed with TBST (6 × 5 min). Blots were then probed with anti-rabbit or anti-mouse HRP-conjugated secondary antibody for 1 h, washed with TBST (6 × 5 min) and developed using Clarity Western ECL Substrate (Bio-Rad, California, US). Images were captured with ChemiDoc XRS+ system (Bio-Rad, California, US). β-actin or histone H3 served as an internal control.

3.7. Chromatin immunoprecipitation (ChIP)

Cells were fixed with formaldehyde (0.75%) for 10 minutes. The reaction was stopped by 125 mM glycine for 10 minutes. Cells were washed and scraped with ice-cold DPBS, pelleted, lysed with lysis buffer (50 mM HEPES pH 7.5, 150 mM KCl, 1 mM EDTA, 10% glycerol, 0.1% NP-40), and sonicated. Lysates were clarified by centrifugation, and supernatants containing chromatin were collected. Chromatin samples were precleared with Protein G plus-Agarose beads precoated with Bovine serum albumin (BSA) and herring sperm DNA for 2 h. 5% of the pre-cleared chromatin samples were separated as input. The remaining portions were incubated with ERα antibody or rabbit IgG antibody for 2 or 4 h. Samples were immunoprecipitated by incubating with 20 µL of coated beads for 2 h, followed by centrifugation. Immunoprecipitates were washed with a series of wash buffers[255] and eluted in 300 µL elution buffer containing proteinase-K for 2 h at 55 °C, and overnight incubation at 65 °C. Immunoprecipitated DNA was column purified and ERα occupancy was assessed by PCR with primers specific to pS2 (positive control) or Region 2 of CSTA locus (Appendix I).

3.8. ChIP-Seq analysis

Raw data of ChIP-Seq experiments were retrieved from Sequence Read Archival (SRA) and analyzed using Galaxy, a web-based platform[256]. ChIP-Seq data (SRA accession ID: ERP000380) of chromatin samples from MCF-7 cells treated with E2 (ID: ERR022026), tamoxifen (ID: ERR022027) or vehicle (ID: ERR022025) and immuno-precipitated with ERα was chosen for this study. IgG control ChIP-seq data was used as negative control. Using FASTQC tool[257], the quality of input reads was assessed. After converting the quality score to sanger quality type by FASTQ Groomer[258], reads were mapped to reference human genome (hg19) using "Map with Bowtie for Illumina" tool[259]. Unmapped reads were discarded by "Filter SAM (Sequence Alignment/Map) or BAM (Binary version of SAM), output SAM or BAM" tool[260]. Genomic regions with enriched sequencing reads were identified by MACS (Model-based analysis of ChIP-Seq) tool[261]. Resultant Wig files were converted to bigWig files using "Wig/BedGraph-to-bigWig" tool and the peaks representing ERα occupancy were visualized using UCSC genome browser[262].

3.9. Bisulfite sequencing

Genomic DNA (gDNA) was isolated from breast cancer cell lines. Two μg of gDNA was bisulfite converted and purified with EpiJET Bisulfite Conversion Kit. Fifty ng of converted gDNA was used for PCR with specific primers (Appendix I) that amplified Region 1 and Region 2 (described in chapter 6), which encompassed few CpG dinucleotides of the upstream region and intron-2, respectively. The primers were designed to amplify only the bisulfite converted DNA. Importantly, the priming region did not contain any CpG dinucleotides. The PCR products were gel purified and cloned in pMD20 vector and sequenced. The inserts of 12 - 15 independent clones per cell line were sequenced. The sequencing results were analyzed to determine methylated and unmethylated CpG sites and represented as lollipop plots. The proportion of CpGs methylated in Region 1 or Region 2 for each cell line was determined.

3.10. Cloning of CSTA ORF in mammalian expression vector

CSTA ORF-specific primers were designed with 25 bp of overlapping vector sequence. Details of primers are provided in Appendix I. CSTA ORF was amplified using MCF-7 cDNA as template. Vectors were digested with B*am*HI and S*al* I and gel purified to remove primer dimers. Purified insert was cloned in pBABE-puro vector by incubating 50 ng of vector and 120 ng of insert with Gibson Assembly master mix at 50 ºC for 50 minutes[263]. Following incubation, the mixture was transformed into DH5α cells. The clone was confirmed by PCR followed by Sanger sequencing.

3.11. Establishment of stable cell lines

MDA-MB-231 cells were seeded in 35 mm dish (4 x 10^5 cells per dish) and incubated for 24 h. Then, cells were transfected with CSTA expression construct, or empty pBABE-puro vector using Lipofectamine 3000 as per manufacturer's instruction for 24 h. 2.5 μg of DNA was used for transfecting cells per dish. Lipofectamine 3000 and DNA were individually diluted in Opti-MEM. P3000 enhancer was added to the diluted DNA in the ratio of 2:1. An equal volume of both the components was mixed and incubated at room temperature for 10 minutes. 250 μL of the mixture was added dropwise to each well and swirled gently. Cells were then incubated for 48 h. After 48 h, puromycin (2 μg/mL) was added to the medium for the selection of stably transfected cells. Subsequently, colonies were picked and propagated to obtain cells stably expressing CSTA. Total RNA was extracted from each clone and DNase

digested to eliminate gDNA contamination followed by qRT-PCR analysis. Similarly, total protein was isolated and subjected to western blotting analysis to detect the expression levels of CSTA in different clones.

3.12. Functional assays

3.12.1. MTT assay

Cells were seeded in 96 well plates (5000 cells/well) in M1 medium and grown for 72 h . The spent medium was removed and 100 µL of MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide) reagent (0.5 mg/mL) was added, followed by incubation at 37 ºC for 3 h. After 3 h, MTT reagent was removed and formazan crystals were dissolved in 100 µL of DMSO. Absorbance was measured at 570 nm and 690 nm (for background correction). The difference in the absorbance was considered as a measure of cell proliferation.

3.12.2. Scratch wound healing assay

Cells were seeded in 35 mm dish (4 x 10^5 cells per dish) and grown to form a monolayer. Then the cells were serum-starved with 0.5% serum containing medium for 12 h followed by mitomycin C (5 µg/mL) treatment for 3.5 h to inhibit cell proliferation. Three scratches were made in each plate using 200 µL micropipette tip. Cells were washed to remove detached cells. Scratches were imaged at 0 h using light microscope at 10X magnification. Fresh medium containing 0.5% FBS was added and incubated for 24 h. After 24 h, scratches were imaged again. The extent of wound closure was quantified using ImageJ software. Relative wound area was calculated using the formula $(A_0 - A_{24})/A_0$, where A_0 is the area of wound at 0 h and A_{24} is the area of wound after 24 h.

3.12.3. Transwell invasion assay

Transwell invasion assay was performed using Transwell® polycarbonate membrane cell culture inserts with 8 µm membrane (Corning, NY, US). Inserts were coated with collagen IV (100 ng/mL in 0.05 N HCl) for 4 h at 37 ºC. Cells were seeded (40000 cells per insert) in the upper chamber of collagen IV-coated transwell insert in serum-free medium and placed in 24-well plates. The lower chamber contained either serum-free medium or medium containing 10% FBS. The cells were allowed to migrate through the pores of membrane for 16 h. The upper side of the membrane was wiped with cotton swabs to remove non-migrated cells. The cells migrated to the other side of the membrane were fixed with paraformaldehyde

and stained with 0.025% crystal violet. The stained cells were imaged in the microscope and then crystal violet stain was dissolved in 1% SDS solution at 25 ºC for 10 minutes. Absorbance was measured at 595 nm.

3.13. Survival analysis

Survival analysis was performed by using the KM Plotter online tool (www.kmplot.com)[264]. Tumor samples were divided into CSTA-high and CSTA-low groups using the "auto select best cutoff" option. The association between CSTA expression and survival was graphically represented as Kaplan-Meier (KM) plots for overall survival (OS), relapse-free survival (RFS) and distant metastasis-free survival (DMFS) using the default settings. Survival analysis was also performed for each of the molecular subtypes of breast tumors, namely luminal A, luminal B, HER2+ and basal.

3.14. TCGA data analysis

The Cancer Genome Atlas (TCGA) is a publicly available cancer genomics project in which 20,000 primary cancer and matched normal samples across 33 cancer types are molecularly characterized[265]. Gene expression data from RNA-Seq (Illumina HiSeq 2000 RNA Sequencing platform) and methylation data generated with Illumina Infinium® Human Methylation 450K BeadChip array from TCGA-BRCA (Breast Invasive Carcinoma) dataset were used for analysis. RNA-Seq data were in terms of $\log_2$(RPKM+1) and methylation data were in terms of beta values. Datasets were accessed through the UCSC Xena browser (https://xenabrowser.net/)[266].

3.14.1. Analysis of CSTA expression in normal breast tissues and primary breast tumors

Samples were segregated into normal breast tissues and primary breast tumors. CSTA mRNA expression values corresponding to 113 normal breast tissues and 1095 primary breast tumors were analyzed by Welch two-sample *t*-test. The distribution of CSTA expression was represented as box plots.

3.14.2. Analysis of CSTA expression in different subtypes and stages of breast tumors

Primary breast tumors were classified based on either molecular subtype (PAM50) or stages (AJCC stage). The distribution of CSTA and ERα expression was represented as box

plots and difference in the mean CSTA and ERα expression in the groups were analyzed by one-way ANOVA (analysis of variance) followed by Tukey's HSD.

3.14.3. Analysis of association between CSTA expression and histopathological parameters

To ascertain the relationship between CSTA and ERα mRNA expression, the primary breast tumors were divided into two groups, namely ERα-high and ERα-low based on RNA-Seq data with the median ERα expression value serving as the threshold. The difference in CSTA expression in these two groups was analyzed by Welch two-sample *t*-test and the expression data were represented as box plots. Further, tumor samples were also classified as positive or negative groups based on immunohistochemical expression of ER, PR or HER2. CSTA expression in these groups were analyzed by Welch two-sample *t*-test. The distribution of CSTA expression in the groups was represented as box plots. To evaluate the association of CSTA expression with histopathological parameters, tumors were classified as CSTA-high and CSTA-low using median value as cut-off and analyzed by non-parametric chi-square test in Microsoft Excel.

3.14.4. Expression-methylation correlation (EMC) analysis

Methylation data (generated with Illumina Infinium® Human Methylation 450K BeadChip array) and CSTA expression (RNA-Seq) data were retrieved from TCGA database[265] using the UCSC Xena browser[266]. CSTA expression data were available for 1218 samples. Out of these, 873 samples (85 normal and 788 tumors) had both expression and methylation data. Only the tumor samples were used for analysis. Data were processed in MS Excel. Correlation between methylation (beta values) and CSTA expression ($\log_2$(RPKM+1)) was assessed by Spearman's rank correlation test. Further, tumors were classified into hyper-methylated and hypo-methylated based on a threshold beta-value of 0.3. CSTA expression in hyper-methylated and hypo-methylated tumors was represented as box plots. The difference in CSTA expression in these two groups was analyzed by Welch two-sample *t*-test.

3.15. Statistical analysis

All the quantitative data are represented as mean ± S.D. Agarose gel images and chemiluminescence images were quantified by ImageJ software[251]. The association of CSTA expression with histopathological parameters was analyzed by non-parametric chi-square test. The difference in the means of two groups was analyzed by Welch two-sample *t*-test. Multiple groups were analyzed by one-way ANOVA followed by Tukey's HSD. Bisulfite sequencing

data were analyzed for statistical differences in proportions of methylated CpGs for every pair of cell lines. The total number of CpGs sampled for each cell line was greater than 30. Methylation at each CpG site was assumed to be independent of the adjacent CpGs. Under the null hypothesis, the mean of the sampling distribution of the difference in proportions follows a standard normal (z) distribution. The test statistic d was calculated as,

$$d = \frac{k_1 - k_2}{\sqrt{(k)(1-k)(\frac{1}{n_1} + \frac{1}{n_2})}}$$

where, $k_1 - k_2$ is the observed difference in proportion in a pair of cell lines, k is the proportion for the pooled data of the pair of cell lines, and n_1 and n_2 are the number of CpGs sampled. The probability of d was obtained from z distribution. The *p*-value obtained for each pair of cell lines was subjected to Bonferroni correction. Correlation between methylation score and CSTA expression was analyzed by Spearman's correlation test. All the statistical analyses were performed using R statistical package. In all the statistical tests, $p < 0.05$ was considered as significant.

CHAPTER 4

CSTA expression in breast cancer

4.1. Introduction

Tumor invasion and metastasis are the major causes of cancer-related mortality. Despite recent advances in the treatment of cancer, metastasis remains a steadfast challenge attributing to the death of many cancer patients[267]. Proteolytic cleavage of basement membrane and remodeling of surrounding ECM are prerequisites in invasion and metastasis of solid tumors[268,269]. Breakdown of normal ECM and replacement with tumor ECM in the microenvironment primes malignant progression. This is facilitated by matrix-degrading enzymes such as MMPs, ADAMs, ADAMTSs, cathepsins, heparanases, hyaluronidases, matriptases, uPA, and tPA. In non-neoplastic conditions, activities of matrix remodeling enzymes are tightly regulated by their endogenous inhibitors such as TIMPs, PAIs, and cystatins. In pathological conditions including cancer, aberrant expression of proteases for prolonged period disturbs the intricate balance between proteases and their respective inhibitors, resulting in degradation of ECM and basement membrane. This, in turn, contributes to invasion and metastasis[148].

CSTA, a member of the class I family of cystatins, is expressed in diverse cell types and tissues [207,209,211]. It is a physiological reversible inhibitor of cysteine cathepsins B, H, L, and papain. CSTA protects the degradation of cytosolic and cytoskeleton proteins from accidentally released cathepsins from lysosomes[154]. Expression of CSTA is diminished or lost

in various forms of cancer, including head and neck squamous cell carcinomas[270], brain tumors[20], prostate[19], and esophageal squamous cell carcinoma[271]. On the other hand, increased expression and activity, or mislocalization of cathepsins is known in several types of cancer[173]. Dysregulated expression of cysteine cathepsins when not balanced by cystatins are likely to play an important role in the malignant progression of tumors[272].

Literature presents contradictory views on the role of CSTA in breast cancer. CSTA mRNA and protein is reduced in the majority of breast carcinoma tissue compared to matched normal breast tissues[30,226]. Parker and co-workers reported that CSTA expression correlates with disease-free survival, and distant metastasis-free survival. These studies suggest tumor suppressor role for CSTA[34]. On the contrary, based on immunohistochemical analysis, Kuopio and co-workers[31] showed an association of CSTA expression with an aggressive phenotype. Levicar and co-workers reported a 1.9 fold higher levels of CSTA in cytosols of primary invasive breast tumors compared to normal breast parenchyma[32].

Prediction of clinical outcome is of prime importance in the management of any cancer. This makes the identification of genomic factors that have a prognostic impact on clinical outcomes essential. A prognostic factor predicts the chance of recovery from the disease or the chance of disease relapse[273]. Matrix remodeling proteases and their inhibitors are considered as potential prognosticators in survival analysis[274]. In line with this, CSTA has been contemplated as an important prognostic factor[34]. However, the reported information of prognostic value of CSTA is not consistent across the literature. Therefore, to appraise the prognostic potential of CSTA, in the present study, it was independently assessed by taking various molecular subtypes into consideration.

Therapeutic decisions for breast cancer are based on the status of ER, PR and HER2 expression[29]. About two-thirds of breast tumors are ER- or PR-positive[28]. These molecular markers are well-defined predictive factors which predicts for responsiveness of tumors to endocrine therapy[275]. However, so far, no study has been done on the association of CSTA with these molecular markers of breast cancer.

This study mined the TCGA-BRCA data to independently assess the prognostic potential of CSTA, by analyzing its expression in primary breast tumors, and its association with histopathological markers.

4.2. Results

4.2.1. Association of CSTA expression with breast cancer prognosis

To assess the prognostic value of CSTA, survival analysis was performed with respect to CSTA expression. The Kaplan-Meier survival analysis of TCGA-BRCA data showed that higher CSTA expression in breast tumors is associated with reduced OS, RFS and DMFS with hazard ratio of 1.47 (95% CI = 1.17-1.86, $p < 0.001$), 1.37 (95% CI = 1.22-1.54, $p < 0.0001$) and 1.4 (95% CI = 1.14-1.71, $p = 0.0012$), respectively (Figure 4.1). This analysis was performed by considering all the tumors irrespective of the tumor subtype.

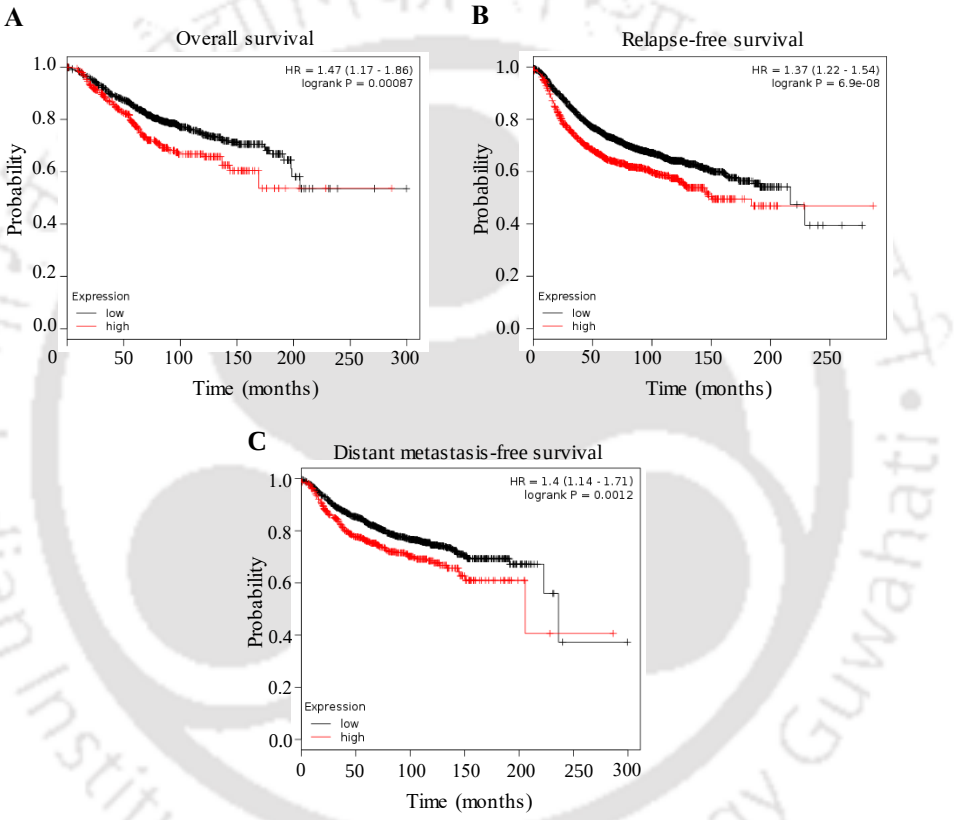

Figure 4.1. Kaplan-Meier survival analysis for OS, RFS and DMFS with respect to CSTA. Plots were generated in using Kaplan-Meier plotter (www.kmplot.com). The breast tumors in each of the groups were divided into two groups, CSTA-high and CSTA-low using "auto select best cutoff" option.

Survival analyses with respect to CSTA expression in the various molecular subtypes of breast tumors produced interesting results. CSTA expression was not associated with OS, RFS, or DMFS in HER2+ and basal tumor subtypes. In luminal A, higher CSTA expression was associated with reduced OS and RFS with hazard ratio of 1.74 (95% CI = 1.21–2.5, $p = 0.0027$) and 1.36 (95% CI = 1.14–1.62, $p = 0.00048$), respectively (Figure 4.2). Interestingly, in luminal B, higher CSTA expression is associated with prolonged OS and DMFS with hazard ratio of 0.63 (95% CI = 0.43–0.92, $p = 0.015$) and 0.69 (95% CI = 0.49–0.99, $p = 0.041$),

respectively (Figure 4.2). Thus, the effect of CSTA on survival appears to be tumor subtype dependent.

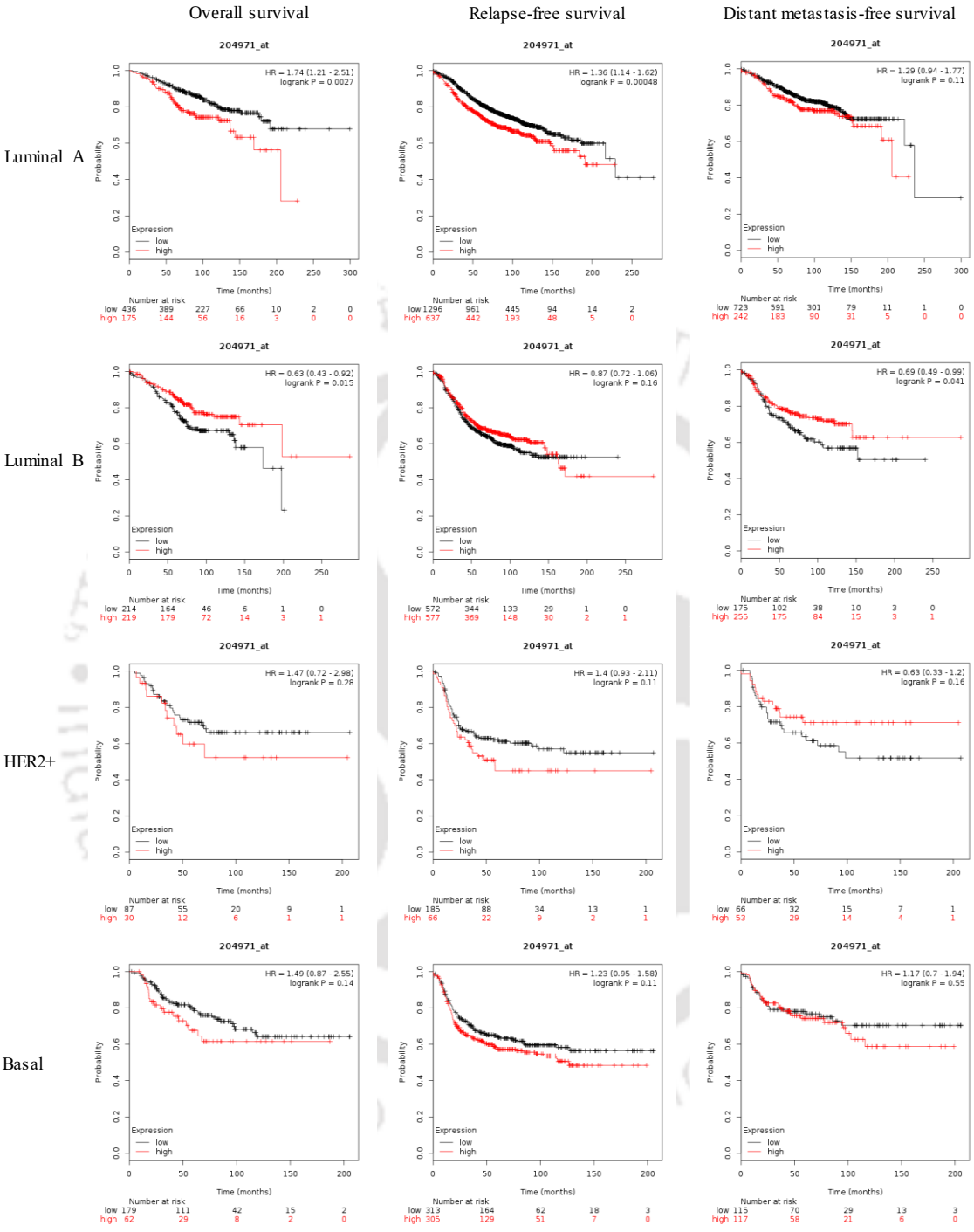

Figure 4.2. Kaplan-Meier survival analysis for OS, RFS and DMFS with respect to CSTA in breast tumors of various molecular subtypes. Plots were generated for molecular subtypes of breast tumors, namely luminal A, luminal B, HER2+ and basal using Kaplan-Meier plotter. The breast tumors in each of the groups were divided into two groups, CSTA-high and CSTA-low using "auto select best cutoff" option.

4.2.2. Analysis of CSTA expression in normal breast tissues and primary breast tumors

CSTA expression data (log_2(RPKM+1) values) corresponding to normal breast tissues and primary tumors of TCGA-BRCA dataset was assessed through UCSC Xena browser. The mean CSTA expression in normal breast tissues (7.85 ± 0.82) was significantly higher than in primary breast tumors (6.52 ± 1.77) (Figure 4.3).

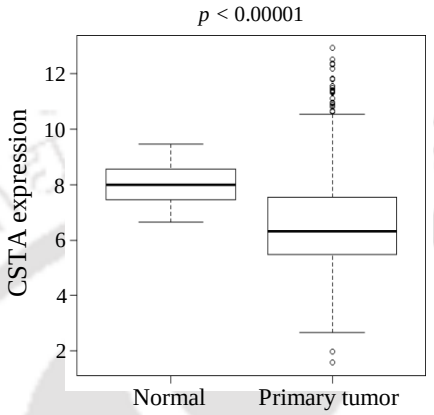

Figure 4.3. Expression of CSTA mRNA in normal breast tissues and breast tumors. Box plots showing the distribution of CSTA mRNA expression in normal breast tissues and breast tumors. The difference in the mean expression values of groups was analyzed by Welch two-sample *t*-test. *p*-value is mentioned above the figure.

4.2.3. Analysis of CSTA expression in molecular subtypes and stages of breast tumors

CSTA and ERα mRNA expression in molecular subtypes of primary breast tumors was analyzed. The mean CSTA expression was significantly lower in luminal A (6.12 ± 1.49) and luminal B (6.38 ± 2.03) subtypes compared to HER2+ (7.57 ± 2.23) and basal (7.20 ± 1.64). The mean ERα expression was significantly higher in luminal A (13.40 ± 1.30) and luminal B (13.60 ± 1.08) compared to HER2+ (8.31 ± 2.32), and basal (6.50 ± 2.10). Mean CSTA expression in normal-like was significantly higher than luminal A ($p < 0.00001$) and luminal B ($p < 0.00001$). The mean ERα expression in normal-like was lower than luminal A ($p < 0.00001$) and luminal B ($p < 0.00001$). The distribution of CSTA and ERα expression in the molecular subtypes are shown as box plots in Figure 4.4. The data were analyzed by ANOVA followed by Tukey's HSD test. The adjusted *p*-values for the pairwise comparison of group means are provided in Table 4.1 and 4.2. CSTA expression was analyzed in various stages of breast cancer. No significant difference in mean CSTA expression across different stages was observed (Figure 4.5).

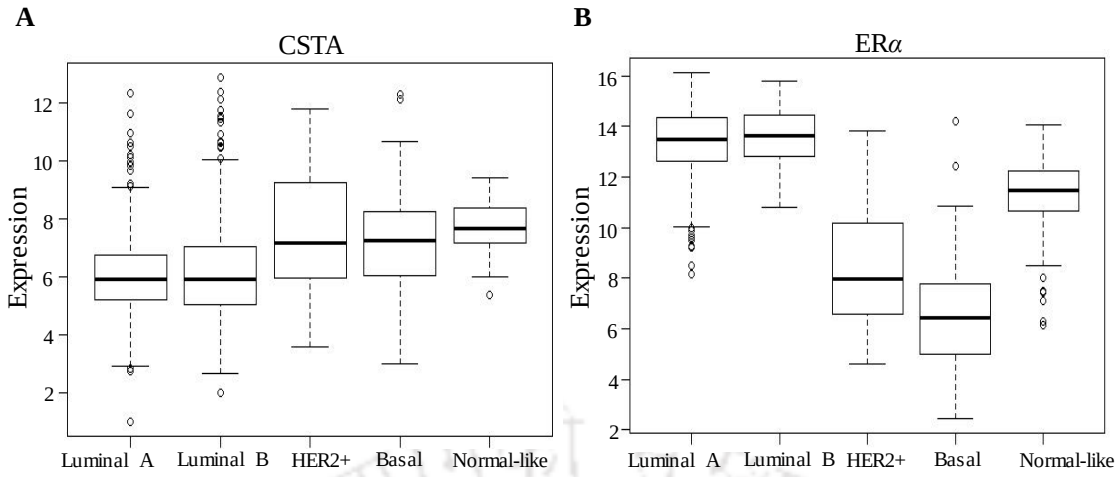

Figure 4.4. Expression of CSTA and ERα mRNA in molecular subtypes of breast tumors. Box plots showing the distribution of CSTA (A) and ERα (B) mRNA expression in the indicated subtypes of primary breast tumors. The data were analyzed by ANOVA followed by Tukey's HSD.

Table 4.1. Analysis of CSTA expression in molecular subtypes of breast tumors.

Comparison	Diff	Lwr	Upr	*p*.adj
HER2+-Basal	0.372329	-0.28845	1.033112	0.5366320
Luminal A-Basal	-1.08232	-1.51333	-0.65131	**0.0000000**
Luminal B-Basal	-0.81409	-1.30646	-0.32172	**0.0000685**
Normal.like-Basal	0.51433	-0.03975	1.068407	0.0834569
Luminal A-HER2+	-1.45465	-2.03985	-0.86945	**0.0000000**
Luminal B-HER2+	-1.18642	-1.81817	-0.55466	**0.0000034**
Normal.like-HER2+	0.142001	-0.53895	0.822948	0.9793901
Luminal B-Luminal A	0.268229	-0.11681	0.653265	0.3158598
Normal.like-Luminal A	1.596647	1.135318	2.057977	**0.0000000**
Normal.like-Luminal B	1.328418	0.809301	1.847535	**0.0000000**

Diff: difference between means of the two groups, Lwr, Upr: the lower and the upper end point of the confidence interval at 95%. *p*.adj: adjusted *p*-value. The significant *p*.adj values are indicated in bold.

Table 4.2. Analysis of ERα expression in molecular subtypes of breast tumors.

Comparison	Diff	Lwr	Upr	*p*.adj
HER2+-Basal	1.805579	1.191903	2.419255	**0.0000000**
Luminal A-Basal	6.891932	6.491646	7.292217	**0.0000000**
Luminal B-Basal	7.096611	6.639341	7.55388	**0.0000000**
Normal.like-Basal	4.755636	4.241059	5.270213	**0.0000000**
Luminal A-HER2+	5.086353	4.542872	5.629833	**0.0000000**
Luminal B-HER2+	5.291032	4.704313	5.87775	**0.0000000**
Normal.like-HER2+	2.950057	2.317655	3.582459	**0.0000000**
Luminal B-Luminal A	0.204679	-0.15291	0.562266	0.5207686
Normal.like-Luminal A	-2.1363	-2.56474	-1.70785	**0.0000000**
Normal.like-Luminal B	-2.34097	-2.82308	-1.85887	**0.0000000**

Diff: difference between means of the two groups, Lwr, Upr: the lower and the upper end point of the confidence interval at 95%. *p*.adj: adjusted *p*-value. The significant *p*.adj values are indicated in bold.

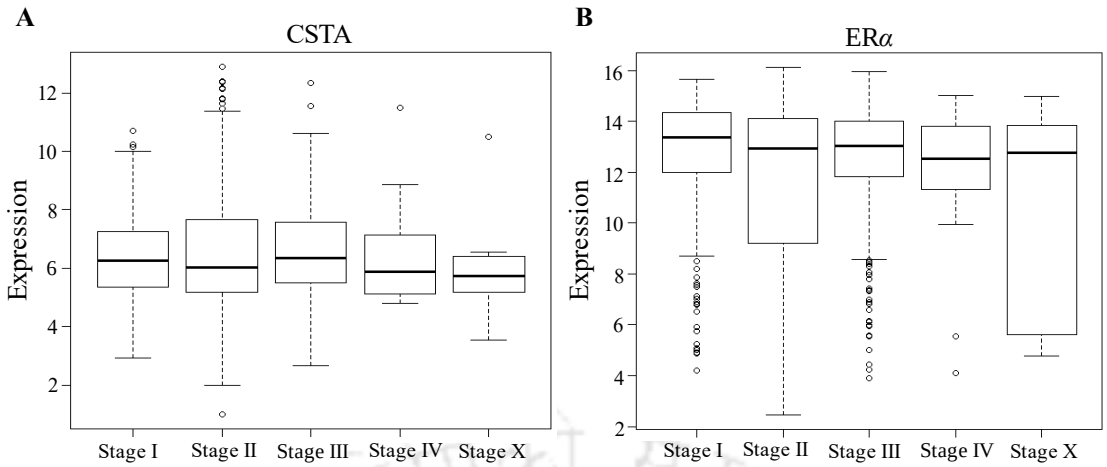

Figure 4.5. Expression of CSTA and ERα mRNA in various stages of breast cancer. Box plots showing the distribution of CSTA (A) and ERα (B) mRNA expression in the different stages of breast cancer. The data were analyzed by ANOVA followed by Tukey's HSD.

4.2.4. Association of CSTA expression with histopathological parameters

CSTA expression was analyzed in primary tumors classified based on the immunohistochemistry (IHC) data for ERα, PR or HER2 status. The boxplots in Figure 4.6B-D show the distribution of CSTA expression in ER-positive and -negative, PR-positive and -negative, and HER2-positive and -negative tumors. CSTA expression of ERα-positive tumors (6.31 ± 1.71) was significantly lower than ERα-negative tumors (7.28 ± 1.73) ($p < 0.00001$). Similarly, PR-positive tumors (6.27 ± 1.65) expressed significantly lower levels of CSTA compared to PR-negative tumors (7.06 ± 1.87) ($p < 0.00001$). However, no significant difference was observed between HER2-positive tumors (6.76 ± 1.91) and HER2-negative tumors (6.51 ± 1.74) ($p = 0.14$). Besides the IHC data, tumors were also segregated as ERα-high and ERα-low based RNA-Seq data using median as cut-off and CSTA expression was analyzed. CSTA expression in ERα-high tumors (6.00 ± 1.69) was significantly lower than that in ERα-low tumors (7.03 ± 1.70) ($p < 0.00001$) (Figure 4.6A).

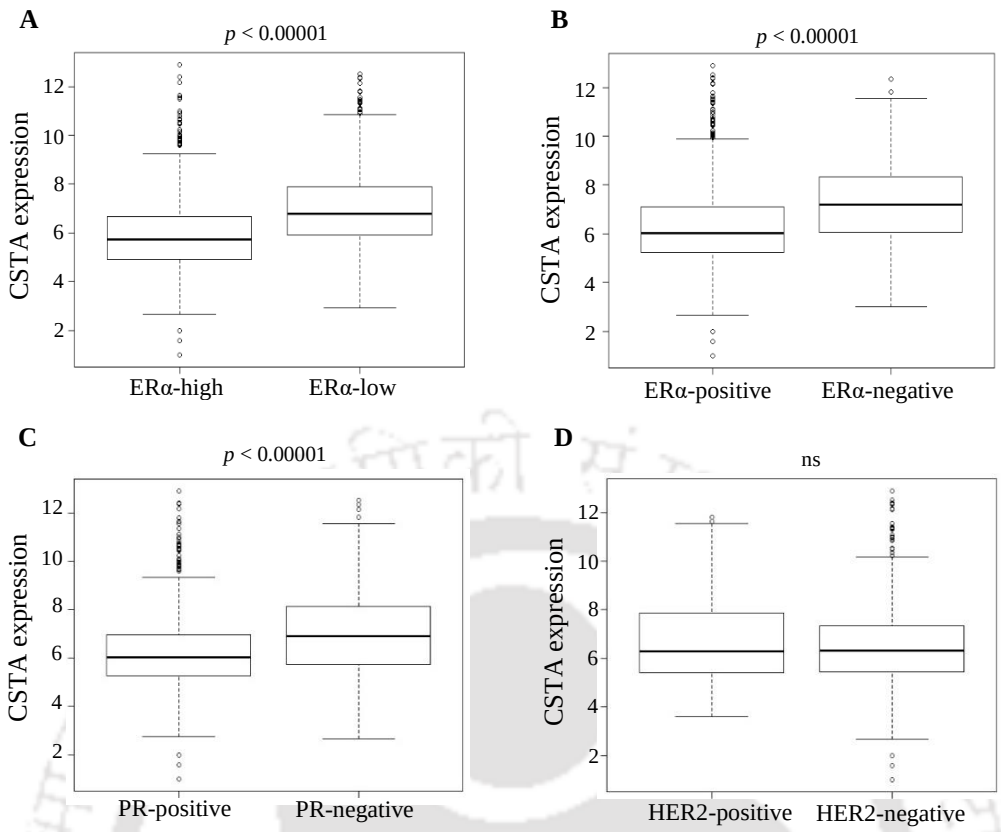

Figure 4.6. Expression of CSTA mRNA in primary breast tumors. Box plots showing the distribution of CSTA mRNA expression in primary breast tumors: A. ERα-high and ERα-low (tumors were divided based on RNA-Seq data), B. ER-positive and ER-negative, C. PR-positive and PR-negative, D. HER2-positive and HER2-negative. Tumors were divided based on IHC data in B-D. The difference in the mean CSTA expression in two groups was analyzed by Welch two-sample *t*-test. *p*-value is mentioned above the panels. ns = not significant

Further, the primary tumors were divided into two groups: CSTA-high and CSTA-low, using median value as a cutoff. Then, the association of CSTA with histopathological parameters was analyzed by non-parametric chi-square test. The mean age of patients in the two groups were not significantly different. CSTA-high tumors were associated with ERα-negative (70.38%) or PR-negative status (63.98%) (Table 4.3). CSTA-low tumors were more frequent in luminal A (61.04%), luminal B (55.72%) subtype, while CSTA-high tumors were frequent in HER2+ and basal subtype. CSTA expression was significantly associated with ERα ($p < 0.0001$), PR ($p < 0.0001$), and molecular subtypes ($p < 0.0001$) of breast cancer. No significant association of CSTA expression was found with HER2 status or tumor stage (Table 4.3).

Table 4.3. Association of CSTA expression with various histopathological parameters.

	CSTA-high	CSTA-low	*p*-value
Age			
Mean ± S.D.	57.7 ± 13.0	59.1 ± 13.2	T: 0.0729
Median	57	59	
Range	26-90	26-90	
ERα			
ERα-positive	353 (44.51)	440 (55.48)	**<0.0001**
ERα-negative	164 (70.38)	69 (29.61)	
PR			
PR-positive	301 (43.81)	386 (56.18)	**<0.0001**
PR-negative	215 (63.98)	121 (36.01)	
HER2			
HER2-positive	84 (52.17)	77 (47.82)	0.9860
HER2-negative	290 (52.25)	265 (47.74)	
Molecular type			
Luminal A	164 (38.95)	257 (61.04)	**<0.0001**
Luminal B	85 (44.27)	107 (55.72)	
Basal	97 (68.72)	44 (31.20)	
HER2-enriched	48 (71.64)	19 (28.35)	
Normal-like	21 (91.30)	2 (8.69)	
Tumor Stage			
Stage I	96 (52.74)	86 (47.25)	0.2522
Stage II	298 (48.85)	312 (51.14)	
Stage III	130 (53.27)	114 (46.72)	
Stage IV	8 (42.10)	11 (57.89)	
Stage X	4 (28.57)	10 (71.42)	

Number within the braces indicates percentage of CSTA-high or -low in various categories. *p*-values were obtained from non-parametric chi-square test except for age wherein *p*-value (T) was obtained from student's *t*-test. In all the tests, $p < 0.05$ was considered as significant

4.3. Discussion

CSTA is the least studied cystatin. The scanty literature on CSTA presents contradictory views on its role in breast cancer. Kuopio and co-workers reported association of CSTA expression with the aggressive phenotype[31]. On the other hand, a study of an independent cohort of breast tumors by Parker and co-workers (2008) showed an association of CSTA with prolonged DMFS[34]. The ambiguity in the apparent role of CSTA in breast cancer and its effect on prognosis is evident. Therefore, in this chapter, the prognostic value of CSTA was independently analyzed by Kaplan-Meier survival analysis with respect to CSTA. Higher CSTA expression in breast tumors, regardless of the expression of markers, is associated with reduced OS, RFS and DMFS (Figure 4.1). This is consistent with the reported association of CSTA expression with the aggressive phenotype and poor prognosis by Kuopio and co-workers[31].

Survival analyses with respect to CSTA expression of each subtype of breast tumors produced interesting results. CSTA expression was not associated with OS, RFS or DMFS in HER2+ and basal tumor subtypes. In luminal A, higher CSTA expression was associated with reduced OS and RFS (Figure 4.2). This mirrors the results of the survival analysis of the entire primary tumors data and also emulates the results reported by Kuopio and co-workers[31]. Interestingly, in luminal B, higher CSTA expression is associated with prolonged OS and DMFS (Figure 4.2). This mirrors the results reported by Parker and co-workers[34]. Thus, the effect of CSTA on survival appears to be tumor subtype dependent. The disparate reports on CSTA are possibly due to the inherent differences in the cohorts under study and the methodologies. Unlike luminal A, a luminal B subtype has high proliferation index and a subcategory of luminal B is HER2-positive[60]. Whether this phenotypic difference is the underlying reason behind the observed difference in survival of patients with luminal A and luminal B tumors is a matter of conjecture that must be investigated.

Genes playing dual roles (prevention and promotion) at different stages of disease progression are known in the literature; for instance, TGF-β and Runt-related transcription factor (RUNX) family of proteins[276-279]. Could CSTA have a dual role in breast cancer progression depending on the stage or molecular profile of tumors? CSTA expression does not appear to be very different in various stages of breast tumors (Figure 4.5). However, the mean CSTA expression was found to be different in the molecular subtypes. Luminal A and luminal B subtypes expressed lower levels of CSTA compared to HER2+ and basal tumors (Figure 4.4). The ambiguity in the apparent role of CSTA in breast cancer development and progression possibly indicates a dual role: as a tumor suppressor and as a promoter of aggressive phenotype, depending on the breast cancer subtype.

TCGA-BRCA dataset revealed that the mean CSTA expression in normal breast tissues is significantly higher than that in primary breast tumors (Figure 4.3). A similar observation was reported by other research groups that the expression of CSTA is lost during tumorigenesis in different types of cancer, including, prostate, brain, head and neck cancer[20,178,226,270]. Differential expression of CSTA in normal breast tissues and primary tumors, and other clinical and *in vivo* studies[30,106,226] uphold the tumor suppressor role of CSTA in breast cancer. This may appear to contradict the inference from the survival analyses of breast tumors irrespective of the molecular subtype (Figure 4.1). However, it is important to note that differential expression of CSTA in normal breast tissues and primary tumors does not have any bearing on survival analysis, which is performed on data corresponding to primary

tumors only. The results of the survival analyses only indicate that subjects with CSTA-low primary breast tumors are expected to survive longer than those with CSTA-high primary tumors.

Attempts have been made earlier to correlate CSTA expression with the known histopathological and clinical markers of breast cancer[30-32]. However, despite the fact that three-quarters of the newly diagnosed breast tumors are ERα-positive[280], the correlation between CSTA and ERα expression had not been studied. This study demonstrates the inverse correlation between CSTA and ERα expression in breast tumors. The mean CSTA expression in ERα-high primary tumors was significantly lower than that observed in ERα-low primary tumors. Furthermore, the inverse relationship is also apparent from the relative levels of ERα and CSTA expression in the molecular subtypes of breast tumors (Figure 4.4). Analysis of the association of CSTA mRNA expression with histopathological parameters revealed that CSTA was significantly associated with ERα, PR status and with various molecular subtypes of breast cancer (Table 4.3). Moreover, CSTA-high tumors were more frequently observed in HER2+ and basal subtype but less in luminal A or luminal B subtype. Luminal A and luminal B subtypes are ER and PR positive but not basal or HER2+ subtype. Inverse correlation of CSTA expression with ERα and PR expression may be the underlying reason for the frequent occurrence of CSTA-high tumor in HER2+ and basal subtype but less so in luminal A or luminal B subtype.

Taken together, the present study revealed that the association between CSTA expression and survival is dependent on molecular subtype of the tumor. Further, this study provides compelling evidences in favor of a functional link between CSTA and ERα and offers a rationale for investigating estrogen-mediated regulation of CSTA.

CHAPTER 5

Estrogen-mediated regulation of CSTA expression in breast cancer

5.1. Introduction

Estrogens are a group of steroid hormones that play a fundamental role in the development and maintenance of female reproductive system. Unfortunately, estrogen is also a major determinant in the etiology of breast tumors[21,281]. Proliferation and metastasis are two essential features of tumor initiation and progression. Previous studies on the role of estrogen in breast tumorigenesis have significantly advanced the understanding of its mitogenic effects[68,282]. However, beyond abnormal proliferation of breast cancer cells, much remains to be understood on the role of estrogen in tumor invasion and metastasis. Several studies have demonstrated the negative impact of estrogen on invasiveness and its reversal by tamoxifen[283-285]. Estrogen regulates tumor-stromal interaction, which is the molecular basis for ECM homeostasis[286]. Moreover, estrogen modulates the expression of syndecan, MMP2, MMP9, TIMP1 and TIMP2[287,288], indicating the perilous role of estrogen in invasion and metastatic progression of tumors.

Over the past five decades, breast cancer mortality has substantially declined though there is not much reduction in the incidence[289-291]. This notable achievement is due to increased early-stage detection and treatment, which requires better comprehension of the molecular basis of breast cancer initiation and progression. Since ECM remodeling is the crucial event in the tumor progression, understanding the mechanism of estrogenic regulation on ECM remodeling genes may help in delineating pathways underlying the E2-mediated regulation of invasion and metastasis.

An earlier report on the gene expression profile of E2-treated MCF-7 breast cancer cells[292] formed the basis of this study. A set of 189 E2 regulated ECM remodeling genes were identified, which includes CSTA, as an estrogen-repressed gene[293]. Since CSTA is an ECM remodeling gene with an inverse relationship with ERα in breast tumors, it is worth elucidating the mechanism of its regulation by E2 in breast cancer cells. This could provide an understanding of the probable mechanism by which ECM remodeling genes are regulated by E2, thereby affecting tumor progression

In this chapter, the mechanism of estrogen-mediated CSTA regulation in breast cancer cells is presented. Furthermore, the estrogen responsive region in the CSTA locus was identified using *in silico* and *in vitro* approaches.

5.2. Results

5.2.1. E2 suppresses CSTA expression in MCF-7 breast cancer cells

To evaluate the dose-dependent effect of E2 on CSTA expression, MCF-7 cells were treated with different concentrations of E2 (0.1 nM to 100 nM) for 72 h. CSTA expression was significantly downregulated in all the tested concentrations of E2 to the same extent (Figure 5.1A). The time-dependent effect of E2 stimulation on the expression of CSTA was analyzed by treating MCF-7 cells with 10 nM E2 for 24, 48 and 72 h. A significant reduction in CSTA mRNA was observed in all the tested time points with respect to vehicle treated control (Figure 5.1B). Alternatively, MCF-7 cells were treated with 10 nM E2 for 6 h to 96 h with individual time-matched vehicle-treated controls. A significant reduction in the CSTA expression was observed from 12 h onwards. The response was roughly similar in all the remaining time points (Figure 5.1C).

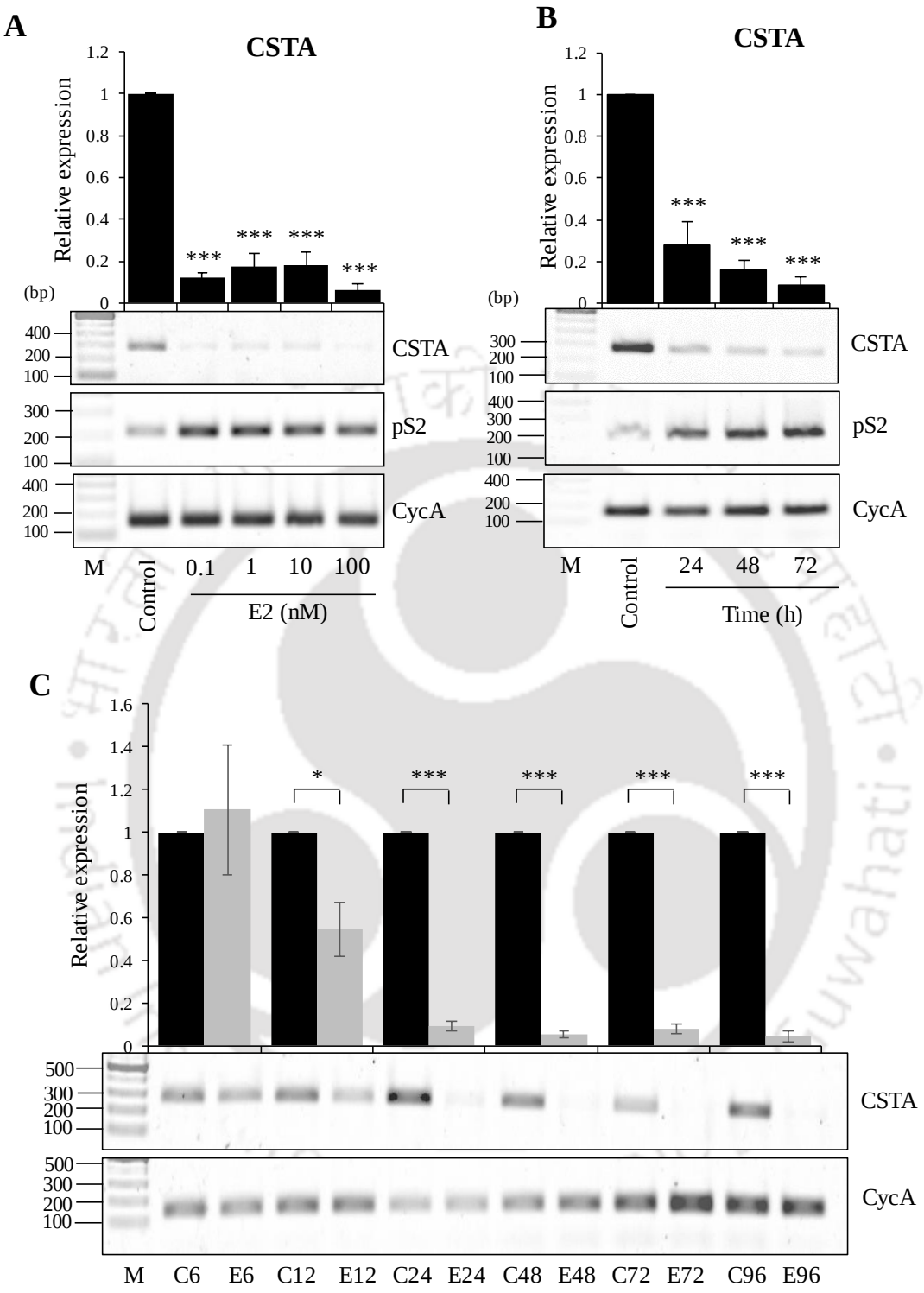

Figure 5.1. A time course and dose-response study of the regulation of CSTA by E2. A. MCF-7 cells were treated with the indicated concentration of E2 or vehicle (control) for 72 h. CSTA and pS2 mRNA levels were analyzed by semi-quantitative RT-PCR. pS2 was used as positive control for E2 treatment. B. MCF-7 cells were treated with 10 nM E2 or vehicle (control) for indicated periods. CSTA and pS2 mRNA levels were analyzed by semi-quantitative RT-PCR. pS2 was used as positive control for E2 treatment. ***$p < 0.001$, ANOVA followed by Tukey's HSD, n = 3. C. MCF-7 cells were treated with 10 nM E2 for indicated periods and each group contained time-matched control. CSTA mRNA level was assessed by semi-quantitative RT-PCR. C: Control (vehicle); E: E2; number indicates the duration of treatment in hours. Bars represent mean relative expression ± S.D. of CSTA mRNA with respect to vehicle-treated control. CycA served as an internal control. *$p < 0.05$, ***$p < 0.001$, Welch two-sample *t*-test, n = 3.

5.2.2. E2-mediated suppression of CSTA expression in MCF-7 cells involves ERα

Since CSTA expression is significantly suppressed by 12 h of E2 treatment, it was hypothesized that CSTA is directly regulated via ERα at the transcriptional level. To examine the involvement of ERα in the E2-mediated suppression of CSTA, MCF-7 cells were treated with a selective ERα-specific agonist, PPT for 72 h. PPT significantly suppressed CSTA expression to the same extent as E2 confirming that ERα is involved in the E2-mediated suppression of CSTA (Figure 5.2A, B). CSTA was not regulated by dexamethasone, progesterone, and testosterone propionate (Figure 5.2A).

Further, involvement of ERα in the E2-mediated suppression of CSTA was assessed using tamoxifen (SERM) and fulvestrant (SERD). MCF-7 cells were treated with 10 nM E2, 1 μM tamoxifen, or both for 24 h. In qRT-PCR analysis, no significant difference in CSTA expression was observed between E2-treated cells and, E2- and tamoxifen-treated cells. However, CSTA suppression by E2 in tamoxifen-treated and untreated cells is 42.56% and 73%, respectively, denoting the E2-mediated suppression of CSTA expression is partially rescued by tamoxifen (Figure 5.3A, bars 1, 2 and 4).

MCF-7 cells were treated with 100 nM fulvestrant for various duration to check the efficacy of proteasomal degradation of ERα. Western blotting analysis showed 80-84% reduction of ERα protein in fulvestrant-treated cells for 3 h to 24 h (Appendix II- Figure A2.1). When MCF-7 cells were pretreated with fulvestrant for 3 h prior to E2 treatment, no significant difference was observed in the CSTA expression in E2-treated groups with or without fulvestrant pretreatment (Figure 5.3B, grey-colored bars 2 and 4). CSTA suppression in E2-treated groups with or without 3 h fulvestrant pretreatment was 80% and 87.22% respectively (Figure 5.3B, grey-colored bars 2 and 4). Nevertheless, CSTA suppression in E2-treated groups with or without 24 h fulvestrant pretreatment was 42.59% and 72.65%, respectively (Figure 5.3B, black-colored bars 2 and 4). This denotes that fulvestrant pretreatment partially blocks the E2-mediated suppression of CSTA. Interestingly, tamoxifen or fulvestrant treatment alone induced CSTA expression (Figure 5.3A, B; bars 1 and 3).

MCF-7 cells were transfected with ERα-specific siRNA followed by E2 treatment. Western blotting analysis showed complete depletion of ERα protein in ERα siRNA transfected cells compared to scrambled siRNA transfected cells (Figure 5.4A). qRT-PCR analysis showed that the differences in fold change in CSTA and pS2 mRNA expression post E2 treatment between scrambled siRNA-treated and ERα siRNA-treated cells was statistically significant (Welch two-sample *t*-test). ERα siRNA significantly blocked E2-mediated

suppression of CSTA mRNA (Figure 5.4B). This clearly showed that the E2-mediated suppression of CSTA involves ERα in MCF-7 cells. Importantly, ERα siRNA alone induced the levels of CSTA mRNA (Figure 5.4B bars 1 and 3), which mirrors the inverse correlation between ERα and CSTA expression observed by analysis of TCGA-BRCA data (Figure 4.6A).

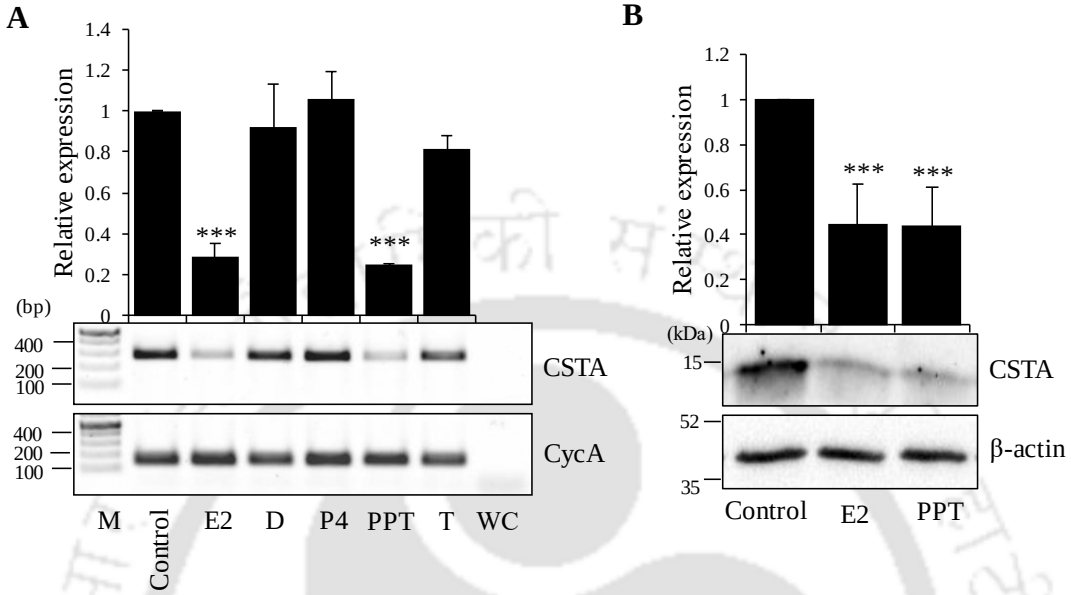

Figure 5.2. Effect of various hormones and ERα-specific agonist on expression of CSTA. A. MCF-7 cells were treated with various hormones: 10 nM E2, 10 nM dexamethasone (D), 10 nM progesterone (P4), 10 nM testosterone propionate (T), ERα-specific agonist: 10 nM PPT or vehicle (control) for 72 h and then analyzed by semi-quantitative RT-PCR. Bars represent mean relative expression ± S.D. of CSTA mRNA with respect to vehicle-treated control. CycA served as an internal control. WC: Water control. ***$p < 0.001$, ANOVA followed by Tukey's HSD, n = 3. B. MCF-7 cells were treated with 10 nM E2 and 10 nM PPT for 24 h. Total protein was extracted and subjected to western blotting analysis with custom generated polyclonal CSTA antibody (Appendix III). Bars represent mean relative expression ± S.D. of CSTA protein with β-actin as an internal control. ***$p <$ 0.001, ANOVA followed by Tukey's HSD, n = 7.

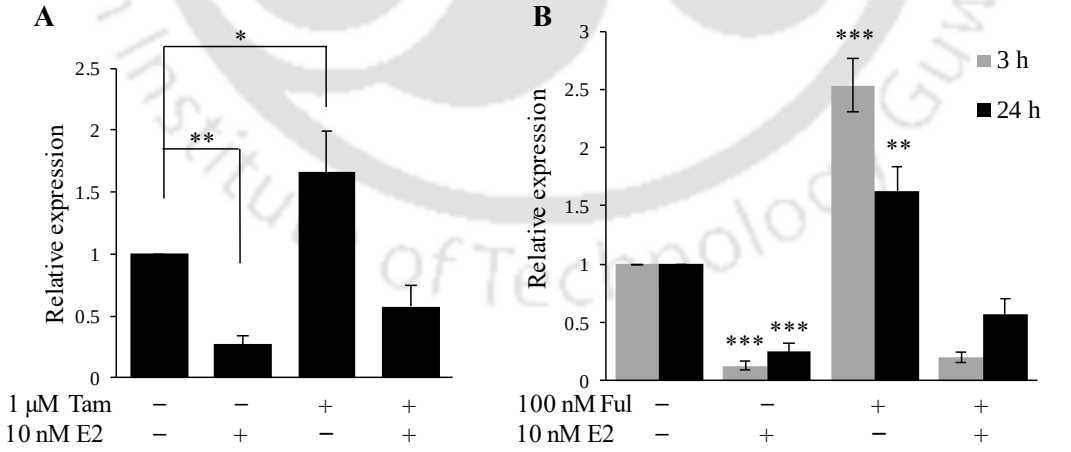

Figure 5.3. E2-mediated suppression of CSTA mRNA is blocked by tamoxifen, fulvestrant. A. MCF-7 cells were treated for 24 h with 10 nM E2, 1 μM tamoxifen (Tam) or both. Total RNA was isolated and CSTA expression was analyzed by qRT-PCR. Bars represent mean relative expression ± S.D. of CSTA mRNA with respect to vehicle-treated control. *$p < 0.05$, **$p < 0.01$, ANOVA followed by Tukey's HSD, n = 3. B. MCF-7 cells were pretreated with fulvestrant (Ful) or vehicle for 3 h (grey bars) or 24 h (black bars), followed by treatment with E2 or vehicle for 24 h. Total RNA was isolated and CSTA expression relative to vehicle-treated control was analyzed by qRT-PCR. ***$p < 0.001$, **$p < 0.01$, ANOVA followed by Tukey's HSD, n = 3. CycA served as an internal control

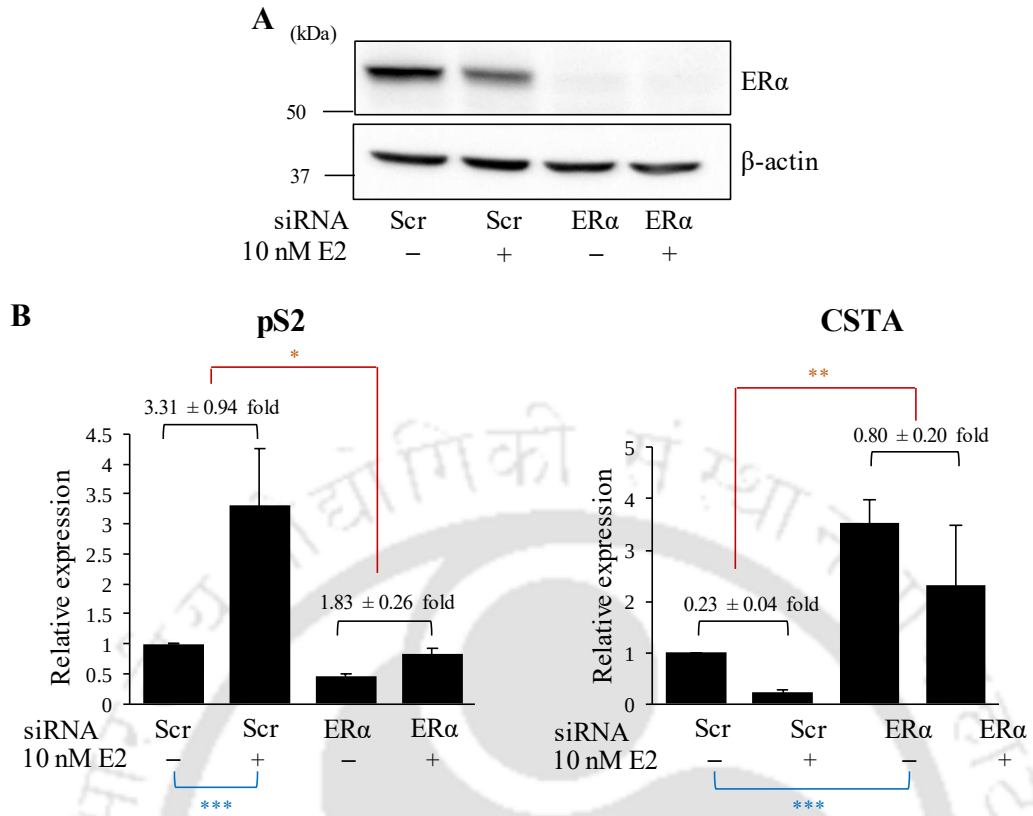

Figure 5.4. E2-mediated suppression of CSTA mRNA is blocked by ERα knockdown. MCF-7 cells were transfected with scrambled (Scr) or ERα siRNA for 24 h followed by recovery for a period of 24 h. The cells were then treated with ethanol or 10 nM E2 for 24 h. A. Western blotting analysis of ERα knockdown in MCF-7 cells. Total protein was isolated from phenol phase of the RNA extraction reagent prepared in house, after RNA isolation and subjected to western blotting analysis using ERα antibody. β-actin served as an internal control. B. Total RNA was isolated and expression levels of pS2 and CSTA mRNA relative to control (scrambled siRNA + vehicle-treated) were determined by qRT-PCR. ***$p < 0.001$, ANOVA followed by Tukey's HSD, n = 4. The differences in fold change in CSTA and pS2 mRNA expressions post E2 treatment between scrambled siRNA-treated and ERα siRNA-treated cells is statistically significant. **$p < 0.01$, *$p < 0.05$, Welch two-sample *t*-test, n = 4. CycA served as an internal control.

5.2.3. Estrogen enhances ERα occupancy in the intron-2 region of CSTA in MCF-7 cells

In silico analysis of the CSTA locus using JASPAR[294] revealed the presence of ERE in the intron-2 region (Figure 5.5A). ChIP-Seq data of chromatin samples from vehicle-, E2- or tamoxifen-treated MCF-7 cells, which were immunoprecipitated with ERα-specific antibody, were retrieved from SRA and analyzed in Galaxy platform. The results were viewed in the UCSC genome browser. A robust peak of ERα occupancy was observed in the intron-2 region of CSTA in E2-treated MCF-7 cells (indicated by the red rectangle, Figure 5.5B). This peak was diminished or negligible in tamoxifen or vehicle-treated MCF-7 cells. Notably, the peak of ERα binding overlapped with ERE predicted by JASPAR.

ChIP experiments on ethanol (vehicle)- or E2-treated MCF-7 cells using an ERα-specific antibody was performed to validate the *in silico* observations. Enrichment of the ERE

containing sequence in the pS2 locus (positive control) following E2 treatment validated the ChIP protocol (Figure 5.6). As shown in Figure 5.5B, the intron-2 region of CSTA was enriched in immunoprecipitated chromatin samples from E2-treated MCF-7 cells. These observations confirm that E2-mediated regulation of CSTA occurs via binding of ERα to the intron-2 region of CSTA.

A

Model ID	Model name	Score	Relative score	Start	End	Strand	predicted site sequence
MA0112.1	ESR1	11.844	0.808	233	250	-1	CCAGGACACTCTGACACC

B

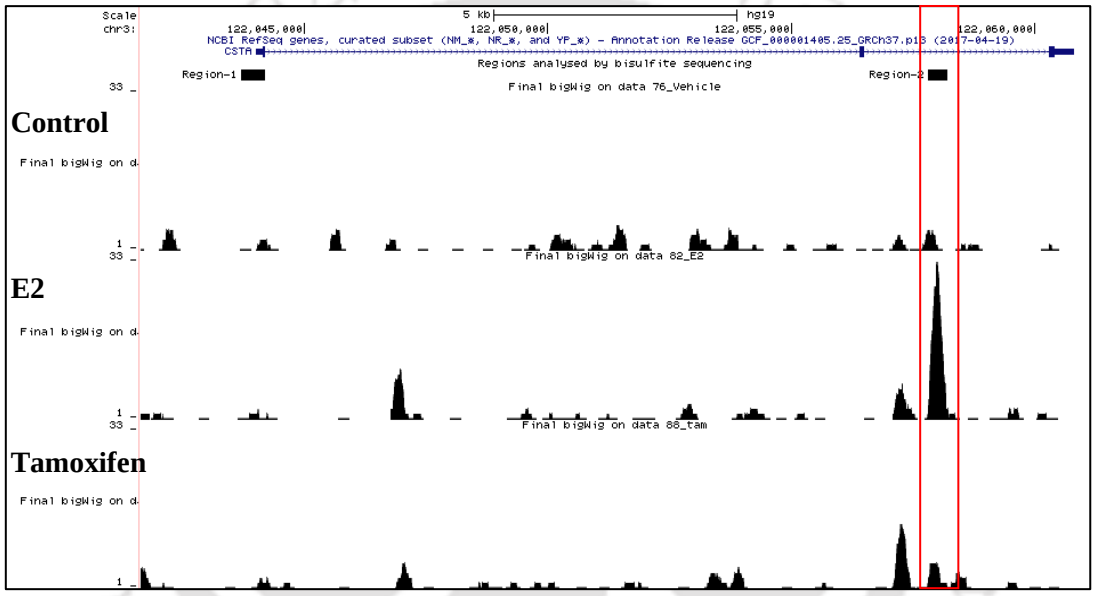

Figure 5.5. Possible involvement of intron-2 in the E2-mediated regulation of CSTA expression. A. JASPAR analysis of intron-2 of CSTA. B. Analysis of ChIP-seq data in Galaxy platform to study the ERα occupancy in the CSTA locus in MCF-7 cells treated with vehicle, estrogen (E2) and tamoxifen. E2 treatment increases the ERα occupancy in the intron-2 region of CSTA (red rectangle)

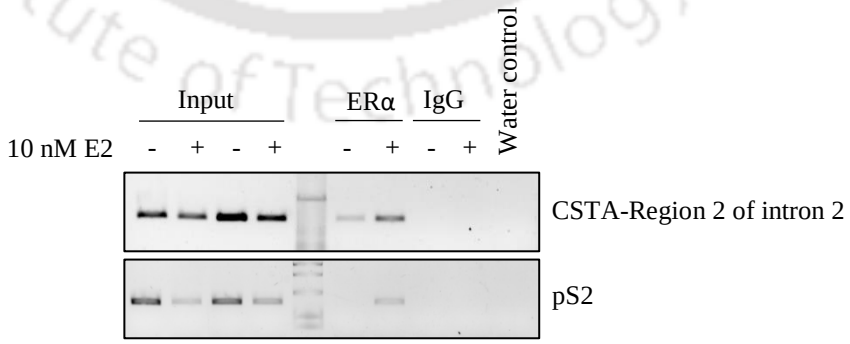

Figure 5.6. E2 enhances ERα occupancy in the intron-2 region of CSTA in MCF-7 cells. MCF-7 cells were treated with E2 for 24 h. Cross-linked chromatin samples from the treated and control cells were fragmented and immunoprecipitated with polyclonal ERα- or IgG-specific antibodies. Immunoprecipitated DNA was reverse cross-linked, purified and subjected to PCR analysis using primers flanking the intron-2 ERE (Region 2). pS2, a known E2 induced gene, served as a positive control. Data shown are representative of three independent experiments.

5.2.4. Regulation of CSTA expression by estrogen in other breast cancer cell lines

To understand the universality of the phenomenon of estrogen-mediated suppression of CSTA in breast cancer cells, two ERα-positive cell lines, T47D and ZR-75-1, were treated with E2 or PPT. Estrogen-mediated suppression of CSTA mRNA was observed in ZR-75-1 at 48 h post-stimulation (Figure 5.7A). In T47D, estrogen-mediated suppression of CSTA was not observed (Figure 5.7B).

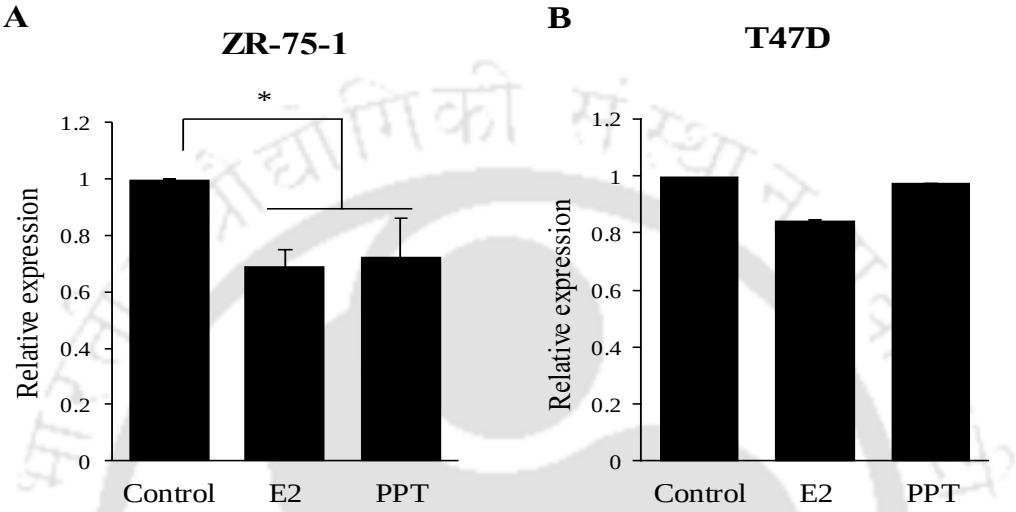

Figure 5.7. Differential regulation of CSTA by E2 and PPT in other ERα-positive breast cancer cells. ZR-75-1 (A) and T47D (B) cells were treated with 10 nM E2 or 10 nM PPT for 48 h. Total RNA was extracted and CSTA expression was analyzed by qRT-PCR. Bars represent mean relative expression ± S.D. of CSTA mRNA with respect to vehicle-treated control. CycA served as an internal control. *$p < 0.05$, ANOVA followed by Tukey's HSD, n = 3.

5.3. Discussion

In this study, the mechanism of E2-mediated suppression of CSTA in breast cancer cells was investigated. The gene expression profile of E2-treated MCF-7 cells was analyzed previously, in which, CSTA appeared as an estrogen repressed transcript[292]. ERα-positive MCF-7 cells were used as a model system representing luminal A subtype of breast tumor. *In vitro* experiments to explore the CSTA regulation revealed the time-dependent and dose-independent manner of suppression of CSTA by E2. Involvement of ERα in the suppression of CSTA by E2 was investigated by three strategies: i) activating ERα with a specific agonist, ii) blocking ERα action with SERM, SERD, and iii) ERα knockdown. PPT, an ERα agonist suppressed CSTA to a same extent as E2. On the other hand, tamoxifen, fulvestrant partially rescued E2-mediated suppression of CSTA and ERα-specific siRNA significantly blocked E2-mediated suppression of CSTA. Importantly, blocking ERα itself induced the CSTA

mRNA expression reinforcing the inverse relationship of CSTA with ERα, as observed in the analysis of clinical samples of TCGA data.

Further, to investigate the region involved in the E2-mediated suppression of CSTA, both *in silico* and *in vitro* approaches were used. JASPAR identified one potential ERE in the intron-2 region of CSTA. Moreover, analysis of publically available ChIP-Seq data revealed the enrichment of ERα at intron-2 upon E2 stimulation. This enrichment was not observed in tamoxifen-treated cells. In addition to that, *in vitro* ChIP experiments also further confirmed the involvement of intron-2 in the E2-mediated regulation of CSTA. Altogether these data suggest that ERα on stimulation with E2 regulates CSTA expression by binding to intron-2 region of CSTA.

Suppression of CSTA expression by estrogen in MCF-7 breast cancer cells was known[27,295]; however, this work, for the first time, demonstrated the role of ERα. The precipitous fall in CSTA mRNA within 24 h of E2 stimulation is noteworthy. Estrogen induces a mitogenic response in MCF-7 cells. Thus, estrogen-mediated CSTA suppression in MCF-7 cells is consistent with the proposed tumor suppressor role for CSTA[34]. Whether suppression of CSTA is essential for estrogen-mediated initiation of breast tumorigenesis is a question worth addressing in future studies.

Metastasis directly impacts survival. CSTA, a cathepsin inhibitor, probably plays a crucial role in metastasis. However, the possibility of considering CSTA alone as an independent predictor of metastatic progression needs to be explored. Furthermore, the metastatic progression of breast tumors overlaps temporally with the acquisition of hormone independence, which is often associated with loss of ERα expression. Given the inverse correlation between ERα and CSTA, is increased CSTA expression in some subtypes of primary breast tumors, an indication of the impending endocrine resistance and metastasis? In this context, the induction of CSTA mRNA expression by fulvestrant and tamoxifen in MCF-7 cells is of utmost relevance. Tamoxifen and fulvestrant are used for the treatment of ERα-positive and estrogen-responsive breast tumors[47]. Therefore, given the reported association of CSTA expression with the aggressive phenotype[31], it is worth investigating whether the endocrine treatment actually promotes aggressive behavior of breast tumors and metastasis via induction of CSTA expression.

Further, to understand the universality of estrogen-mediated suppression of CSTA, similar experiments were performed in two ERα-positive breast cancer cell lines, ZR-75-1 and

T47D. In ZR-75-1, estrogen-mediated suppression of CSTA mRNA was observed at 48 h post-stimulation. However, the extent of suppression was much less than that observed in MCF-7 cells. This might be due to the differential expression of ERα in both the cells. ZR-75-1 express relatively less ERα compared to MCF-7. Earlier Kolar and co-workers[295] have reported estrogen regulation of CSTA in ZR-75-1 cells. Surprisingly, in T47D, despite having high levels of ERα, estrogen-mediated suppression of CSTA was not observed. The probable reason behind the absence of E2-mediated regulation of CSTA expression in T47D cells is addressed in the next chapter.

CHAPTER 6

DNA methylation-dependent expression and regulation of CSTA

6.1. Introduction

Literature presents contradictory views on the relationship between CSTA expression, and breast cancer progression and prognosis[30,31,34,226]. However, most of the reports state that CSTA expression is lost or reduced in malignant tumors compared to benign tumors or normal tissues[42,211,225,296]. Forced expression of CSTA reduces migration and invasion in cell lines, and metastasis in xenograft models[34,42,43], suggesting a possible tumor suppressor role for CSTA. However, the mechanisms that lead to the loss of CSTA expression in breast tumors are not known.

Epigenetic silencing of tumor suppressor genes is a common phenomenon associated with tumor initiation and progression[36,297]. Methylation of tumor suppressor genes is an early event in the development and progression of tumors. The methylation status of tumor suppressor genes is a promising marker for the early detection and prognosis of cancer[298,299]. Besides the well-known tumor suppressor genes, members of the cystatin superfamily are also reported to be epigenetically silenced by DNA methylation in breast tumors and malignant

glioma[245,246]. The probable effect of DNA methylation in the silencing of CSTA in breast cancer cells is not yet studied.

In the previous chapter, estrogen-mediated suppression of CSTA expression in breast cancer cells was demonstrated. Aberration in estrogen signaling is the primary cause of breast cancer development[300]. Estrogen target genes are tightly regulated by interaction between ER and various co-activators/co-repressors, which also includes DNMTs[301]. Putnik and co-workers have reported the set of genes co-regulated both by DNA methylation and estrogen regulation in MCF-7 breast cancer cells[302]. DNA hypermethylation of estrogen-responsive enhancers is associated with reduced ERα binding in tamoxifen-resistant MCF-7 breast cancer cells[38]. Therefore, considering the tumor suppressor role of CSTA, it is worth addressing the plausible connection between DNA methylation and estrogen signaling in the regulation of CSTA.

In this chapter, the possibility of DNA methylation-dependent silencing of CSTA in breast cancer cell lines, and breast tumors of the TCGA cohort was explored. CSTA expression showed an inverse relationship with DNA methylation. Interestingly, estrogen regulation via ERα and DNA methylation-dependent silencing converges in the intron-2 of CSTA.

6.2. Results

6.2.1. Differential expression of CSTA in breast cancer cell lines

CSTA mRNA and protein expression in a panel of breast cancer cell lines were analyzed using RT-PCR and western blotting techniques, respectively. CSTA mRNA expression was highest in ZR-75-1, followed by MCF-7 cells. T47D cells expressed low but detectable levels of CSTA mRNA. CSTA mRNA expression in MDA-MB-231 and MDA-MB-453 was undetectable (Figure 6.1A). CSTA protein expression matched the mRNA expression in these cell lines (Figure 6.1B). T47D cells expressed the highest levels of ERα, followed by MCF7. ZR-75-1 cells expressed low but detectable levels of ERα. ERα expression was undetectable in the remaining cell lines (Figure 6.1A, B).

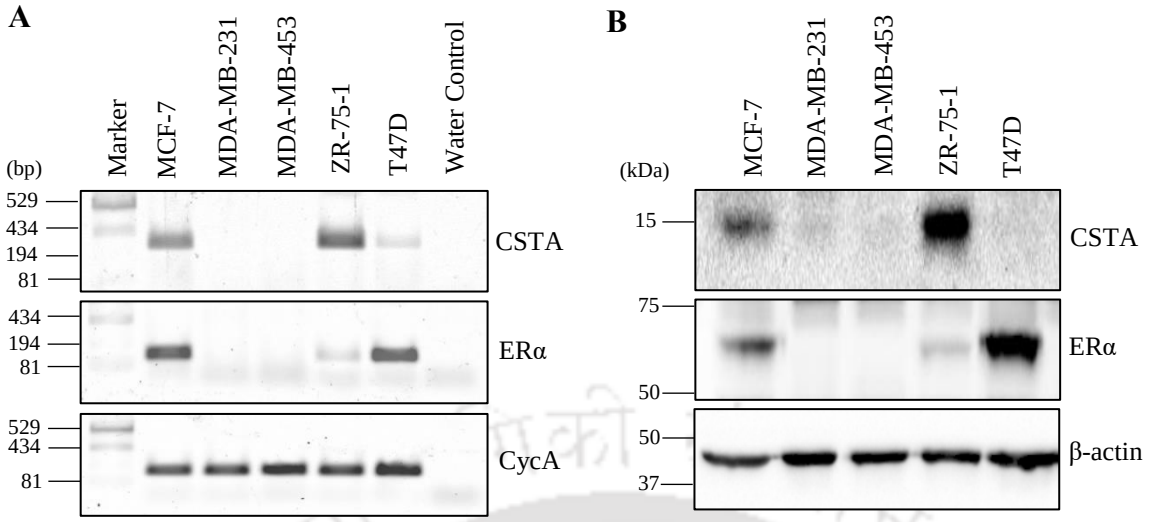

Figure 6.1. Expression of CSTA in breast cancer cell lines. A. Total RNA was isolated from the indicated breast cancer cell lines and subjected to RT-PCR analysis using primers specific for CSTA, ERα, and CycA. CycA served as an internal control. B. Total protein was isolated from the indicated breast cancer cell lines and subjected to western blot analysis using primary antibodies specific to CSTA, ERα, and β-actin. β-actin served as an internal control.

6.2.2. 5-aza induces CSTA expression in MDA-MB-231 cells

Treatment with 5-aza, a DNMT1 inhibitor, causes genome-wide demethylation. The effect of 5-aza on CSTA expression in MDA-MB-231 cells was tested. As shown in Figure 6.2, 5-aza treatment-induced CSTA mRNA expression. This suggested that low or absence of CSTA expression in breast cancer cells could be a result of DNA methylation in the CSTA locus.

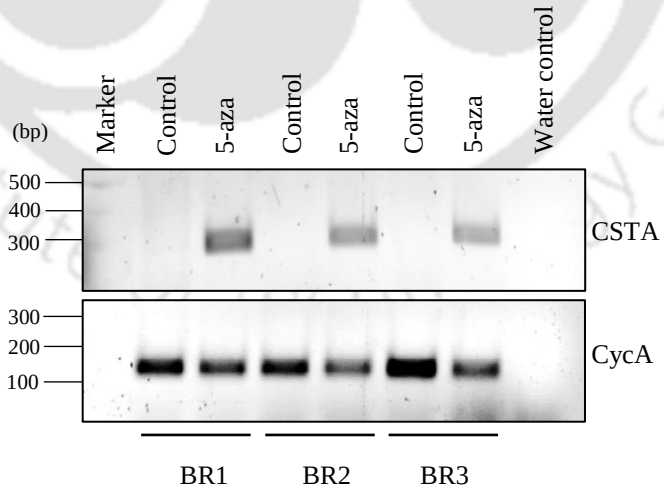

Figure 6.2. 5-aza induces the expression of CSTA mRNA in MDA-MB-231 cells. MDA-MB-231 cells were treated with DMSO (control) or 10 μM 5-aza for 5 days in M1 medium. Total RNA was isolated and reverse transcribed. The expression of CSTA was analyzed by routine RT-PCR. BR1, BR2, and BR3 are three biological replicates. Each biological replicate comprised of one dish each for DMSO and 5-aza treated cells.

6.2.3. *In silico* analysis of DNA methylation in the CSTA locus

CSTA does not have CpG islands[14]. Methylation data obtained from 450K BeadChip arrays (ENCODE project) was analyzed to ascertain the status of methylation in five CpGs in the CSTA locus. The location of these CpGs is indicated by the colored vertical lines in Figure 6.3A. One of these mapped to the intron-2. Among the remaining (referred to as the upstream CpGs), two mapped to the exon-1, and two were present in the upstream region close to the transcription start site. In MCF-7 cells, the four upstream CpGs appeared methylated. In T47D cells, out of the four upstream CpGs, one was methylated, one was unmethylated and the remaining two were partially methylated. The intron-2 CpG appeared unmethylated in MCF-7, but partially methylated in T47D cells (Figure 6.3A). The *in silico* analysis indicated that methylation in the intron-2 CpG had an inverse relationship with CSTA expression in the two cell lines.

6.2.4. Bisulfite sequencing of upstream and intron-2 regions in the CSTA locus

gDNA was isolated from the panel of breast cancer cells indicated in Figure 6.1. They were subjected to bisulfite sequencing analysis of Region 1 and Region 2 (Figure 6.3A), which harbor the CpG dinucleotides interrogated with the ENCODE data. Region 1 and Region 2 have a total of 7 and 3 CpG dinucleotides, respectively. Eighty-four CpGs were interrogated across 12 independent TA clones in the Region 1. Thirty-nine CpGs were interrogated across 13 independent clones in Region 2. The lollipop models for methylated and unmethylated CpG sites in Regions -1 and -2 are shown in Figure 6.3B and C, respectively. The proportion of methylated CpGs in each region in the cell lines was determined. In Region 1, MDA-MB-453 cells showed the highest proportion of methylated CpGs, followed by MCF-7, MDA-MB-231, T47D and ZR-75-1, in the decreasing order. In Region 2, MDA-MB-453 cell showed the highest proportion of methylated CpGs, followed by T47D, MDA-MB-231, ZR-75-1 and MCF-7 cells, in the decreasing order. There were significant differences in the proportions of methylated CpGs in both the regions in the indicated pair of cell lines (Figure 6.3B and C). Generally, methylation in both regions appeared to be inversely related to CSTA expression in breast cancer cells.

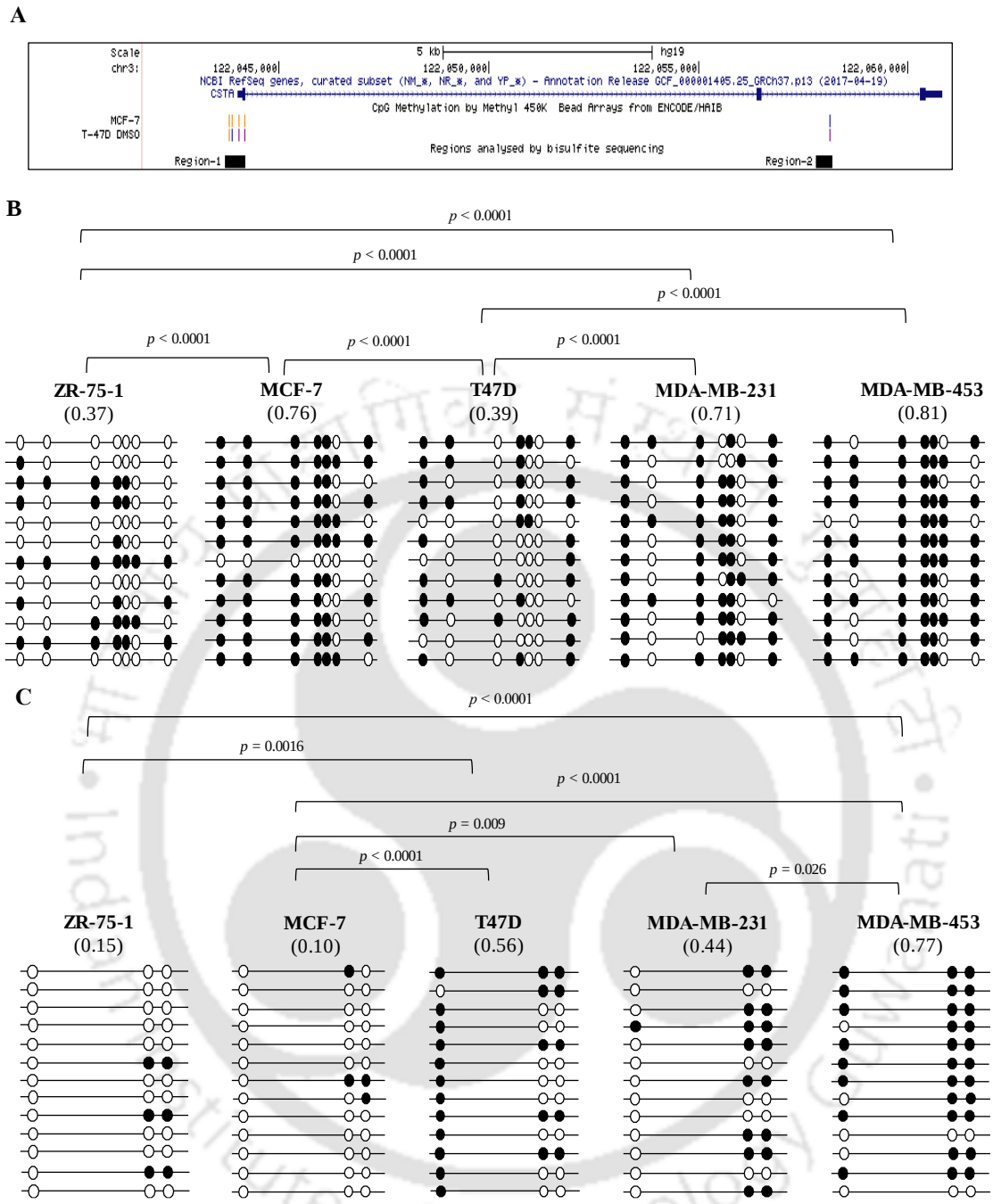

Figure 6.3. Differential methylation of upstream and intron-2 CpG sites of the CSTA locus in breast cancer cell lines. A. Snapshot from UCSC genome browser displaying the location of Methylation 450K BeadChip array probes with respect to the CSTA locus. The colored vertical lines along the MCF-7 and T47D tracks indicate the extent of methylation of the CpG sites; orange = methylated (beta value >= 0.6), purple = partially methylated (0.2 < beta value < 0.6), bright blue = unmethylated (0 < beta value <= 0.2). Black solid rectangles labeled as Region 1 and Region 2 indicate the regions analyzed by bisulfite sequencing. B, C. Bisulfite sequencing of Region 1 and Region 2, respectively. gDNA samples isolated from the indicated breast cancer cell lines were bisulfite converted and used for PCR reactions with primers specific to Region 1 and Region 2. The PCR amplified products were cloned in TA vector and sequenced. The inserts from 12 or 13 independent TA clones per cell line were analyzed for methylated and unmethylated CpG sites in Region 1 and Region 2, respectively. The methylation pattern is represented by lollipop plots. Filled circles represent methylated CpGs, and open circles represent unmethylated CpGs. The numbers below each cell line indicate the proportion of methylated CpGs. Two proportions from each pair of cell lines were tested for a significant difference. The indicated *p*-values are adjusted *p*-values obtained following Bonferroni correction, which are indicated only for the pair of cell lines with a significant difference in the proportion of methylated CpGs.

6.2.5. Upstream CpGs methylation inversely correlates with CSTA expression in breast tumors

CSTA expression (RNA-Seq; $\log_2$(RPKM+1)) and methylation data (generated with Illumina Infinium® Human Methylation 450K BeadChip array) for the primary breast tumors in the TCGA breast cancer dataset were accessed using the UCSC Xena browser[266]. Methylation data were available for 4 probes in the CSTA locus. These probes correspond to the four CpG sites in Region 1 shown in Figure 6.3A. Firstly, probe-wise analysis of the correlation between methylation (beta value) and CSTA expression in primary tumors was performed. Methylation at each of the 4 CpG sites inversely correlated with CSTA expression (Table 6.1). The primary tumors were divided into two groups, namely hypo- and hyper-methylated, using a beta value of 0.3 as a cut-off. CSTA expression in hypo-methylated tumors was significantly higher than those in hyper-methylated tumors (Figure 6.4A-D). A composite methylation score for each sample was generated by averaging the beta values for all the probes. The composite methylation score correlated inversely with CSTA expression ($\rho = -0.582$, $p < 0.0001$, Figure 6.4E). Furthermore, when the tumors were divided into hypo- and hyper-methylated groups, based on a cut-off composite methylation score of 0.3, the hypo-methylated tumors showed significantly higher expression of CSTA compared to hyper-methylated tumors (Figure 6.4F).

Table 6.1. Correlation between CSTA expression and methylation in CSTA locus in breast tumors of the TCGA cohort.

Probe ID	Spearman's ρ	*p*-value
cg14664412	-0.5482637	< 0.0001
cg18618429	-0.5683092	< 0.0001
cg21932814	-0.4717445	< 0.0001
cg26928972	-0.5154166	< 0.0001

6.2.6. Intron-2 region of CSTA encompasses a potential ERE

In chapter 5, estrogen-mediated suppression of CSTA, which occurs via binding of ERα to intron-2 region of CSTA, was demonstrated. Analysis of the CSTA locus using JASPAR[294] revealed a potential ERE in the intron-2 region. (Figure 5.5 and Figure 5.6). Interestingly, this ERE was located amidst the differentially methylated CpG sites analyzed in Region 2 (Figure 6.5).

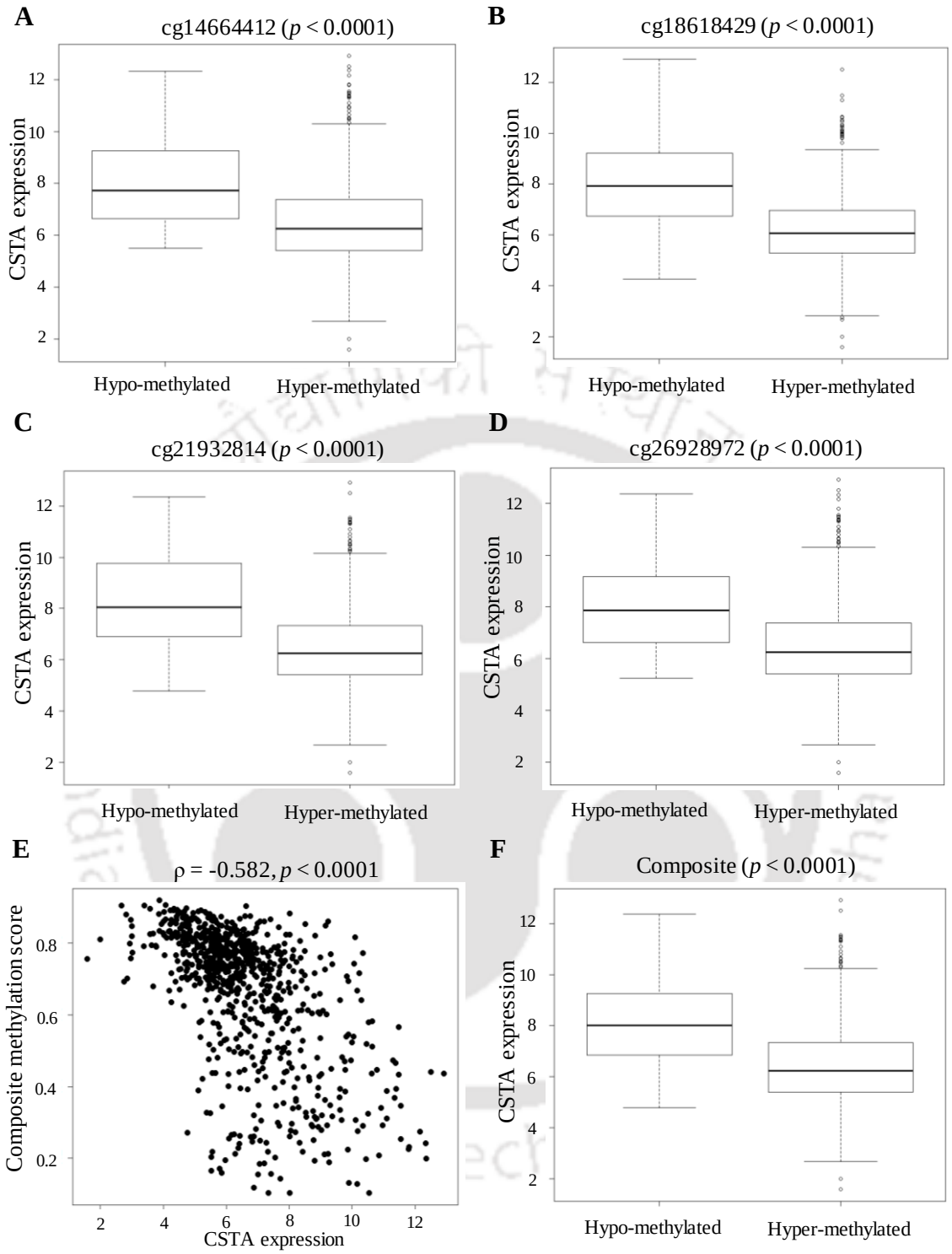

Figure 6.4. Inverse correlation between CpG methylation and CSTA expression in breast tumors of the TCGA cohort. A-D. Probe-wise analysis of the correlation between methylation and CSTA expression. The tumors were segregated into hypo-methylated or hyper-methylated groups based on the threshold beta value of 0.3 for each probe. The distribution of CSTA expression in hypo-methylated and hyper-methylated tumors are represented as box plots. E-F. A composite methylation score, which is the average beta value of all the probes, was determined for each tumor sample. The scatter plot of the composite methylation score versus CSTA expression is shown in E. The tumors were segregated into hypo-methylated or hyper-methylated groups based on the threshold composite score of 0.3. The distribution of CSTA expression ($\log_2$(RPKM+1)) in hypo-methylated and hyper-methylated tumors is represented as a box plot (F). The difference in mean CSTA expression in hypo-methylated and hyper-methylated tumors was analyzed by Welch two-sample *t*-tests.

TTCTGCTATCAAACTTTTCCTACTGGATCTCAGCCAC**CG**ATCCCAGTTCCCTTTTACTTC
CTGGTAGTCTGGCTGTTGATCCCTTTGCTCTGAGGCACTCTAGATTTAAGGTCTTGCCAG
TGATGTGACCTTCTCTATGTATTTCAAGTACCTATCAAGAGGTAGGTGGTAGAATGGAAG
GACCACAAGCTTAGGTGTCAGAGTGTCCTGGGTTTGAACCCTTGTTCAATTTGTTCTATG
GGAAGCTCCTCCTCCTCTCTGAGCCTTCATTCCCTTATCTGCACAATGAGGGTAATAATC
TACTT**CG**CAG**CG**TGTTGTGAGGAATAAATAAGCTGGAAATTTATTGAGCACTTATAATTC
ACTATGCACTATTCTAAGAACAGGGCTT

Figure 6.5. *In silico* analysis of intron-2 region of CSTA locus for potential ERα binding site. The location of ERE predicted by JASPAR in Region 2 is underlined. The CpG sites analyzed by bisulfite sequencing are indicated by bold letters.

6.2.7. Global demethylation restores estrogen regulation of CSTA in MDA-MB-231 and T47D cells

The effect of global demethylation on estrogen regulation of CSTA was studied in MDA-MB-231 and T47D cells. ERα and CSTA expression in 5-aza-untreated or -pretreated MDA-MB-231 and T47D cells, which were stimulated with vehicle or 10 nM E2 was examined. 5-aza caused a significant loss of methylation in Region 2 (p = 0.041 in MDA-MB-231, p = 0.034 in T47D) (Figure 6.6). As expected, in 5-aza-untreated MDA-MB-231 cells, ERα protein expression was not detectable after vehicle or E2 treatment (Figure 6.7A, lanes 1 and 2). There was no significant difference in CSTA mRNA expression (Figure 6.7C, bars 1 and 2, ANOVA followed by Tukey's HSD). On the other hand, in 5-aza-pretreated cells, an immunoreactive protein was detected on western blots with ERα-specific antibody (Figure 6.7A, lanes 3 and 4). This immunoreactive protein had a higher molecular mass compared to the expected 66 kDa for ERα. Notwithstanding this discrepancy, induction of PR, and further enhancement of its expression with E2 confirmed the generation of a functional ERα in 5-aza pretreated cells (Figure 6.7A, B, lanes 3 and 4). 5-aza significantly induced CSTA mRNA expression in MDA-MB-231 cells (Figure 6.7C, bars 1 and 3; ANOVA followed by Tukey's HSD). E2 suppressed the 5-aza induced levels of CSTA mRNA, although the difference was not statistically significant when analyzed by ANOVA. However, the levels of CSTA mRNA in 5-aza pretreated cells with and without E2 treatment were significantly different when analyzed by the Welch two-sample t-test in (Figure 6.7C, bars 3 and 4, p = 0.0098). Western blots failed to demonstrate CSTA protein in MDA-MB-231 cells. E2 treatment resulted in ERα occupancy in intron-2 in 5-aza pretreated cells (Figure 6.7D, lanes 8 and 9). These results show that demethylation of intron-2 CpGs leads to restoration of ERα and CSTA expression and estrogen suppression of CSTA in MDA-MB-231 cells.

T47D cells express a very low or undetectable level of CSTA. Without 5-aza pretreatment, T47D cells treated with E2 showed decreased levels of ERα protein and

increased levels of PR, as expected (Figure 6.8A, B, lanes 1 and 2)[303,304]. There was no observable effect on CSTA protein (Figure 6.8A, lanes 1 and 2). However, an increase in CSTA mRNA was observed, although the increase was not statistically significant (Figure 6.8C, bars 1 and 2; ANOVA followed by Tukey's HSD). 5-aza pretreatment alone caused a decrease in ERα protein in T47D cells (Figure 6.8A, lanes 1 and 3) in a manner similar to that reported in MCF-7 cells[305]. This also led to increased CSTA (Figure 6.8A, lanes 1 and 3) and decreased PR protein expression (Figure 6.8B, lanes 1 and 3). E2 induction of PR in 5-aza pretreated T47D cells showed that ERα was functional (Figure 6.8B, lanes 3 and 4). E2 not only enhanced CSTA protein expression (Figure 6.8A, lanes 3 and 4) but also significantly enhanced CSTA mRNA in 5-aza pretreated cells (Figure 6.8C, bars 1, 3 and 4; ANOVA followed by Tukey's HSD). E2 treatment resulted in ERα occupancy in the intron-2 region in 5-aza treated cells (Figure 6.8D, lanes 9 and 10). These results show that demethylation of intron-2 CpGs restores estrogen regulation of CSTA in T47D cells.

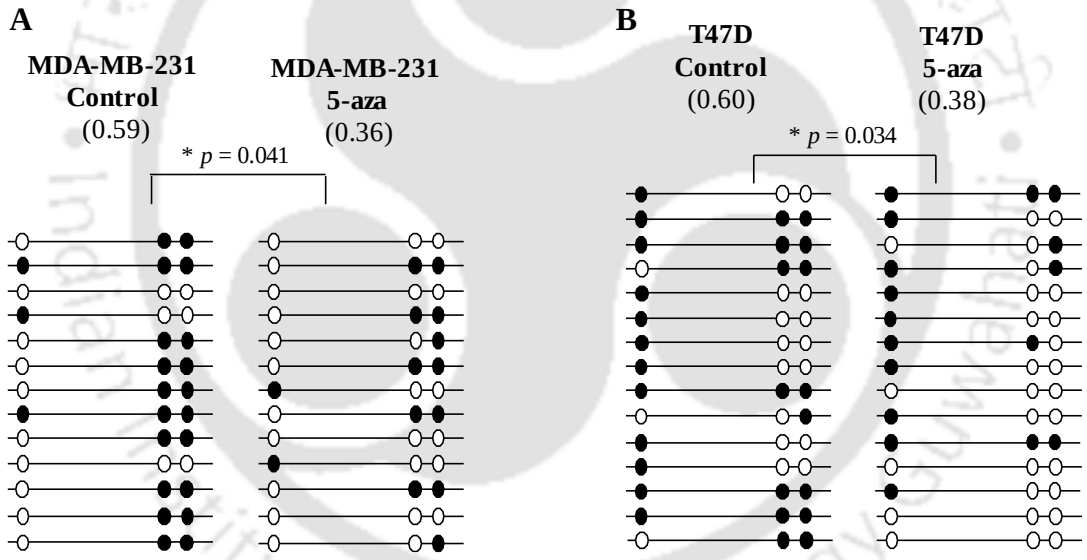

Figure 6.6. 5-aza treatment demethylates Region 2 in MDA-MB-231 and T47D cells. Cells were treated with 10 μM 5-aza for 5 days. gDNA isolated from treated cells were bisulfite converted and used for PCR reactions with Region 2-specific primers. The PCR amplified products were cloned in TA vector and sequenced. 13 and 15 independent TA clones were analyzed for methylated and unmethylated CpG sites in Region 2 of MDA-MB-231 and T47D cells, respectively. The methylation pattern is represented by lollipop plots. Filled circles represent methylated CpGs, and open circles represent unmethylated CpGs. The proportion of methylated CpGs are indicated in parentheses. The proportions were tested for significant difference as described in materials and methods. *p*-values obtained from Welch two-sample *t*-test are indicated above the plots.

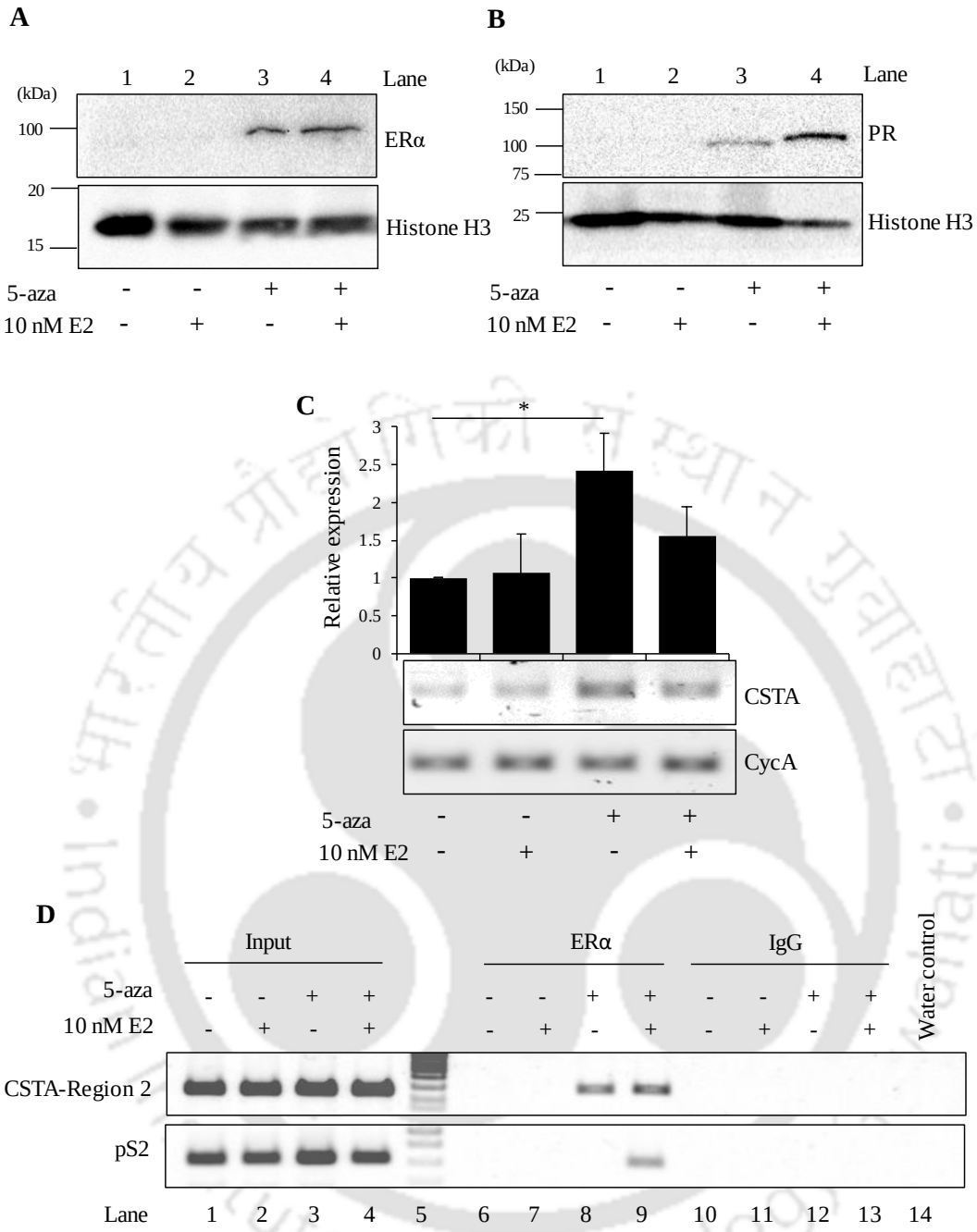

Figure 6.7. Global demethylation restores estrogen regulation of CSTA in MDA-MB-231 cells. A-C. Cells were subjected to global demethylation using 10 μM 5-aza for 5 days. The cells were then stimulated with 10 nM E2 or ethanol (vehicle) for 24 h. ER and PR expression was analyzed by western blotting (A, B) and CSTA expression was analyzed by semi-quantitative RT-PCR (C). CycA was used as an internal control in semi-quantitative RT-PCR. Histone H3 served as an internal control in western blots. CSTA mRNA expression in the control samples (without 5-aza and E2 treatments) were set to 1 and the expression in the other treatment groups was expressed relative to the control. Bars represent mean relative expression ± S.D. Data were analyzed by ANOVA followed by Tukey's HSD (n = 3). *$p < 0.05$. D. Cross-linked chromatin samples from the treated and control cells were fragmented and immunoprecipitated with monoclonal ERα- or IgG-specific antibodies. Immunoprecipitated DNA was reverse cross-linked, purified and subjected to PCR analysis using primers specific for Region 2 or pS2. Note the enrichment of the ERE containing sequence in the pS2 locus following E2 treatment (with 5-aza pretreatment), which validated the ChIP protocol.

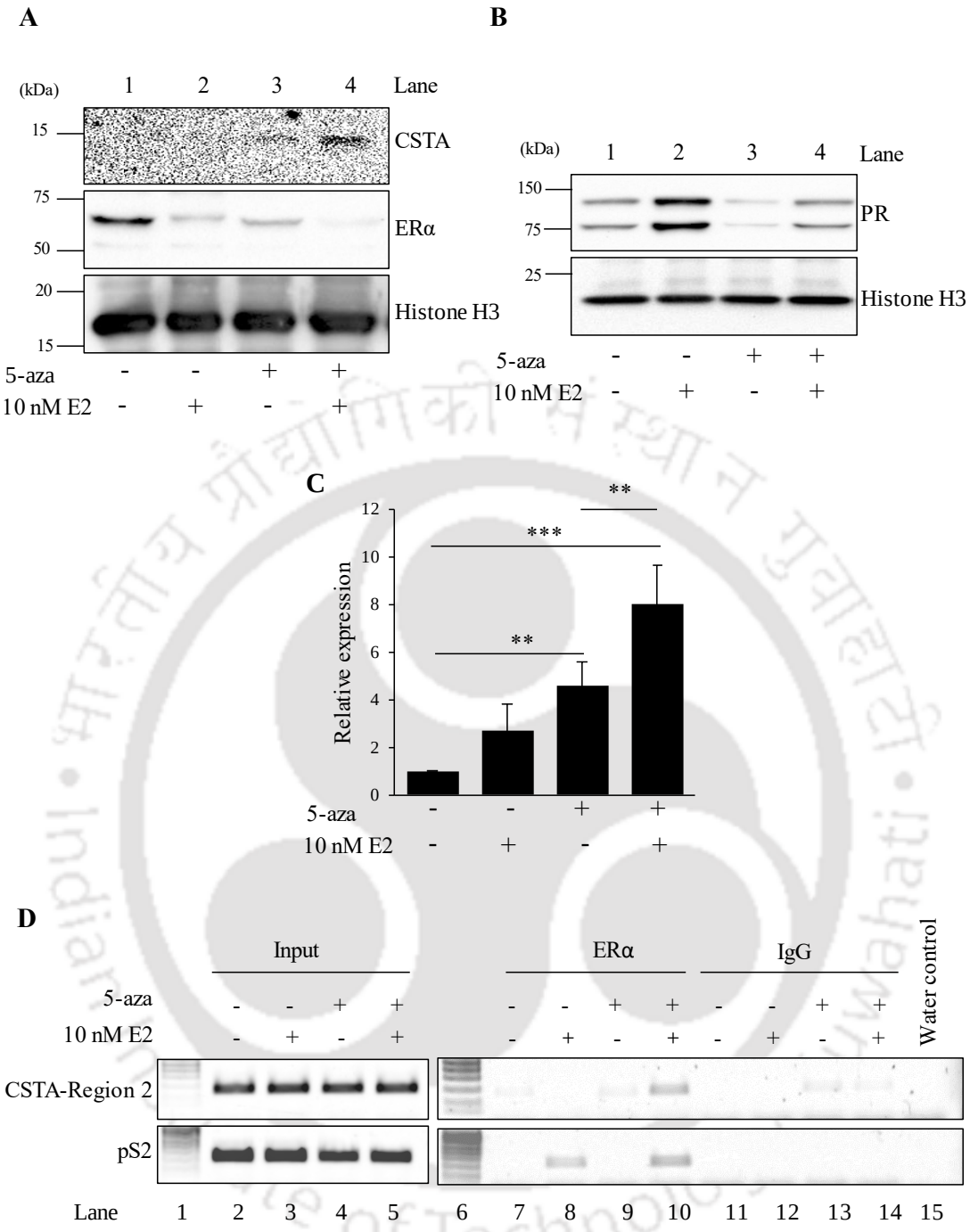

Figure 6.8. Global demethylation restores estrogen regulation of CSTA in T47D cells. A-C. Cells were subjected to global demethylation using 10 μM 5-aza for 5 days. The cells were then stimulated with 10 nM E2 or ethanol (vehicle) for 24 h. CSTA, ER and PR protein expression was analyzed by western blotting (A, B). CSTA mRNA expression was analyzed by qRT-PCR (C). CycA was used as an internal control in qRT-PCR. Histone H3 served as an internal control in western blots. Relative CSTA expression in the control samples (without 5-aza and E2 treatments) were set to 1 and the expression in the other treatment groups was expressed relative to the control. Bars represent mean relative expression ± S.D. Data were analyzed by ANOVA followed by Tukey's HSD (n = 4). ** $p < 0.01$, ***$p < 0.001$. D. Cross-linked chromatin samples from the treated and control cells were fragmented and immunoprecipitated with monoclonal ERα- or IgG-specific antibodies. Immunoprecipitated DNA was reverse cross-linked, purified and subjected to PCR analysis using primers specific for Region 2 or pS2. Note the enrichment of the ERE containing sequence in the pS2 locus following E2 treatment, which validated the ChIP protocol.

Similar experiments in MCF-7 cells showed that global demethylation neither affected CSTA expression nor affected E2-mediated suppression (Figure 6.9). Without 5-aza pretreatment, MCF-7 cells treated with E2 showed a significant reduction in CSTA mRNA (Figure 6.9A, bars 1 and 2) and significant induction in pS2 mRNA (Figure 6.9B, bars 1 and 2). In 5-aza pretreated cells, E2 significantly reduced CSTA mRNA (Figure 6.9A, bars 3 and 4) and induced pS2 mRNA (Figure 6.9B bars 3 and 4). No significant difference in CSTA (Figure 6.9A, bars 2 and 4) and pS2 expression (Figure 6.9B, bars 2 and 4) was observed in E2-treated cells with or without 5-aza pretreatment.

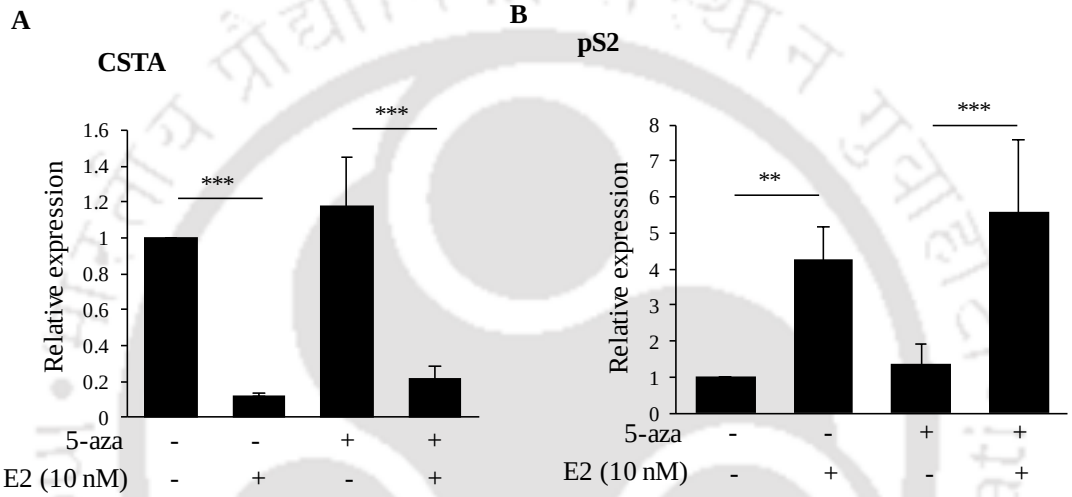

Figure 6.9. E2-mediated suppression of CSTA is unaffected by global demethylation in MCF-7 cells. A, B. MCF-7 cells were subjected to global demethylation using 10 μM 5-aza for 5 days. The cells were then stimulated with 10 nM E2 or ethanol (vehicle) for 24 h. CSTA and pS2 expression were analyzed by qRT-PCR. CycA was used as an internal control. pS2 was used as positive control for E2 treatment. aRelative CSTA and pS2 mRNA expression data are represented as bar graphs. The expression in the control samples (without 5-aza and E2 treatments) was set to 1 and the expression in the other treatment groups was expressed relative to the control. Bars represent mean relative expression ± S.D. Data were analyzed by ANOVA followed by Tukey's HSD (n = 4). ** $p < 0.01$, *** $p < 0.001$.

6.3. Discussion

There were two motivating reasons behind this study of the relationship between CSTA gene expression and methylation in breast cancer. Firstly, CSTA is a proposed tumor suppressor[272], and methylation-dependent silencing of tumor suppressors are well known[306]. Recently, Ma and co-workers showed the association between the loss of CSTA expression in lung cancer cell lines and partial methylation of the CpG dinucleotides in the promoter and exon-1 regions[42]. Secondly, Cystatin M, which is also a reversible inhibitor of cathepsins B and L, is silenced due to methylation of the proximal promoter CpG island in breast cancer cell lines and primary invasive breast tumors[245,307-309]. The present analysis was focused on CpG methylation on two regions of the CSTA locus. This is the first study to demonstrate

the inverse relationship between CSTA expression and methylation in the context of breast cancer. The CSTA locus lacks CpG islands. However, the examples of DNA methylation-dependent regulation of CpG island-less genes[310-316] motivated the investigation of DNA methylation of CSTA locus.

The conclusions of this study are drawn from results obtained through three different approaches; a) analysis of CSTA expression and methylation data from the TCGA breast cancer cohort, b) examination of DNA methylation data (ENCODE project) for the CSTA locus, and c) bisulfite sequencing of DNA isolated from breast cancer cell lines, which express differential levels of CSTA. TCGA methylation data was generated using the Infinium Methylation 450K BeadChip arrays, which do not have probes to interrogate all CpG sites in the CSTA locus. Due to this limitation, no conclusion could be drawn regarding the correlation between CSTA expression and methylation of the intron-2 CpGs. Nevertheless, an inverse correlation between CSTA expression and methylation in the upstream CpG sites was observed (Figure 6.4E). The ENCODE project data corresponding to MCF-7 and T47D cells, and the bisulfite sequencing results clearly demonstrated that CSTA expression, and methylation in the intron-2 CpGs, are inversely correlated. Collectively, these data provide compelling evidences in favor of DNA methylation-dependent silencing of CSTA in breast cancer cells. Altered cathepsin B: CSTA ratio in breast tumors is reported. It also correlates with disease prognosis[30,32,226]. Ablation or inhibition of cathepsin B also inhibits spine and lung metastasis in the animal model[35]. Therefore, it was proposed that DNA methylation-mediated silencing of CSTA in primary breast tumors tips the cathepsin B/CSTA balance in favor of cathepsin B, which in turn facilitates tumor invasion and metastasis. A detailed study on the correlation of DNA methylation in the CSTA locus, and disease progression, treatment outcome and survival, may uncover its potential as a prognostic marker.

CSTA and ERα mRNA expression in breast tumors of the TCGA cohort are inversely correlated (Figure 4.6A). This is consistent with the observed induction of CSTA mRNA in MCF-7 cells following ERα knockdown (Figure 5.4B). The inverse relationship is reiterated in the ERα-positive breast cancer cell lines used in this study. ZR-75-1, which has the highest expression of CSTA, has the least ERα expression, whereas T47D, which has the least expression of CSTA, has the highest expression of ERα. CSTA and ERα expression levels in MCF-7 cells are in between these two extremes. Estrogen suppresses CSTA expression in MCF-7 cells via ERα (Figure 5.5 and Figure 5.6). However, the mechanistic role of ERα was not known. The prediction of an ERE by JASPAR, the peak of ERα binding revealed by ChIPseq data, and the validation of increased ERα occupancy in the intron-2 following

estrogen treatment of MCF-7 cells suggest that estrogen suppresses CSTA expression at the level of transcription. The precise events post-ERα binding that lead to transcriptional shut-off are worth addressing in future investigations. However, the mechanism of estrogen-mediated regulation of CSTA is more complex. Estrogen does not produce similar effects on CSTA expression in ERα-positive breast cancer cell lines. Estrogen suppresses CSTA expression in ZR-75-1 cells. However, the extent of suppression is much lower than that observed in MCF-7 cells. In T47D cells, estrogen does not modulate CSTA expression. Subsequent analysis revealed the probable reason behind the lack of estrogen-mediated regulation of CSTA in T47D cells.

Here two specific regions, namely Region 1 and Region 2, that encompass few of the upstream and intron-2 CpG sites, respectively, were analyzed. While it is worth analyzing methylation at every CpG site in the CSTA locus, this study shows that intron-2 is the site of convergence of estrogen regulation and DNA methylation-dependent silencing. Interestingly, the ERα binding site is located amidst the intron-2 CpGs (Region 2). The global demethylation experiments with MDA-MB-231 and T47D cells revealed the conflict between ERα binding and DNA methylation in Region 2. Due to methylation-dependent silencing, MDA-MB-231 cells do not express ERα[317,318] and CSTA (Figure 6.1A, B). Global demethylation in MDA-MB-231 cells established functional ERα (as revealed by PR induction upon estrogen stimulation), and CSTA mRNA expression (Figure 6.7A, B). Furthermore, estrogen tends to suppress 5-aza induced CSTA mRNA, resembling estrogen regulation of CSTA in MCF-7 cells. This was possible because demethylation of Region 2 CpGs made the intron-2 ERE accessible to ERα. The conflict is also supported by the results from T47D cells. It must be noted that T47D cells show a significantly greater level of Region 2 methylation than MCF-7 cells, which likely prevents ERα binding to the ERE. This is arguably the reason why despite detectable levels of CSTA and functional ERα in T47D, estrogen does not regulate CSTA. 5-aza not only increased CSTA expression in T47D but also made it amenable to estrogen regulation. Although it is not clear why the direction of CSTA regulation in T47D is opposite to that observed in MCF-7, MDA-MB-231 and ZR-75-1 cells, these are enticing evidence that indicates the crucial role of intron-2 in CSTA expression and regulation.

The relationship between DNA methylation and transcription is not a one-way interaction. In a given genomic locus, the transcriptional activity can limit DNA methylation. In Arabidopsis, this is evident from the distribution of methylated and transcriptionally active loci[319]. Pharmacological inhibition of RNA polymerase II induces repressive histone modification, which results in epigenetic silencing. It is possible that repressive histone

modification also leads to DNA methylation. Thurman and co-workers studied methylation of transcription factor binding sites and transcription factor abundance in DNAse I hypersensitive sites. They found an inverse correlation between the expression level of a given transcription factor and methylation of the cognate binding site. This suggests a model of "passive DNA methylation"[320]. On the other hand, methylation of cytosine residues in CpG dinucleotides prevents binding of transcription factors to their cognate response elements on DNA[321-324], thereby interfering with gene expression. An exception to this model is the binding of Sp1 to the methylated cognate site, leading to enhanced gene transcription[325]. Alternatively, methylated CpG sites attract MCBPs[326]. MCBPs, in turn, recruit histone deacetylases and methylases that cause remodeling and compaction of the local chromatin and transcriptional shutoff. This is the converse model of methylation-mediated blockade of transcription factor access. In the context of the interaction between transcriptional activity and transcription factor binding, ERα is not an exception. Ung and co-workers have analyzed DNA methylation in relation to ERα expression and binding[327]. They found an inverse correlation between ERα expression and CpG methylation within ERα binding sites. Methylation of these CpGs was therefore interpreted as being dependent on ERα activity, consistent with the passive model of DNA methylation; more the ERα binding, lesser the methylation. Except for T47D, the cell lines used in this study have patterns of ERα expression and Region 2 methylation that is consistent with the passive model. However, that methylation of Region 2 in MDA-MB-231 and MDA-MB-453 cells is entirely due to lack of ERα expression, cannot be stated with certainty. Parallelly, results from T47D cells are consistent with the converse model, wherein despite ERα expression and detectable levels of CSTA, estrogen could not significantly modulate CSTA expression. The combined results from MDA-MB-231 and T47D showed that demethylation could restore estrogen regulation at the CSTA locus. This is most likely due to the restoration of the ERα access to the ERE.

Taken together, the present study shows that CSTA expression in breast cancer cells is inversely related to DNA methylation in the CSTA locus. It explains the loss of CSTA expression in breast tumors. DNA methylation of CSTA may be exploited for predicting metastatic progression of breast tumors. Furthermore, this study has uncovered an interesting interplay between ERα binding and transcriptional regulation. The proposed model from this study is that CSTA expression in breast cancer cells is an integrated result of estrogen regulation and DNA methylation-dependent silencing converging on the intron-2.

CHAPTER 7

Functional role of CSTA in breast cancer

7.1. Introduction

Metastasis accounts for approximately 90% of cancer-related deaths[328]. Metastasis is the major hurdle in the treatment of breast cancer, necessitating therapeutic intervention to specifically target metastasis. A better understanding of molecular mechanisms of tumor progression may unravel the key molecules against which therapeutic strategies could be designed. The basic strategies of intervention are a) blocking the molecules which promote metastasis using specific antibodies, b) restoring the expression of suppressors, or c) mimicking the function of suppressors using small molecules[329]. Besides metallo and serine proteases, cysteine proteases such as cathepsins are also known to play a causal role in tumor progression and metastasis. They are involved in invasion either by directly cleaving the ECM components or by indirectly activating other proteases such as pro-uPA and MMPs[172]. Hence, cystatins are likely to regulate tumor invasion and metastasis by inhibiting cysteine cathepsins. Besides the cathepsin inhibitory property of CSTA, its involvement in apoptosis and differentiation makes it an interesting point of investigation for exploring novel avenues in cystatins-based treatment strategy[330].

During tumor progression, polarized epithelial cells undergo multiple biochemical changes resulting in a mesenchymal cell phenotype with enhanced migratory potential. This process is called EMT. Members of the cystatin superfamily were proposed to impede EMT

by inhibiting cathepsin-mediated proteolytic cleavage of ECM components. Cystatin C inhibits TGF-β induced EMT by antagonizing TGF-β type II receptor in breast cancer[331]. In colon cancer, cystatin D represses the EMT inducers and promotes adhesive phenotype[332]. Loss of function mutation in the CSTA gene resulted in impaired cell-cell adhesion[234]. On the contrary, cystatin SN promotes tumor migration and invasion by inducing EMT in breast cancer cells[333]. Nevertheless, the precise role of CSTA in the transition of epithelial to mesenchymal phenotype during tumor progression in the context of breast cancer is not yet established.

In the present study, attempts were made to understand the possible role of CSTA in breast cancer cells by stably expressing CSTA into MDA-MB-231 cells or transiently knocking down CSTA from MCF-7 cells. This study revealed that CSTA modulates migration and invasion of breast cancer cells without affecting proliferation.

7.2. Results

7.2.1. Cloning of CSTA ORF into mammalian expression vector

CSTA ORF was amplified from MCF-7 cDNA and cloned along with FLAG tag into mammalian expression vector, pBABE-puro, as described in section 3.10 of chapter 3. The positive clones were confirmed by PCR (Figure 7.1) and Sanger sequencing (Figure 7.2).

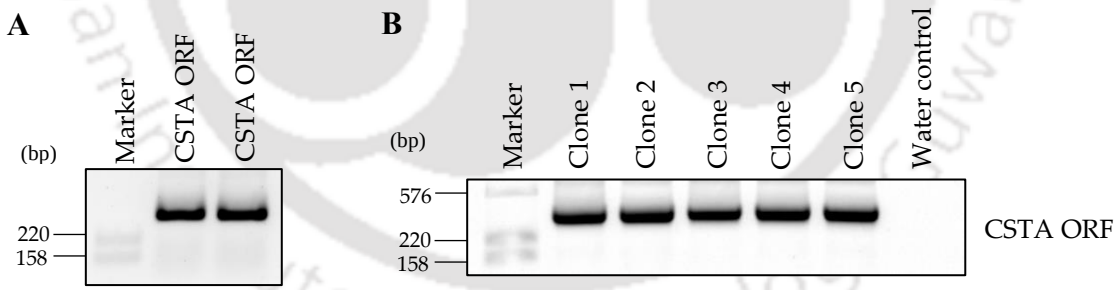

Figure 7.1. Cloning of CSTA ORF in a mammalian expression vector. A. PCR amplification of CSTA ORF from cDNA of MCF-7 breast cancer cells. B. Clone confirmation by PCR with CSTA-specific primer.

```
Clone          101 CCTTTATCCAGCCCTCACTCCTTCTCTAGGCGCCGGCCGGATCCGCCACC    150

Reference        1 ----------------------------------------------------      0

Clone          151 ATGATACCTGGAGGCTTATCTGAGGCCAAACCCGCCACTCCAGAAATCCA    200
                   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Reference        1 ATGATACCTGGAGGCTTATCTGAGGCCAAACCCGCCACTCCAGAAATCCA     50

Clone          201 GGAGATTGTTGATAAGGTTAAACCACAGCTTGAAGAAAAAACAAATGAGA    250
                   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Reference       51 GGAGATTGTTGATAAGGTTAAACCACAGCTTGAAGAAAAAACAAATGAGA    100

Clone          251 CTTACGGAAAATTGGAAGCTGTGCAGTATAAAACTCAAGTTGTTGCTGGA    300
                   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Reference      101 CTTACGGAAAATTGGAAGCTGTGCAGTATAAAACTCAAGTTGTTGCTGGA    150

Clone          301 ACAAATTACTACATTAAGGTACGAGCAGGTGATAATAAATATATGCACTT    350
                   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Reference      151 ACAAATTACTACATTAAGGTACGAGCAGGTGATAATAAATATATGCACTT    200

Clone          351 GAAAGTATTCAAAAGTCTTCCCGGACAAAATGAGGACTTGGTACTTACTG    400
                   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Reference      201 GAAAGTATTCAAAAGTCTTCCCGGACAAAATGAGGACTTGGTACTTACTG    250

Clone          401 GATACCAGGTTGACAAAAACAAGGATGACGAGCTGACGGGCTTTGACTAC    450
                   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Reference      251 GATACCAGGTTGACAAAAACAAGGATGACGAGCTGACGGGCTTTGACTAC    300

Clone          451 AAAGACGATGACGACAAGTAGGTCGACCCTGTGGAATGTGTGTCAGTTAG    500
                   ||||||||||||||||||||
Reference      301 AAAGACGATGACGACAAGTAG----------------------------    321

Clone          501 GGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATG    550

Reference      322 ----------------------------------------------------    321
```

Figure 7.2. Sanger sequencing of CSTA expression construct. The DNA sequence of the clone (Query) was compared by pairwise alignment with CSTA mRNA sequence (Reference). Blue colored sequences in the beginning and the end are the flanking vector sequences.

7.2.2. Generation of MDA-MB-231 cells stably expressing CSTA

MDA-MB-231, a highly invasive cell line, does not express CSTA, which makes it a suitable model system for assessing the function of CSTA. MDA-MB-231 cells stably expressing CSTA were generated as described in section 3.11 of chapter 3. Total RNA was extracted from the cells and gDNA-free RNA was subjected to qRT-PCR analysis. CSTA mRNA expression was analyzed in 7 clones by qRT-PCR (Figure 7.3A). CSTA protein level was analyzed in 3 clones (Figure 7.3B). MCF-7 was used as a positive control. Among all clones, KC7 expressed the highest level of CSTA mRNA and protein. Therefore, the KC7 clone was used in further experiments to evaluate the role of CSTA in proliferation, migration, and invasion.

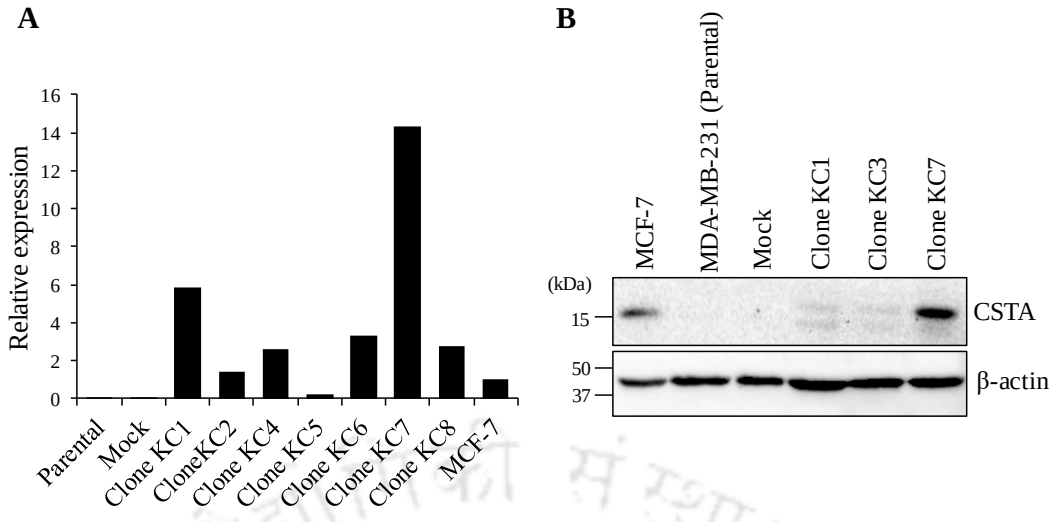

Figure 7.3. Screening of stable cell lines for CSTA expression. Total RNA and protein were extracted from various clones of MDA-MB-231 stably transfected with CSTA expression construct, mock and parental cells and subjected to qRT-PCR (A) and western blotting analysis (B). MCF-7 was used as a positive control. Bars represent the relative expression of CSTA with respect to MCF-7. CycA and β–actin served as an internal control for qRT-PCR and western blots, respectively.

7.2.3. CSTA expression reduces the migration of MDA-MB-231 cells

The effect of CSTA expression on migration of MDA-MB-231 cells was assessed by scratch wound assay. Scratches were made on parental, mock, and KC7 monolayers and imaged at 0 h and 24 h. There was 48% reduction in the wound area in KC7 compared to parental cells (Figure 7.4A, B). This indicates that CSTA expression reduces migration in MDA-MB-231 cells. Further, MTT assay was performed on parental, mock, and KC7 cells to investigate the effect of CSTA expression on proliferation. No significant difference in the viability of KC7, mock or parental cells was observed (Figure 7.4C). This indicates CSTA expression does not affect the proliferation in MDA-MB-231 cells.

7.2.4. CSTA expression reduces the invasion of MDA-MB-231 cells

CSTA is the endogenous inhibitor of cathepsin B, an ECM remodeling enzyme involved in the tumor invasion and metastasis. Therefore, the invasive potential of CSTA expressing stable transfectants was assessed using serum as a chemoattractant by transwell invasion assay. Crystal violet staining of the membrane showed that the number of invaded cells in KC7 was relatively less compared to parental and mock. No invasion was observed in the cells seeded in the serum-free condition on both sides of the membrane. To obtain quantitative data, crystal violet stain was dissolved in SDS and absorbance was measured. The mean absorbance of KC7 samples was significantly less compared to mock or parental clone samples ($p < 0.05$). This denotes the potential of CSTA in reducing the invasion of MDA-MB-231 cells (Figure 7.5A, B).

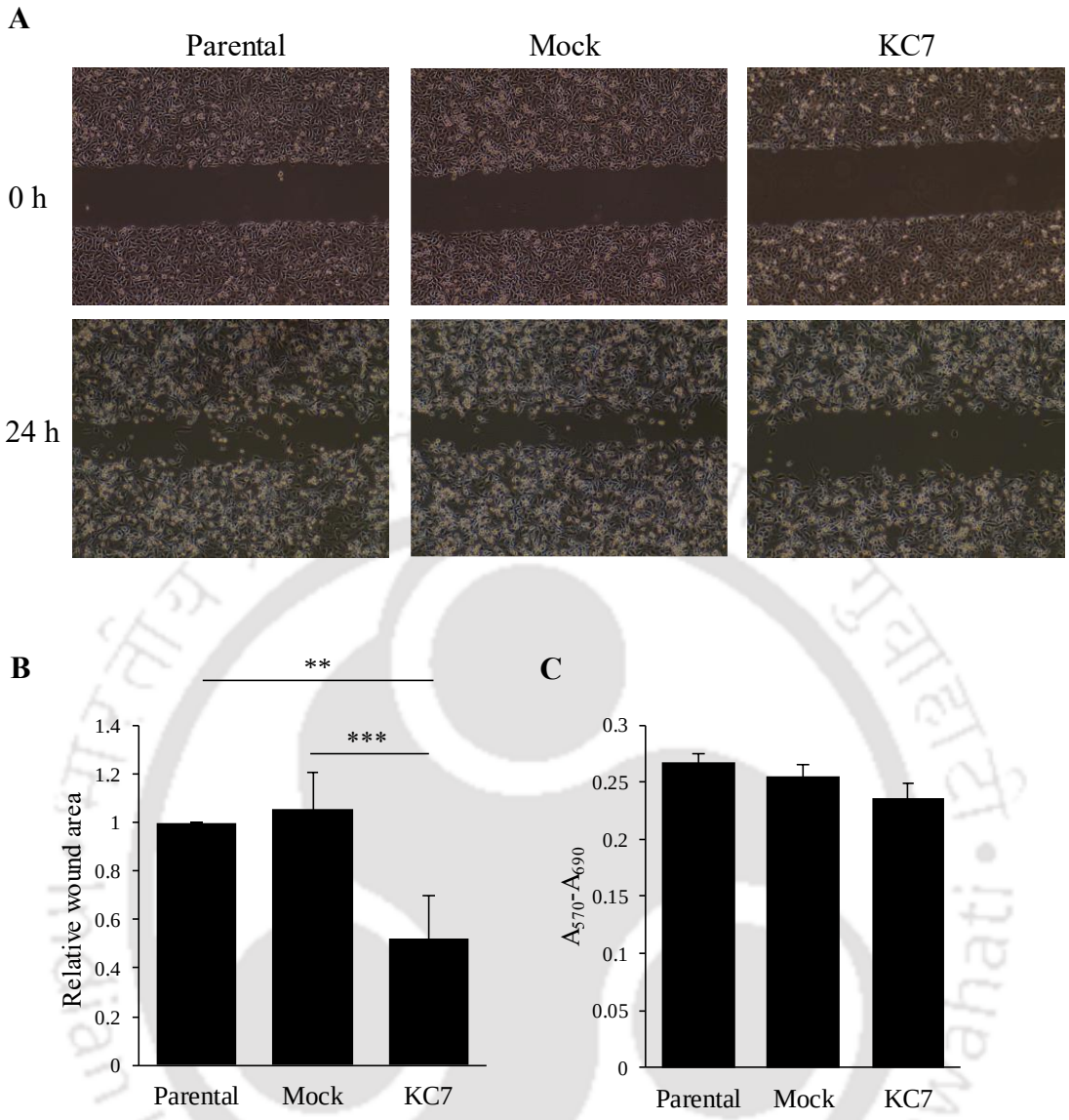

Figure 7.4. CSTA overexpression reduced the migration of MDA-MB-231 cells without affecting proliferation. A, B. Cells were seeded and grown up to confluence. The monolayer was serum-starved, treated with mitomycin C and scratches were made. Representative images of scratches imaged at 0 h and 24 h (A). The extent of wound closure was quantified using ImageJ. Bars represent the mean wound area ± S.D. from three independent experiments (B). ***$p < 0.001$, **$p < 0.01$, ANOVA followed by Tukey's HSD (n = 3). C. Parental, mock, stable transfectants were seeded in 96-well plates and grown for 72 hours. After 72 hours of incubation, MTT reagent was added and incubated for 3 hours. Formazan crystals were dissolved with DMSO and absorbance was measured at 570 nm and 690 nm. Bars represent the mean absorbance ± S.D. (n = 8).

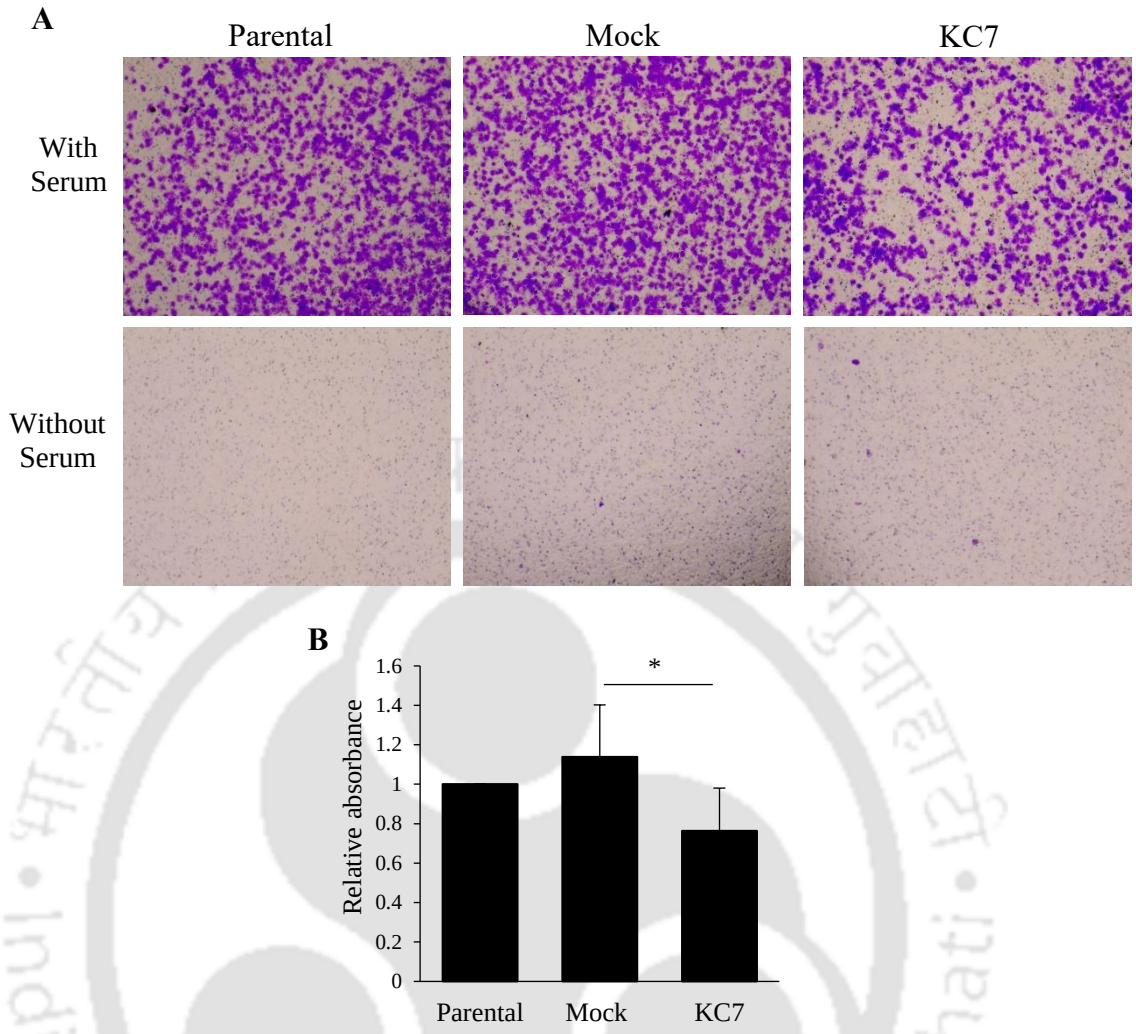

Figure 7.5. CSTA overexpression inhibited cell invasion in transwell invasion assay. A. Parental, mock, stable transfectants were seeded in collagen IV coated transwell insert and allowed to invade for 16 h. Cells with serum-free conditions on both sides of the membrane served as negative control. Representative images of stained invaded cells in parent, mock and stable transfectants. B. The invaded cells were quantified by dissolving the stain in 1% SDS. Absorbance was measured at 595 nm and bars represent relative absorbance ± S.D. *$p < 0.05$, one-way ANOVA followed by Tukey's test.

7.2.5. Effect of CSTA expression on EMT markers

To explore the effect of CSTA expression on EMT, the expression of EMT markers were assessed in mock and KC7 cells. Generally, mesenchymal phenotype is characterized by increased expression of vimentin, transcription factor-8 (TCF-8), Snail, β-catenin, N-cadherin, while epithelial phenotype is characterized by increased expression of Zonula occludens-1 (ZO-1), E-cadherin[334]. Western blotting analysis showed that KC7 showed decreased expression of mesenchymal markers (Vimentin, TCF-8, Snail, β-catenin and N-cadherin) (Figure 7.6A-E) and increased expression of epithelial markers (ZO-1) with respect to mock (Figure 7.6A).

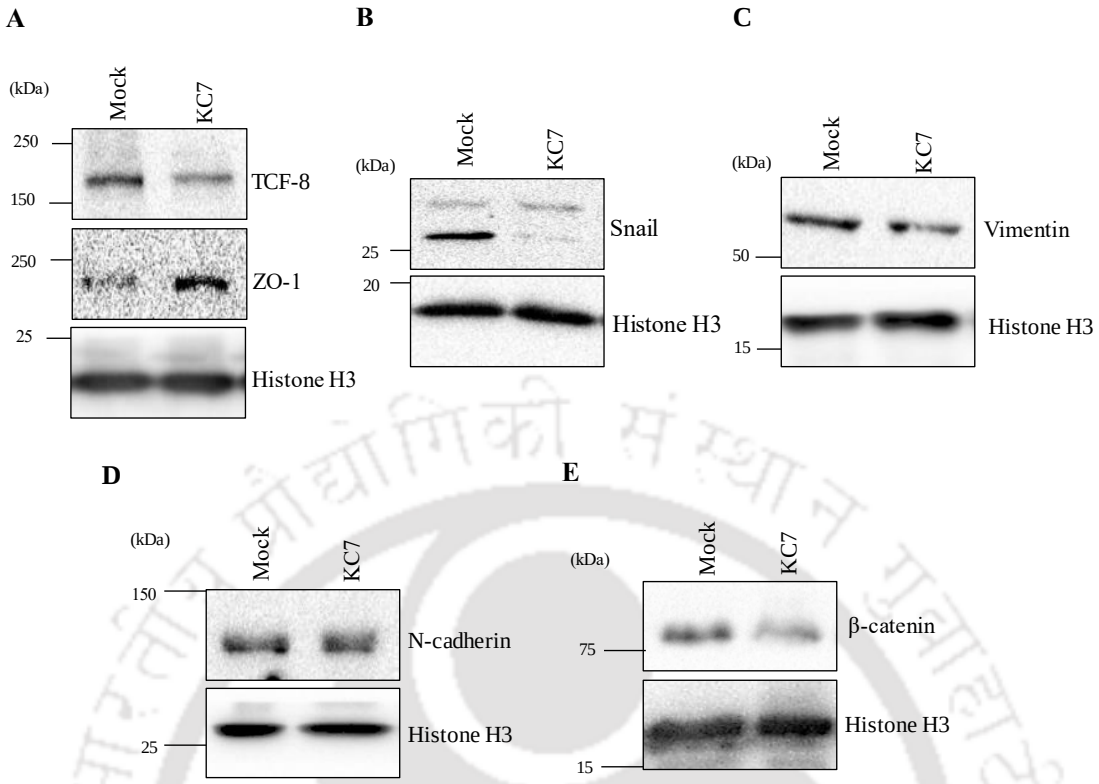

Figure 7.6. CSTA overexpression altered the expression of EMT markers in stable transfectants. A-E. MDA-MB-231 cells were stably transfected with CSTA expression construct or pBABE-puro vector (mock). Stable transfectants were selected with puromycin. Total protein was extracted from KC7 and mock and subjected to western blotting analysis with antibodies specific for EMT markers. Histone H3 was used as an internal control.

7.2.6. Effect of CSTA knockdown on EMT markers

Further, the expression of EMT markers was studied in MCF-7 cells transfected with scrambled or CSTA siRNA. CSTA knockdown was confirmed at mRNA and protein level using qRT-PCR and western blotting analysis, respectively. CSTA mRNA was significantly less (0.25 ± 0.03) in CSTA siRNA transfected cells compared to scrambled siRNA transfected cells (Figure 7.7A). CSTA protein was also reduced in MCF-7 cells upon transfection with CSTA siRNA (Figure 7.7B). Western blotting analysis showed increased expression of mesenchymal markers (Snail and β-catenin) in CSTA siRNA transfected cells (Figure 7.7C, D) compared to scrambled siRNA transfected cells. On the other hand, decreased expression of epithelial markers (ZO-1 and E-cadherin) was observed in CSTA siRNA transfected cells (Figure 7.7B, D).

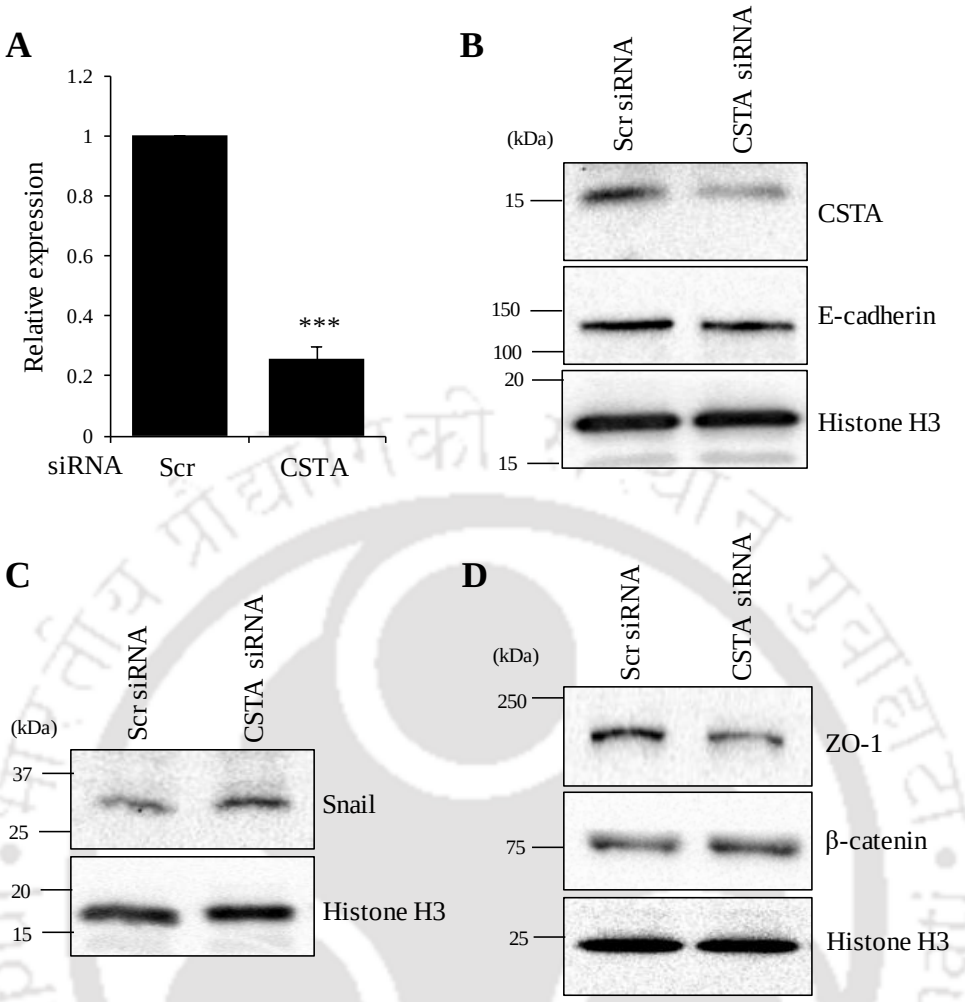

Figure 7.7. CSTA knockdown altered the expression of EMT markers. MCF-7 cells were transfected with scrambled siRNA or CSTA-specific siRNA for 24 h. Total RNA or protein was extracted subjected to qRT-PCR or western blotting analysis. Bars represent mean relative expression ± S.D. of CSTA with CycA mRNA serving as an internal control. ***$p < 0.001$, Welch two-sample *t*-test, n = 3. Histone H3 served as an internal control for western blots.

7.3. Discussion

In the present study, the functional role of CSTA in breast cancer was explored using two breast cancer cell lines, which differ in terms of invasiveness and CSTA expression. MDA-MB-231, which doesn't express CSTA, was used for generating CSTA expressing stable cell lines. On the other hand, since MCF-7 expresses CSTA, it was used in knockdown studies. The knockdown of CSTA from MCF-7 cells resulted in a mesenchymal phenotype characterized by an increase in expression of mesenchymal markers and a decrease in epithelial markers. Overexpression of CSTA in MDA-MB-231 cells resulted in an epithelial pattern of MET markers expression. Further, overexpression of CSTA reduced invasion and migration of cells, without affecting proliferation. This inhibitory effect of CSTA on migration indicated its possible role in MET. A similar observation about other cystatins such as cystatin C, D

and M are reported suggesting their involvement in inhibiting EMT. Transfection of cystatin C in melanoma cells (B16F10) reduced motility and *in vitro* invasiveness[335]. In MDA-MB-435S breast cancer cells, overexpression of cystatin M significantly reduced migration and invasion[336]. Implantation of cystatin M expressing cells into mice reduced tumor growth and metastatic burden[192]. Ectopic expression of cystatin D in colon cancer cells reduced migration and anchorage-independent growth[332]. Moreover, in esophageal squamous cell carcinoma cells, overexpression of CSTA delayed the *in vitro* and *in vivo* growth and metastasis. This is primarily through inhibition of cathepsin B activity. Further, a significant reduction in angiogenesis characterized by reduced factor VIII staining was observed in tumors of CSTA-expressing xenograft bearing mice[43]. Forced expression of CSTA in highly metastatic 4T1.2 drastically reduced bone metastasis[34]. In the same model, when cathepsin B was knockdown or selectively inhibited, metastasis to bone and lung was reduced[35]. These studies hints that the observed CSTA inhibition of tumor migration invasion is probably due to the inhibition of cathepsin B. However, in the present study, CSTA expression did not have a profound effect on the proliferation of breast cancer cells. Nevertheless, its effect on proliferation could be more appreciable in the tumor microenvironment.

Interestingly, in addition to the primary function of cystatins to regulate cathepsins, recent evidence indicate cystatins may affect tumor progression in a proteolysis-independent manner. Ma and co-workers reported that CSTA overexpression induces MET in lung cancer by inhibiting ERK/MAPK pathway[42]. Moreover, cystatin C reduces TGF-β-mediated tumor progression by partly inhibiting smad2, p38 MAPK and ERK1/2 phosphorylation[337]. Cst5, a gene encoding cystatin D is the putative target of p53. Calcitriol (an active metabolite of vitamin D) activates p53, which in turn induces cystatin D resulting in the repression of snail, an EMT inducing transcription factor[338]. One of the striking observations made by Ferrer-Mayorga and co-workers is the detection of the proportion of cystatin D in the cell nucleus at the transcriptionally active sites of chromatin. In addition to its regulatory role on RUNX1 (Runt-related transcription factor 1), RUNX2, and MEF2C (myocyte-specific enhancer factor 2C), cystatin D also reduces the secretion of pro-tumor cytokines[339].

Considering these results in the light of existing literature, it can be suggested that, like other cystatins, CSTA is likely to play a crucial role in breast tumor invasion and metastasis either by inhibiting cathepsin activity directly or by modulating other signaling pathways. Further investigations on the mechanism behind the inhibition of migration and invasion by CSTA may unravel its therapeutic potential.

CHAPTER 8

Conclusion

The most perilous attribute of malignant tumors is to metastasize, which directly impacts survival. The potential of tumor cells to invade and metastasize is based on their ability to degrade surrounding components of the ECM. Overexpression of proteolytic enzymes is significantly associated with the metastatic progression of tumors cells. Cathepsins, the lysosomal cysteine proteases, have been reported to be increasingly expressed in various types of tumors. Altered expression of cysteine cathepsins tilts the homeostatic balance to favor ECM remodeling, thereby promoting tumor progression, invasion, and metastasis. Endogenous inhibitors called cystatins naturally regulate the activity of these proteases. Due to the inhibitory activity against cysteine cathepsins, cystatins are considered as tumor suppressors.

The knowledge of the exact role of cystatins in cancer has been expanding over the years. However, scanty literature presents contradictory views on the role of the first member of cystatin superfamily, CSTA, in breast cancer. The possible reason could be the heterogenicity in the intrinsic subtype of the cohorts. This necessitated an independent study on its prognostic potential by taking into consideration the various molecular subtypes of breast tumors. Further, this study attempted a synthesis of the present results and the existing information to reflect on the ambiguity in the anticipated role of CSTA in breast cancer development and progression. It also pivots on the regulation of CSTA expression in breast cancer.

In silico analysis of publicly available TCGA data brought about interesting observations. In luminal A, higher CSTA expression was correlated with reduced survival. While in luminal B, higher CSTA expression is correlated with prolonged survival. HER2+ and basal tumor subtypes, CSTA expression, was not associated with survival, indicating that the effect of CSTA on survival is tumor subtype dependent. This study highlights that the ambiguity in the apparent role of CSTA in breast cancer development and progression possibly indicates a dual role: as a tumor suppressor and as a promoter of aggressive phenotype, depending on the breast cancer molecular subtype. An inverse correlation was observed between CSTA and ERα expression in primary breast tumors, which provided compelling evidence in favor of a functional link between CSTA and ERα and offered a rationale for investigating estrogen-mediated regulation of CSTA.

This study unveils the essential role of ERα in estrogen-mediated suppression of CSTA expression in breast cancer cells. *In vitro* experiments showed estrogen suppresses CSTA expression in MCF-7 and ZR-75-1 cells. Estrogen-mediated suppression of CSTA expression in breast cancer cells occurs by binding of ERα to the intron-2 region of CSTA. However, the extent of suppression varies across cell lines. The reason behind the differential effects of estrogen on CSTA expression in ERα-positive breast cancer cell lines was partially due to variation in the relative ERα levels in these cell lines. However, in T47D cells which express high ERα, estrogen did not modulate CSTA expression.

Further, TCGA-BRCA analysis revealed that CSTA expression in primary breast tumors is significantly less than normal breast tissues. DNA methylation in the intron-2 of CSTA locus is inversely related to CSTA expression in breast cancer cells, explaining the reason behind the loss of CSTA expression in breast tumors. This result may be exploited for predicting CSTA expression in the metastatic progression of breast tumors. Moreover, global demethylation restores estrogen regulation of CSTA in T47D and MDA-MB-231 cells. This unveiled the interesting interplay between ERα binding and transcriptional regulation in the CSTA locus. Therefore, the proposed model of this study is, CSTA expression in breast cancer cells is an integrated result of estrogen regulation and DNA methylation-dependent silencing converging on the intron-2.

This study also attempted to understand the possible role of CSTA in breast cancer using stable cell lines. Overexpression of CSTA in breast cancer cells reduced migration and invasion of breast cancer cells without affecting proliferation. Moreover, the expression of

CSTA promoted the epithelial phenotype, and knockdown promoted mesenchymal phenotype in breast cancer cells.

Taken together, this work offers novel insights into the regulation of CSTA expression in breast cancer. This is the first study to provide detailed molecular insights into the estrogen-mediated regulation of CSTA. Further, it provides enough evidence that DNA methylation is the probable reason for the loss of CSTA expression in breast tumors. This work also proposes the potential interplay between ERα binding and DNA methylation in the regulation of CSTA expression.

Considering the subtype-dependent effect of CSTA on survival, CSTA can be used to predict the tumor relapse and survival of breast cancer patients. Methylation at intron-2 can serve as a prognostic marker to assess the clinical behavior of breast tumors. Given the critical role of CSTA in the inhibition of cathepsins, epigenetic restoration of CSTA expression using pharmacological agents can suppress tumor progression in luminal B tumors. Besides, the potential of CSTA to downmodulate breast tumor invasion and metastasis has therapeutic significance and requires further detailed investigation.

R

References

1 Goldfarb, R. H. & Liotta, L. A. Proteolytic enzymes in cancer invasion and metastasis. *Semin Thromb Hemost* **12**, 294-307, doi:10.1055/s-2007-1003570 (1986).

2 Liotta, L. A. & Stetler-Stevenson, W. G. Tumor invasion and metastasis: an imbalance of positive and negative regulation. *Cancer Res* **51**, 5054s-5059s (1991).

3 Palermo, C. & Joyce, J. A. Cysteine cathepsin proteases as pharmacological targets in cancer. *Trends Pharmacol Sci* **29**, 22-28, doi:10.1016/j.tips.2007.10.011 (2008).

4 Vidak, E., Javorsek, U., Vizovisek, M. & Turk, B. Cysteine Cathepsins and their Extracellular Roles: Shaping the Microenvironment. *Cells* **8**, doi:10.3390/cells8030264 (2019).

5 Khaket, T. P., Kwon, T. K. & Kang, S. C. Cathepsins: Potent regulators in carcinogenesis. *Pharmacol Ther* **198**, 1-19, doi:10.1016/j.pharmthera.2019.02.003 (2019).

6 Turk, V. *et al.* Cysteine cathepsins: from structure, function and regulation to new frontiers. *Biochim Biophys Acta* **1824**, 68-88, doi:10.1016/j.bbapap.2011.10.002 (2012).

7 Turk, B., Turk, D. & Turk, V. Lysosomal cysteine proteases: more than scavengers. *Biochim Biophys Acta* **1477**, 98-111, doi:10.1016/s0167-4838(99)00263-0 (2000).

8 Fonovic, M. & Turk, B. Cysteine cathepsins and extracellular matrix degradation. *Biochim Biophys Acta* **1840**, 2560-2570, doi:10.1016/j.bbagen.2014.03.017 (2014).

9 Vizovisek, M., Fonovic, M. & Turk, B. Cysteine cathepsins in extracellular matrix remodeling: Extracellular matrix degradation and beyond. *Matrix Biol* **75-76**, 141-159, doi:10.1016/j.matbio.2018.01.024 (2019).

10 Pogorzelska, A., Zolnowska, B. & Bartoszewski, R. Cysteine cathepsins as a prospective target for anticancer therapies-current progress and prospects. *Biochimie* **151**, 85-106, doi:10.1016/j.biochi.2018.05.023 (2018).

11 Kolwijck, E. *et al.* The balance between extracellular cathepsins and cystatin C is of importance for ovarian cancer. *Eur J Clin Invest* **40**, 591-599, doi:10.1111/j.1365-2362.2010.02305.x (2010).

12 Paraoan, L. *et al.* Cathepsin S and its inhibitor cystatin C: imbalance in uveal melanoma. *Front Biosci (Landmark Ed)* **14**, 2504-2513, doi:10.2741/3393 (2009).

13 Yano, M. *et al.* Expression of cathepsin B and cystatin C in human breast cancer. *Surg Today* **31**, 385-389, doi:10.1007/s005950170126 (2001).

14 Rivenbark, A. G. & Coleman, W. B. Epigenetic regulation of cystatins in cancer. *Front Biosci (Landmark Ed)* **14**, 453-462, doi:10.2741/3254 (2009).

15 Thomssen, C. *et al.* Prognostic value of the cysteine proteases cathepsins B and cathepsin L in human breast cancer. *Clin Cancer Res* **1**, 741-746 (1995).

16 Kos, J. *et al.* Cathepsins B, H, and L and their inhibitors stefin A and cystatin C in sera of melanoma patients. *Clin Cancer Res* **3**, 1815-1822 (1997).

17 Duivenvoorden, H. M. *et al.* Myoepithelial cell-specific expression of stefin A as a suppressor of early breast cancer invasion. *J Pathol* **243**, 496-509, doi:10.1002/path.4990 (2017).

18 Jones, C. *et al.* Expression profiling of purified normal human luminal and myoepithelial breast cells: identification of novel prognostic markers for breast cancer. *Cancer Res* **64**, 3037-3045, doi:10.1158/0008-5472.can-03-2028 (2004).

19 Sinha, A. A. *et al.* Prediction of pelvic lymph node metastasis by the ratio of cathepsin B to stefin A in patients with prostate carcinoma. *Cancer* **94**, 3141-3149, doi:10.1002/cncr.10604 (2002).

20 Strojnik, T. *et al.* Cathepsin B and its inhibitor stefin A in brain tumors. *Pflugers Arch* **439**, R122-123 (2000).

21 Clemons, M. & Goss, P. Estrogen and the risk of breast cancer. *N Engl J Med* **344**, 276-285, doi:10.1056/nejm200101253440407 (2001).

22 Bernstein, L. & Ross, R. K. Endogenous hormones and breast cancer risk. *Epidemiol Rev* **15**, 48-65, doi:10.1093/oxfordjournals.epirev.a036116 (1993).

23 McEwen, B. S. & Alves, S. E. Estrogen actions in the central nervous system. *Endocr Rev* **20**, 279-307, doi:10.1210/edrv.20.3.0365 (1999).

24 Turner, R. T., Riggs, B. L. & Spelsberg, T. C. Skeletal effects of estrogen. *Endocr Rev* **15**, 275-300, doi:10.1210/edrv-15-3-275 (1994).

25 Tostes, R. C., Nigro, D., Fortes, Z. B. & Carvalho, M. H. Effects of estrogen on the vascular system. *Braz J Med Biol Res* **36**, 1143-1158, doi:10.1590/s0100-879x2003000900002 (2003).

26 Lang, T. J. Estrogen as an immunomodulator. *Clin Immunol* **113**, 224-230, doi:10.1016/j.clim.2004.05.011 (2004).

27 Vendrell, J. A. *et al.* Estrogen regulation in human breast cancer cells of new downstream gene targets involved in estrogen metabolism, cell proliferation and cell transformation. *J Mol Endocrinol* **32**, 397-414, doi:10.1677/jme.0.0320397 (2004).

28 Dai, X., Xiang, L., Li, T. & Bai, Z. Cancer hallmarks, biomarkers and breast cancer molecular subtypes. *Journal of Cancer* **7**, 1281 (2016).

29 Zhang, S. J. *et al.* Expression and significance of ER, PR, VEGF, CA15-3, CA125 and CEA in judging the prognosis of breast cancer. *Asian Pac J Cancer Prev* **14**, 3937-3940, doi:10.7314/apjcp.2013.14.6.3937 (2013).

30 Lah, T. T. *et al.* The expression of lysosomal proteinases and their inhibitors in breast cancer: possible relationship to prognosis of the disease. *Pathol Oncol Res* **3**, 89-99, doi:10.1007/bf02907801 (1997).

31 Kuopio, T. *et al.* Cysteine proteinase inhibitor cystatin A in breast cancer. *Cancer Res* **58**, 432-436 (1998).

32 Levicar, N. *et al.* Comparison of potential biological markers cathepsin B, cathepsin L, stefin A and stefin B with urokinase and plasminogen activator inhibitor-1 and clinicopathological data of breast carcinoma patients. *Cancer Detect Prev* **26**, 42-49, doi:10.1016/s0361-090x(02)00015-6 (2002).

33 Buzdar, A. U. & Hortobagyi, G. N. Tamoxifen and toremifene in breast cancer: comparison of safety and efficacy. *J Clin Oncol* **16**, 348-353, doi:10.1200/jco.1998.16.1.348 (1998).

34 Parker, B. S. *et al.* Primary tumour expression of the cysteine cathepsin inhibitor Stefin A inhibits distant metastasis in breast cancer. *J Pathol* **214**, 337-346, doi:10.1002/path.2265 (2008).

35 Withana, N. P. *et al.* Cathepsin B inhibition limits bone metastasis in breast cancer. *Cancer Res* **72**, 1199-1209, doi:10.1158/0008-5472.Can-11-2759 (2012).

36 Esteller, M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet* **8**, 286-298, doi:10.1038/nrg2005 (2007).

37 Garinis, G. A., Patrinos, G. P., Spanakis, N. E. & Menounos, P. G. DNA hypermethylation: when tumour suppressor genes go silent. *Hum Genet* **111**, 115-127, doi:10.1007/s00439-002-0783-6 (2002).

38 Stone, A. *et al.* DNA methylation of oestrogen-regulated enhancers defines endocrine sensitivity in breast cancer. *Nat Commun* **6**, 7758, doi:10.1038/ncomms8758 (2015).

39 Anastasi, A. *et al.* Cystatin, a protein inhibitor of cysteine proteinases. Improved purification from egg white, characterization, and detection in chicken serum. *Biochem J* **211**, 129-138, doi:10.1042/bj2110129 (1983).

40 Blaydon, D. C. *et al.* Mutations in CSTA, encoding Cystatin A, underlie exfoliative ichthyosis and reveal a role for this protease inhibitor in cell-cell adhesion. *Am J Hum Genet* **89**, 564-571, doi:10.1016/j.ajhg.2011.09.001 (2011).

41 Jones, B., Roberts, P. J., Faubion, W. A., Kominami, E. & Gores, G. J. Cystatin A expression reduces bile salt-induced apoptosis in a rat hepatoma cell line. *Am J Physiol* **275**, G723-730, doi:10.1152/ajpgi.1998.275.4.G723 (1998).

42 Ma, Y. *et al.* Cystatin A suppresses tumor cell growth through inhibiting epithelial to mesenchymal transition in human lung cancer. *Oncotarget* **9**, 14084-14098, doi:10.18632/oncotarget.23505 (2018).

43 Li, W. *et al.* Overexpression of stefin A in human esophageal squamous cell carcinoma cells inhibits tumor cell growth, angiogenesis, invasion, and metastasis. *Clin Cancer Res* **11**, 8753-8762, doi:10.1158/1078-0432.Ccr-05-0597 (2005).

44 Ferlay, J. *et al. Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer*, <https://gco.iarc.fr/today> (2018).

45 Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **68**, 394-424, doi:10.3322/caac.21492 (2018).

46 Malvia, S., Bagadi, S. A., Dubey, U. S. & Saxena, S. Epidemiology of breast cancer in Indian women. *Asia Pac J Clin Oncol* **13**, 289-295, doi:10.1111/ajco.12661 (2017).

47 Agarwal, G. & Ramakant, P. Breast cancer care in India: The current scenario and the challenges for the future. *Breast Care (Basel)* **3**, 21-27, doi:10.1159/000115288 (2008).

48 Leong, S. P. *et al.* Is breast cancer the same disease in Asian and Western countries? *World J Surg* **34**, 2308-2324, doi:10.1007/s00268-010-0683-1 (2010).

49 Formenti, S. C., Arslan, A. A. & Love, S. M. Global breast cancer: the lessons to bring home. *Int J Breast Cancer* **2012**, 249501 (2012).

50 National Cancer Registry Programme (ICMR). *Consolidated report of Hospital Based Cancer Registries 2012-2014*, <https://ncdirindia.org/ncrp/ALL_NCRP_REPORTS/HBCR_REPORT_2012_2014/index.htm> (2016).

51 National Centre for Disease Informatics and Research. *A Report on Cancer Burden in North Eastern States of India*, <https://www.ncdirindia.org/All_Reports/Reports_Ne/NE2012_2014/Files/NE_2012_14.pdf> (2017).

52 Sharma, J. D., Kataki, A. C., Barman, D., Sharma, A. & Kalita, M. Cancer statistics in Kamrup urban district: Incidence and mortality in 2007-2011. *Indian J Cancer* **53**, 600-606, doi:10.4103/0019-509x.204764 (2016).

53 Gucalp, A. *et al.* Male breast cancer: a disease distinct from female breast cancer. *Breast Cancer Res Treat* **173**, 37-48, doi:10.1007/s10549-018-4921-9 (2019).

54 Sinn, H. P. & Kreipe, H. A brief overview of the WHO classification of breast tumors, 4th edition, focusing on issues and updates from the 3rd edition. *Breast Care (Basel)* **8**, 149-154, doi:10.1159/000350774 (2013).

55 Sorlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98**, 10869-10874, doi:10.1073/pnas.191367098 (2001).

56 Hammond, M. E., Hayes, D. F., Wolff, A. C., Mangu, P. B. & Temin, S. American society of clinical oncology/college of american pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J Oncol Pract* **6**, 195-197, doi:10.1200/jop.777003 (2010).

57 Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747-752, doi:10.1038/35021093 (2000).

58 Prat, A. *et al.* Clinical implications of the intrinsic molecular subtypes of breast cancer. *Breast* **24 Suppl 2**, S26-35, doi:10.1016/j.breast.2015.07.008 (2015).

59 Hu, Z. *et al.* The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* **7**, 96, doi:10.1186/1471-2164-7-96 (2006).

60 Goldhirsch, A. *et al.* Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol* **22**, 1736-1747, doi:10.1093/annonc/mdr304 (2011).

61 Denoix, P. Tumor, node and metastasis (TNM). *Bull Inst Nat Hyg* **1**, 1-69 (1944).

62 Sobin, L. H. TNM: evolution and relation to other prognostic factors. *Semin Surg Oncol* **21**, 3-7, doi:10.1002/ssu.10014 (2003).

63 Ludwig, J. A. & Weinstein, J. N. Biomarkers in cancer staging, prognosis and treatment selection. *Nat Rev Cancer* **5**, 845-856, doi:10.1038/nrc1739 (2005).

64 Rakha, E. A. *et al.* Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Res* **12**, 207, doi:10.1186/bcr2607 (2010).

65 Elston, C. W. & Ellis, I. O. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* **19**, 403-410, doi:10.1111/j.1365-2559.1991.tb00229.x (1991).

66 Kelsey, J. L., Gammon, M. D. & John, E. M. Reproductive factors and breast cancer. *Epidemiol Rev* **15**, 36-47, doi:10.1093/oxfordjournals.epirev.a036115 (1993).

67 Henderson, B. E., Ross, R. & Bernstein, L. Estrogens as a cause of human cancer: the Richard and Hinda Rosenthal Foundation award lecture. *Cancer Res* **48**, 246-253 (1988).

68 Pike, M. C., Spicer, D. V., Dahmoush, L. & Press, M. F. Estrogens, progestogens, normal breast cell proliferation, and breast cancer risk. *Epidemiol Rev* **15**, 17-35, doi:10.1093/oxfordjournals.epirev.a036102 (1993).

69 Hulka, B. S. Epidemiologic analysis of breast and gynecologic cancers. *Prog Clin Biol Res* **396**, 17-29 (1997).

70 Zumoff, B. Does postmenopausal estrogen administration increase the risk of breast cancer? Contributions of animal, biochemical, and clinical investigative studies to a resolution of the controversy. *Proc Soc Exp Biol Med* **217**, 30-37, doi:10.3181/00379727-217-44202 (1998).

71 Russo, I. H. & Russo, J. Role of hormones in mammary cancer initiation and progression. *J Mammary Gland Biol Neoplasia* **3**, 49-61, doi:10.1023/a:1018770218022 (1998).

72 Sledge, G. W. *et al.* Past, present, and future challenges in breast cancer treatment. *J Clin Oncol* **32**, 1979-1986, doi:10.1200/jco.2014.55.4139 (2014).

73 Yang, T. J. & Ho, A. Y. Radiation therapy in the management of breast cancer. *Surg Clin North Am* **93**, 455-471, doi:10.1016/j.suc.2013.01.002 (2013).

74 Slamon, D. J. *et al.* Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* **344**, 783-792, doi:10.1056/nejm200103153441101 (2001).

75 Tinoco, G., Warsch, S., Gluck, S., Avancha, K. & Montero, A. J. Treating breast cancer in the 21st century: emerging biological therapies. *J Cancer* **4**, 117-132, doi:10.7150/jca.4925 (2013).

76 Peto, R. *et al.* Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. *Lancet* **379**, 432-444, doi:10.1016/s0140-6736(11)61625-5 (2012).

77 Beatson, G. T. On the treatment of inoperable cases of carcinoma of the mamma: Suggestions for a new method of treatment, with illustrative cases. *Trans Med Chir Soc Edinb* **15**, 153-179 (1896).

78 Allen, E. & Doisy, E. A. An ovarian hormone: Preliminary report on its localization, extraction and partial purification, and action in test animals. *JAMA* **81**, 819-821 (1923).

79 Astwood, E. B. Time relationships in the growth and water exchange of the uterus following estrogenic stimulation. *Anat. Record Suppl.* **70**, 5 (1938).

80 Talbot, N., Lowry, O. H. & Astwood, E. Influence of estrogen on the electrolyte pattern of the immature rat uterus. *J biol Chem* **132**, 1 (1940).

81 Clifton, K. H. & Meyer, R. K. Mechanism of anterior pituitary tumor induction by estrogen. *Anat Rec* **125**, 65-81 (1956).

82 Cui, J., Shen, Y. & Li, R. Estrogen synthesis and signaling pathways during aging: from periphery to brain. *Trends Mol Med* **19**, 197-209, doi:10.1016/j.molmed.2012.12.007 (2013).

83 Russo, J. & Russo, I. H. The role of estrogen in the initiation of breast cancer. *J Steroid Biochem Mol Biol* **102**, 89-96, doi:10.1016/j.jsbmb.2006.09.004 (2006).

84 Hamilton, T. H. Control by estrogen of genetic transcription and translation. *Science* (1968).

85 Yager, J. D. & Leihr, J. Molecular mechanisms of estrogen carcinogenesis. *Annu Rev Pharmacol Toxicol* **36**, 203-232 (1996).

86 Barton, M. *et al.* Twenty years of the G protein-coupled estrogen receptor GPER: Historical and personal perspectives. *J Steroid Biochem Mol Biol* **176**, 4-15, doi:10.1016/j.jsbmb.2017.03.021 (2018).

87 Attar, E. & Bulun, S. E. Aromatase inhibitors: the next generation of therapeutics for endometriosis? *Fertil Steril* **85**, 1307-1318, doi:10.1016/j.fertnstert.2005.09.064 (2006).

88 Harper, M. J. & Walpole, A. L. A new derivative of triphenylethylene: effect on implantation and mode of action in rats. *J Reprod Fertil* **13**, 101-119, doi:10.1530/jrf.0.0130101 (1967).

89 Snyder, K. R., Sparano, N. & Malinowski, J. M. Raloxifene hydrochloride. *Am J Health Syst Pharm* **57**, 1669-1675; quiz 1676-1668 (2000).

90 Wakeling, A. E., Dukes, M. & Bowler, J. A potent specific pure antiestrogen with clinical potential. *Cancer Res* **51**, 3867-3873 (1991).

91 Patel, H. K. & Bihani, T. Selective estrogen receptor modulators (SERMs) and selective estrogen receptor degraders (SERDs) in cancer treatment. *Pharmacol Ther* **186**, 1-24, doi:10.1016/j.pharmthera.2017.12.012 (2018).

92 O'Malley, B. W. & Khan, S. Elwood V. Jensen (1920-2012): father of the nuclear receptors. *Proc Natl Acad Sci U S A* **110**, 3707-3708, doi:10.1073/pnas.1301566110 (2013).

93 Green, S. *et al.* Cloning of the human oestrogen receptor cDNA. *J Steroid Biochem* **24**, 77-83, doi:10.1016/0022-4731(86)90035-x (1986).

94 Kuiper, G. G., Enmark, E., Pelto-Huikko, M., Nilsson, S. & Gustafsson, J. A. Cloning of a novel receptor expressed in rat prostate and ovary. *Proc Natl Acad Sci U S A* **93**, 5925-5930, doi:10.1073/pnas.93.12.5925 (1996).

95 Szego, C. M. & Davis, J. S. Adenosine 3',5'-monophosphate in rat uterus: acute elevation by estrogen. *Proc Natl Acad Sci U S A* **58**, 1711-1718, doi:10.1073/pnas.58.4.1711 (1967).

96 Pietras, R. J. & Szego, C. M. Specific binding sites for oestrogen at the outer surfaces of isolated endometrial cells. *Nature* **265**, 69-72 (1977).

97 McCoy, R. & Permutter, D. Cloning of novel IL8-related receptors from human hepatic tissue. *EMBL/GenBank/DDBJ databases* (1996).

98 Bonini, J. A., Anderson, S. M. & Steiner, D. F. Molecular cloning and tissue expression of a novel orphan G protein-coupled receptor from rat lung. *Biochem Biophys Res Commun* **234**, 190-193, doi:10.1006/bbrc.1997.6591 (1997).

99 Feng, Y. & Gregor, P. Cloning of a novel member of the G protein-coupled receptor family related to peptide receptors. *Biochem Biophys Res Commun* **231**, 651-654, doi:10.1006/bbrc.1997.6161 (1997).

100 Carmeci, C., Thompson, D. A., Ring, H. Z., Francke, U. & Weigel, R. J. Identification of a gene (GPR30) with homology to the G-protein-coupled receptor superfamily associated with estrogen receptor expression in breast cancer. *Genomics* **45**, 607-617, doi:10.1006/geno.1997.4972 (1997).

101 Revankar, C. M., Cimino, D. F., Sklar, L. A., Arterburn, J. B. & Prossnitz, E. R. A transmembrane intracellular estrogen receptor mediates rapid cell signaling. *Science* **307**, 1625-1630, doi:10.1126/science.1106943 (2005).

102 Filardo, E. J., Quinn, J. A., Bland, K. I. & Frackelton, A. R., Jr. Estrogen-induced activation of Erk-1 and Erk-2 requires the G protein-coupled receptor homolog, GPR30, and occurs via trans-activation of the epidermal growth factor receptor through release of HB-EGF. *Mol Endocrinol* **14**, 1649-1660, doi:10.1210/mend.14.10.0532 (2000).

103 Prossnitz, E. R. & Arterburn, J. B. International Union of Basic and Clinical Pharmacology. XCVII. G protein-coupled estrogen receptor and its pharmacologic modulators. *Pharmacol Rev* **67**, 505-540, doi:10.1124/pr.114.009712 (2015).

104 Thomas, P. *et al.* Conserved estrogen binding and signaling functions of the G protein-coupled estrogen receptor 1 (GPER) in mammals and fish. *Steroids* **75**, 595-602, doi:10.1016/j.steroids.2009.11.005 (2010).

105 Eckert, R. L. & Katzenellenbogen, B. S. Physical properties of estrogen receptor complexes in MCF-7 human breast cancer cells. Differences with anti-estrogen and estrogen. *J Biol Chem* **257**, 8840-8846 (1982).

106 Lee, H. R., Kim, T. H. & Choi, K. C. Functions and physiological roles of two types of estrogen receptors, ERalpha and ERbeta, identified by estrogen receptor knockout mouse. *Lab Anim Res* **28**, 71-76, doi:10.5625/lar.2012.28.2.71 (2012).

107 Matthews, J. & Gustafsson, J. A. Estrogen signaling: a subtle balance between ER alpha and ER beta. *Mol Interv* **3**, 281-292, doi:10.1124/mi.3.5.281 (2003).

108 Pearce, S. T. & Jordan, V. C. The biological role of estrogen receptors alpha and beta in cancer. *Crit Rev Oncol Hematol* **50**, 3-22, doi:10.1016/j.critrevonc.2003.09.003 (2004).

109 Delaunay, F., Pettersson, K., Tujague, M. & Gustafsson, J. A. Functional differences between the amino-terminal domains of estrogen receptors alpha and beta. *Mol Pharmacol* **58**, 584-590, doi:10.1124/mol.58.3.584 (2000).

110 Barkhem, T. *et al.* Differential response of estrogen receptor alpha and estrogen receptor beta to partial estrogen agonists/antagonists. *Mol Pharmacol* **54**, 105-112, doi:10.1124/mol.54.1.105 (1998).

111 Paech, K. *et al.* Differential ligand activation of estrogen receptors ERalpha and ERbeta at AP1 sites. *Science* **277**, 1508-1510, doi:10.1126/science.277.5331.1508 (1997).

112 Webb, P. *et al.* The estrogen receptor enhances AP-1 activity by two distinct mechanisms with different requirements for receptor transactivation functions. *Mol Endocrinol* **13**, 1672-1685, doi:10.1210/mend.13.10.0357 (1999).

113 Suen, C. S. *et al.* A transcriptional coactivator, steroid receptor coactivator-3, selectively augments steroid receptor transcriptional activity. *J Biol Chem* **273**, 27645-27653, doi:10.1074/jbc.273.42.27645 (1998).

114 Haldosen, L. A., Zhao, C. & Dahlman-Wright, K. Estrogen receptor beta in breast cancer. *Mol Cell Endocrinol* **382**, 665-672, doi:10.1016/j.mce.2013.08.005 (2014).

115 Leung, Y. K., Mak, P., Hassan, S. & Ho, S. M. Estrogen receptor (ER)-beta isoforms: a key to understanding ER-beta signaling. *Proc Natl Acad Sci U S A* **103**, 13162-13167, doi:10.1073/pnas.0605676103 (2006).

116 Chang, E. C., Frasor, J., Komm, B. & Katzenellenbogen, B. S. Impact of estrogen receptor β on gene networks regulated by estrogen receptor α in breast cancer cells. *Endocrinology* **147**, 4831-4842 (2006).

117 Hanstein, B., Liu, H., Yancisin, M. C. & Brown, M. Functional analysis of a novel estrogen receptor-beta isoform. *Mol Endocrinol* **13**, 129-137, doi:10.1210/mend.13.1.0234 (1999).

118 Poola, I., Abraham, J., Baldwin, K., Saunders, A. & Bhatnagar, R. Estrogen receptors beta4 and beta5 are full length functionally distinct ERbeta isoforms: cloning from human ovary and functional characterization. *Endocrine* **27**, 227-238, doi:10.1385/endo:27:3:227 (2005).

119 Wang, Z. *et al.* A variant of estrogen receptor-{alpha}, hER-{alpha}36: transduction of estrogen- and antiestrogen-dependent membrane-initiated mitogenic signaling. *Proc Natl Acad Sci U S A* **103**, 9063-9068, doi:10.1073/pnas.0603339103 (2006).

120 Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685, doi:10.1038/227680a0 (1970).

121 Lee, L. M. *et al.* ER-alpha36, a novel variant of ER-alpha, is expressed in ER-positive and -negative human breast carcinomas. *Anticancer Res* **28**, 479-483 (2008).

122 Shi, L. *et al.* Expression of ER-{alpha}36, a novel variant of estrogen receptor {alpha}, and resistance to tamoxifen treatment in breast cancer. *J Clin Oncol* **27**, 3423-3429, doi:10.1200/jco.2008.17.2254 (2009).

123 Zhang, X. T. *et al.* A positive feedback loop of ER-alpha36/EGFR promotes malignant growth of ER-negative breast cancer cells. *Oncogene* **30**, 770-780, doi:10.1038/onc.2010.458 (2011).

124 Klinge, C. M. *et al.* Estrogen receptor alpha 46 is reduced in tamoxifen resistant breast cancer cells and re-expression inhibits cell proliferation and estrogen receptor alpha 66-regulated target gene transcription. *Mol Cell Endocrinol* **323**, 268-276, doi:10.1016/j.mce.2010.03.013 (2010).

125 Penot, G. *et al.* The human estrogen receptor-alpha isoform hERalpha46 antagonizes the proliferative influence of hERalpha66 in MCF7 breast cancer cells. *Endocrinology* **146**, 5474-5484, doi:10.1210/en.2005-0866 (2005).

126 Sabbah, M., Radanyi, C., Redeuilh, G. & Baulieu, E. E. The 90 kDa heat-shock protein (hsp90) modulates the binding of the oestrogen receptor to its cognate DNA. *Biochem J* **314(Pt 1)**, 205-213, doi:10.1042/bj3140205 (1996).

127 Le Romancer, M. *et al.* Cracking the estrogen receptor's posttranslational code in breast tumors. *Endocr Rev* **32**, 597-622, doi:10.1210/er.2010-0016 (2011).

128 Edwards, D. P. The role of coactivators and corepressors in the biology and mechanism of action of steroid hormone receptors. *J Mammary Gland Biol Neoplasia* **5**, 307-324, doi:10.1023/a:1009503029176 (2000).

129 Kojetin, D. J. & Burris, T. P. Small molecule modulation of nuclear receptor conformational dynamics: implications for function and drug discovery. *Mol Pharmacol* **83**, 1-8, doi:10.1124/mol.112.079285 (2013).

130 Kushner, P. J. *et al.* Estrogen receptor pathways to AP-1. *J Steroid Biochem Mol Biol* **74**, 311-317, doi:10.1016/s0960-0760(00)00108-4 (2000).

131 Sabbah, M., Courilleau, D., Mester, J. & Redeuilh, G. Estrogen induction of the cyclin D1 promoter: involvement of a cAMP response-like element. *Proc Natl Acad Sci U S A* **96**, 11217-11222, doi:10.1073/pnas.96.20.11217 (1999).

132 Saville, B. *et al.* Ligand-, cell-, and estrogen receptor subtype (alpha/beta)-dependent activation at GC-rich (Sp1) promoter elements. *J Biol Chem* **275**, 5379-5387, doi:10.1074/jbc.275.8.5379 (2000).

133 Kato, S. *et al.* Activation of the estrogen receptor through phosphorylation by mitogen-activated protein kinase. *Science* **270**, 1491-1494, doi:10.1126/science.270.5241.1491 (1995).

134 Bunone, G., Briand, P. A., Miksicek, R. J. & Picard, D. Activation of the unliganded estrogen receptor by EGF involves the MAP kinase pathway and direct phosphorylation. *Embo J* **15**, 2174-2183 (1996).

135 Tremblay, A., Tremblay, G. B., Labrie, F. & Giguère, V. Ligand-independent recruitment of SRC-1 to estrogen receptor β through phosphorylation of activation function AF-1. *J Molecular cell* **3**, 513-519 (1999).

136 Filardo, E. *et al.* Activation of the novel estrogen receptor G protein-coupled receptor 30 (GPR30) at the plasma membrane. *Endocrinology* **148**, 3236-3245, doi:10.1210/en.2006-1605 (2007).

137 Filardo, E. J. Epidermal growth factor receptor (EGFR) transactivation by estrogen via the G-protein-coupled receptor, GPR30: a novel signaling pathway with potential significance for breast cancer. *J Steroid Biochem Mol Biol* **80**, 231-238, doi:10.1016/s0960-0760(01)00190-x (2002).

138 Ignatov, A., Ignatov, T., Roessner, A., Costa, S. D. & Kalinski, T. Role of GPR30 in the mechanisms of tamoxifen resistance in breast cancer MCF-7 cells. *Breast Cancer Res Treat* **123**, 87-96, doi:10.1007/s10549-009-0624-6 (2010).

139 Mo, Z. *et al.* GPR30 as an initiator of tamoxifen resistance in hormone-dependent breast cancer. *Breast Cancer Res* **15**, R114, doi:10.1186/bcr3581 (2013).

140 Ward, H. W. Anti-oestrogen therapy for breast cancer: a trial of tamoxifen at two dose levels. *Br Med J* **1**, 13-14, doi:10.1136/bmj.1.5844.13 (1973).

141 Jordan, V. C. Antiestrogens and selective estrogen receptor modulators as multifunctional medicines. 2. Clinical considerations and new agents. *J Med Chem* **46**, 1081-1111, doi:10.1021/jm020450x (2003).

142 Dixon, J. Endocrine resistance in breast cancer. *New Journal of Science* **2014** (2014).

143 Hiscox, S. *et al.* Tamoxifen resistance in breast cancer cells is accompanied by an enhanced motile and invasive phenotype: Inhibition by gefitinib (Iressa', ZD1839). *J Clinical experimental metastasis* **21**, 201-212 (2004).

144 Geiger, T. R. & Peeper, D. S. Metastasis mechanisms. *Biochim Biophys Acta* **1796**, 293-308, doi:10.1016/j.bbcan.2009.07.006 (2009).

145 Weaver, V. M., Fischer, A. H., Peterson, O. W. & Bissell, M. J. The importance of the microenvironment in breast cancer progression: recapitulation of mammary tumorigenesis using a unique human mammary epithelial cell model and a three-dimensional culture assay. *Biochem Cell Biol* **74**, 833-851, doi:10.1139/o96-089 (1996).

146 Hu, M. & Polyak, K. Microenvironmental regulation of cancer development. *Curr Opin Genet Dev* **18**, 27-34, doi:10.1016/j.gde.2007.12.006 (2008).

147 Khamis, Z. I., Sahab, Z. J. & Sang, Q. X. Active roles of tumor stroma in breast cancer metastasis. *Int J Breast Cancer* **2012**, 574025, doi:10.1155/2012/574025 (2012).

148 Cox, T. R. & Erler, J. T. Remodeling and homeostasis of the extracellular matrix: implications for fibrotic diseases and cancer. *Dis Model Mech* **4**, 165-178, doi:10.1242/dmm.004077 (2011).

149 Lu, P., Weaver, V. M. & Werb, Z. The extracellular matrix: a dynamic niche in cancer progression. *J Cell Biol* **196**, 395-406, doi:10.1083/jcb.201102147 (2012).

150 Gout, S. & Huot, J. Role of cancer microenvironment in metastasis: focus on colon cancer. *Cancer Microenviron* **1**, 69-83, doi:10.1007/s12307-008-0007-2 (2008).

151 De Duve, C., Pressman, B., Gianetto, R., Wattiaux, R. & Appelmans, F. Tissue fractionation studies. 6. Intracellular distribution patterns of enzymes in rat-liver tissue. *Biochemical Journal* **60**, 604 (1955).

152 Mohamed, M. M. & Sloane, B. F. Cysteine cathepsins: multifunctional enzymes in cancer. *Nat Rev Cancer* **6**, 764-775, doi:10.1038/nrc1949 (2006).

153 Lah, T. T., Obermajer, N., Alonso, M. B. D. & Kos, J. in *The Cancer Degradome* 587-625 (Springer, 2008).

154 Brömme, D. & Wilson, S. in *Extracellular matrix degradation* 23-51 (Springer, 2011).

155 Gocheva, V. & Joyce, J. A. Cysteine cathepsins and the cutting edge of cancer invasion. *Cell Cycle* **6**, 60-64, doi:10.4161/cc.6.1.3669 (2007).

156 Turk, V., Turk, B. & Turk, D. Lysosomal cysteine proteases: facts and opportunities. *Embo J* **20**, 4629-4633, doi:10.1093/emboj/20.17.4629 (2001).

157 Barrett, A. J. The cystatins: a diverse superfamily of cysteine peptidase inhibitors. *Biomed Biochim Acta* **45**, 1363-1374 (1986).

158 Hayashi, H., Tokuda, A. & Udaka, K. Biochemical study of cellular antigen-antibody reaction in tissue culture. I. Activation and release of a protease. *J Exp Med* **112**, 237-247, doi:10.1084/jem.112.2.237 (1960).

159 Tokuda, A., Hayashi, H. & Matsuba, K. Biochemical study of cellular antigen-antibody reaction in tissue culture. II. Release of a protease inhibitor. *J Exp Med* **112**, 249-255, doi:10.1084/jem.112.2.249 (1960).

160 Fossum, K. & Whitaker, J. R. Ficin and papain inhibitor from chicken egg white. *Arch Biochem Biophys* **125**, 367-375, doi:10.1016/0003-9861(68)90672-3 (1968).

161 Keilová, H. & Tomášek, V. Inhibition of cathepsin C by papain inhibitor from chicken egg white and by complex of this inhibitor with cathepsin B 1. *Collect Czech Chem Commun* **40**, 218-224 (1975).

162 Barrett, A. J. in *Methods in Enzymology* Vol. 80, 771-778 (Elsevier, 1981).

163 Abe, M., Abe, K., Kuroda, M. & Arai, S. Corn kernel cysteine proteinase inhibitor as a novel cystatin superfamily member of plant origin. Molecular cloning and expression studies. *Eur J Biochem* **209**, 933-937, doi:10.1111/j.1432-1033.1992.tb17365.x (1992).

164 Brzin, J., Kopitar, M., Turk, V. & Machleidt, W. Protein inhibitors of cysteine proteinases. I. Isolation and characterization of stefin, a cytosolic protein inhibitor of cysteine proteinases from human polymorphonuclear granulocytes. *Hoppe Seylers Z Physiol Chem* **364**, 1475-1480, doi:10.1515/bchm2.1983.364.2.1475 (1983).

165 Minakata, K. & Asano, M. New protein inhibitors of cysteine proteinases in human saliva and salivary glands. *Hoppe Seylers Z Physiol Chem* **365**, 399-403, doi:10.1515/bchm2.1984.365.1.399 (1984).

166 Lah, T. T., Kokalj-Kunovar, M. & Turk, V. Cysteine proteinase inhibitors in human cancerous tissues and fluids. *Biol Chem Hoppe Seyler* **371 Suppl**, 199-203 (1990).

167 Turk, V. in *International Symposium on Cysteine Proteinases and Their Inhibitors 1985: Portorož, Slovenia.* (De Gruyter).

168 Barrett, A. J. *et al.* Nomenclature and classification of the proteins homologous with the cysteine-proteinase inhibitor chicken cystatin. *Biochem J* **236**, 312, doi:10.1042/bj2360312 (1986).

169 Ochieng, J. & Chaudhuri, G. Cystatin superfamily. *J Health Care Poor Underserved* **21**, 51-70, doi:10.1353/hpu.0.0257 (2010).

170 Stubbs, M. T. *et al.* The refined 2.4 A X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. *Embo J* **9**, 1939-1947 (1990).

171 Poole, A. R., Tiltman, K. J., Recklies, A. D. & Stoker, T. A. Differences in secretion of the proteinase cathepsin B at the edges of human breast carcinomas and fibroadenomas. *Nature* **273**, 545-547, doi:10.1038/273545a0 (1978).

172 Koblinski, J. E., Ahram, M. & Sloane, B. F. Unraveling the role of proteases in cancer. *Clin Chim Acta* **291**, 113-135, doi:10.1016/s0009-8981(99)00224-7 (2000).

173 Jedeszko, C. & Sloane, B. F. Cysteine cathepsins in human cancer. *Biol Chem* **385**, 1017-1027, doi:10.1515/bc.2004.132 (2004).

174 Kos, J. *et al.* Cysteine proteinase inhibitors stefin A, stefin B, and cystatin C in sera from patients with colorectal cancer: relation to prognosis. *Clin Cancer Res* **6**, 505-511 (2000).

175 Sameni, M. *et al.* Cathepsin B and D are localized at the surface of human breast cancer cells. *Pathol Oncol Res* **1**, 43-53, doi:10.1007/bf02893583 (1995).

176 Tan, G. J., Peng, Z. K., Lu, J. P. & Tang, F. Q. Cathepsins mediate tumor metastasis. *World J Biol Chem* **4**, 91-101, doi:10.4331/wjbc.v4.i4.91 (2013).

177 Andl, C. D., McCowan, K. M., Allison, G. L. & Rustgi, A. K. Cathepsin B is the driving force of esophageal cell invasion in a fibroblast-dependent manner. *Neoplasia* **12**, 485-498, doi:10.1593/neo.10216 (2010).

178 Sinha, A. A., Gleason, D. F., Deleon, O. F., Wilson, M. J. & Sloane, B. F. Localization of a biotinylated cathepsin B oligonucleotide probe in human prostate including invasive cells and invasive edges by in situ hybridization. *Anat Rec* **235**, 233-240, doi:10.1002/ar.1092350207 (1993).

179 Sivaparvathi, M., Sawaya, R., Gokaslan, Z. L., Chintala, S. K. & Rao, J. S. Expression and the role of cathepsin H in human glioma progression and invasion. *Cancer Lett* **104**, 121-126, doi:10.1016/0304-3835(96)04242-5 (1996).

180 Kleer, C. G. *et al.* Epithelial and stromal cathepsin K and CXCL14 expression in breast tumor progression. *Clin Cancer Res* **14**, 5357-5367, doi:10.1158/1078-0432.Ccr-08-0732 (2008).

181 Zhang, Q. *et al.* Downregulation of cathepsin L suppresses cancer invasion and migration by inhibiting transforming growth factorbetamediated epithelialmesenchymal transition. *Oncol Rep* **33**, 1851-1859, doi:10.3892/or.2015.3754 (2015).

182 Pecar Fonovic, U. *et al.* Profilin 1 as a target for cathepsin X activity in tumor cells. *PLoS One* **8**, e53918, doi:10.1371/journal.pone.0053918 (2013).

183 Zhitomirsky, B. & Assaraf, Y. G. Lysosomal sequestration of hydrophobic weak base chemotherapeutics triggers lysosomal biogenesis and lysosome-dependent cancer multidrug resistance. *Oncotarget* **6**, 1143-1156, doi:10.18632/oncotarget.2732 (2015).

184 Sui, H., Shi, C., Yan, Z. & Wu, M. Overexpression of Cathepsin L is associated with chemoresistance and invasion of epithelial ovarian cancer. *Oncotarget* **7**, 45995-46001, doi:10.18632/oncotarget.10276 (2016).

185 Han, M. L. *et al.* Cathepsin L upregulation-induced EMT phenotype is associated with the acquisition of cisplatin or paclitaxel resistance in A549 cells. *Acta Pharmacol Sin* **37**, 1606-1622, doi:10.1038/aps.2016.93 (2016).

186 Zheng, X. *et al.* Cathepsin L inhibition suppresses drug resistance in vitro and in vivo: a putative mechanism. *Am J Physiol Cell Physiol* **296**, C65-74, doi:10.1152/ajpcell.00082.2008 (2009).

187 Zore, I. *et al.* Cathepsin B/cystatin C complex levels in sera from patients with lung and colorectal cancer. *Biol Chem* **382**, 805-810, doi:10.1515/bc.2001.097 (2001).

188 Kothapalli, R. *et al.* Constitutive expression of cytotoxic proteases and down-regulation of protease inhibitors in LGL leukemia. *Int J Oncol* **22**, 33-39 (2003).

189 Breznik, B., Mitrovic, A., T, T. L. & Kos, J. Cystatins in cancer progression: More than just cathepsin inhibitors. *Biochimie* **166**, 233-250, doi:10.1016/j.biochi.2019.05.002 (2019).

190 Lah, T. T. *et al.* Inhibitory properties of low molecular mass cysteine proteinase inhibitors from human sarcoma. *Biochim Biophys Acta* **993**, 63-73, doi:10.1016/0304-4165(89)90144-x (1989).

191 Wegiel, B. *et al.* Cystatin C is downregulated in prostate cancer and modulates invasion of prostate cancer cells via MAPK/Erk and androgen receptor pathways. *PLoS One* **4**, e7953, doi:10.1371/journal.pone.0007953 (2009).

192 Zhang, J. *et al.* Cystatin M: a novel candidate tumor suppressor gene for breast cancer. *Cancer Res* **64**, 6957-6964, doi:10.1158/0008-5472.Can-04-0819 (2004).

193 Zhang, J., Shi, Z., Huang, J. & Zou, X. CSTB downregulation promotes cell proliferation and migration and suppresses apoptosis in gastric cancer SGC-7901 cell line. *Oncol Res* **24**, 487-494, doi:10.3727/096504016x14685034103752 (2016).

194 Song, W., Zhao, C. & Jiang, R. Integrin-linked kinase silencing induces a S/G2/M phases cell cycle slowing and modulates metastasis-related genes in SGC7901 human gastric carcinoma cells. *Tumori* **99**, 249-256, doi:10.1700/1283.14200 (2013).

195 Yang, F. *et al.* Cystatin B inhibition of TRAIL-induced apoptosis is associated with the protection of FLIP(L) from degradation by the E3 ligase itch in human melanoma cells. *Cell Death Differ* **17**, 1354-1367, doi:10.1038/cdd.2010.29 (2010).

196 Ervin, H. & Cox, J. L. Late stage inhibition of hematogenous melanoma metastasis by cystatin C over-expression. *Cancer Cell Int* **5**, 14, doi:10.1186/1475-2867-5-14 (2005).

197 D'Costa, Z. C. *et al.* TBX2 represses CST6 resulting in uncontrolled legumain activity to sustain breast cancer proliferation: a novel cancer-selective target pathway with therapeutic opportunities. *Oncotarget* **5**, 1609-1620, doi:10.1186/1475-2867-5-14 (2014).

198 Leto, G., Crescimanno, M. & Flandina, C. On the role of cystatin C in cancer progression. *Life Sci* **202**, 152-160, doi:10.1016/j.lfs.2018.04.013 (2018).

199 Veena, M. S. *et al.* Inactivation of the cystatin E/M tumor suppressor gene in cervical cancer. *Genes Chromosomes Cancer* **47**, 740-754, doi:10.1002/gcc.20576 (2008).

200 Oh, B. M. *et al.* Cystatin SN inhibits auranofin-induced cell death by autophagic induction and ROS regulation via glutathione reductase activity in colorectal cancer. *Cell Death Dis* **8**, e3053, doi:10.1002/gcc.20576 (2017).

201 Jarvinen, M. Purification and properties of two protease inhibitors from rat skin inhibiting papain and other SH-proteases. *Acta Chem. Scand. B* **30** (1976).

202 Pol, E., Olsson, S. L., Estrada, S., Prasthofer, T. W. & Bjork, I. Characterization by spectroscopic, kinetic and equilibrium methods of the interaction between recombinant human cystatin A (stefin A) and cysteine proteinases. *Biochem J* **311(Pt 1)**, 275-282, doi:10.1042/bj3110275 (1995).

203 Leonardi, A., Turk, B. & Turk, V. Inhibition of bovine cathepsins L and S by stefins and cystatins. *Biol Chem Hoppe Seyler* **377**, 319-321 (1996).

204 Estrada, S. *et al.* The role of Gly-4 of human cystatin A (stefin A) in the binding of target proteinases. Characterization by kinetic and equilibrium methods of the interactions of cystatin A Gly-4 mutants with papain, cathepsin B, and cathepsin L. *Biochemistry* **37**, 7551-7560, doi:10.1021/bi980026r (1998).

205 Hsieh, W. T., Fong, D., Sloane, B. F., Golembieski, W. & Smith, D. I. Mapping of the gene for human cysteine proteinase inhibitor stefin A, STF1, to chromosome 3cen-q21. *Genomics* **9**, 207-209, doi:10.1016/0888-7543(91)90241-6 (1991).

206 Rinne, A., Jarvinen, M. & Rasanen, O. A protein reminiscent of the epidermal SH-protease inhibitor occurs in squamous epithelia of man and rat. *Acta Histochem* **63**, 183-192, doi:10.1016/s0065-1281(78)80024-5 (1978).

207 Rinne, A. *et al.* Demonstration of immunoreactive acid cysteine-proteinase inhibitor in reticulum cells of lymph node germinal centres. *Virchows Arch B Cell Pathol Incl Mol Pathol* **43**, 121-126, doi:10.1007/bf02932949 (1983).

208 Soderstrom, K. O., Rinne, R., Hopsu-Havu, V. K., Jarvinen, M. & Rinne, A. Identification of acid cysteine proteinase inhibitor (cystatin A) in the human thymus. *Anat Rec* **240**, 115-119, doi:10.1002/ar.1092400111 (1994).

209 Green, G. D., Kembhavi, A. A., Davies, M. E. & Barrett, A. J. Cystatin-like cysteine proteinase inhibitors from human liver. *Biochem J* **218**, 939-946, doi:10.1042/bj2180939 (1984).

210 Davies, M. E. & Barrett, A. J. Immunolocalization of human cystatins in neutrophils and lymphocytes. *Histochemistry* **80**, 373-377, doi:10.1007/bf00495420 (1984).

211 Soderstrom, K. O. *et al.* Expression of acid cysteine proteinase inhibitor (ACPI) in the normal human prostate, benign prostatic hyperplasia and adenocarcinoma. *Int J Cancer* **62**, 1-4, doi:10.1002/ijc.2910620102 (1995).

212 Scott, D. K., Lord, R., Muller, H. K., Malley, R. C. & Woods, G. M. Proteomics identifies enhanced expression of stefin A in neonatal murine skin compared with adults: functional implications. *Br J Dermatol* **156**, 1156-1162, doi:10.1111/j.1365-2133.2007.07875.x (2007).

213 Kato, T., Takai, T., Mitsuishi, K., Okumura, K. & Ogawa, H. Cystatin A inhibits IL-8 production by keratinocytes stimulated with Der p 1 and Der f 1: biochemical skin barrier against mite cysteine proteases. *J Allergy Clin Immunol* **116**, 169-176, doi:10.1016/j.jaci.2005.03.044 (2005).

214 Vasilopoulos, Y. *et al.* A nonsynonymous substitution of cystatin A, a cysteine protease inhibitor of house dust mite protease, leads to decreased mRNA stability and shows a significant association with atopic dermatitis. *Allergy* **62**, 514-519, doi:10.1111/j.1398-9995.2007.01350.x (2007).

215 Bowcock, A. M. *et al.* Insights into psoriasis and other inflammatory diseases from large-scale gene expression studies. *Hum Mol Genet* **10**, 1793-1805, doi:10.1093/hmg/10.17.1793 (2001).

216 Hawley-Nelson, P., Roop, D. R., Cheng, C. K., Krieg, T. M. & Yuspa, S. H. Molecular cloning of mouse epidermal cystatin A and detection of regulated expression in differentiation and tumorigenesis. *Mol Carcinog* **1**, 202-211, doi:10.1002/mc.2940010309 (1988).

217 Chwieralski, C. E., Welte, T. & Buhling, F. Cathepsin-regulated apoptosis. *Apoptosis* **11**, 143-149, doi:10.1007/s10495-006-3486-y (2006).

218 Peri, P. *et al.* The cysteine protease inhibitors cystatins inhibit herpes simplex virus type 1-induced apoptosis and virus yield in HEp-2 cells. *J Gen Virol* **88**, 2101-2105, doi:10.1099/vir.0.82990-0 (2007).

219 Takahashi, H. *et al.* Cystatin A suppresses ultraviolet B-induced apoptosis of keratinocytes. *J Dermatol Sci* **46**, 179-187, doi:10.1016/j.jdermsci.2007.02.003 (2007).

220 Dubin, G. Proteinaceous cysteine protease inhibitors. *Cell Mol Life Sci* **62**, 653-669, doi:10.1007/s00018-004-4445-9 (2005).

221 Bjorck, L., Grubb, A. & Kjellen, L. Cystatin C, a human proteinase inhibitor, blocks replication of herpes simplex virus. *J Virol* **64**, 941-943 (1990).

222 Bobek, L. A. & Levine, M. J. Cystatins--inhibitors of cysteine proteinases. *Crit Rev Oral Biol Med* **3**, 307-332, doi:10.1177/10454411920030040101 (1992).

223 Shah, A. & Bano, B. Cystatins in Health and Diseases. *Int J Pept Res Ther* **15**, 43, doi:10.1007/s10989-008-9160-1 (2009).

224 Turk, V., Turk, B., Guncar, G., Turk, D. & Kos, J. Lysosomal cathepsins: structure, role in antigen processing and presentation, and cancer. *Adv Enzyme Regul* **42**, 285-303, doi:10.1016/s0065-2571(01)00034-6 (2002).

225 Dohm, A. *et al.* Identification of CD37, cystatin A, and IL-23A gene expression in association with brain metastasis: analysis of a prospective trial. *Int J Biol Markers* **34**, 90-97, doi:10.1177/1724600818803104 (2019).

226 Lah, T. T. *et al.* Stefins and lysosomal cathepsins B, L and D in human breast carcinoma. *Int J Cancer* **50**, 36-44, doi:10.1002/ijc.2910500109 (1992).

227 Strojan, P. *et al.* Prognostic significance of cysteine proteinases cathepsins B and L and their endogenous inhibitors stefins A and B in patients with squamous cell carcinoma of the head and neck. *Clin Cancer Res* **6**, 1052-1062 (2000).

228 Butler, M. W. *et al.* Modulation of cystatin A expression in human airway epithelium related to genotype, smoking, COPD, and lung cancer. *Cancer Res* **71**, 2572-2581, doi:10.1158/0008-5472.Can-10-2046 (2011).

229 Gole, B. *et al.* The regulation of cysteine cathepsins and cystatins in human gliomas. *Int J Cancer* **131**, 1779-1789, doi:10.1002/ijc.27453 (2012).

230 Sinha, A. A. *et al.* Heterogeneity of cathepsin B and stefin A expression in Gleason pattern 3+3 (score 6) prostate cancer needle biopsies. *Anticancer Res* **27**, 1407-1413 (2007).

231 Takahashi, H., Kinouchi, M., Wuepper, K. D. & Iizuka, H. Cloning of human keratolinin cDNA: keratolinin is identical with a cysteine proteinase inhibitor, cystatin A, and is regulated by Ca2+, TPA, and cAMP. *Journal of investigative dermatology* **108**, 843-847 (1997).

232 Takahashi, H. *et al.* Expression of human cystatin A by keratinocytes is positively regulated via the Ras/MEKK1/MKK7/JNK signal transduction pathway but negatively regulated via the Ras/Raf-1/MEK1/ERK pathway. *J Biol Chem* **276**, 36632-36638, doi:10.1074/jbc.M102021200 (2001).

233 Takahashi, H. *et al.* Transcriptional factor AP-2gamma increases human cystatin A gene transcription of keratinocytes. *Biochem Biophys Res Commun* **278**, 719-723, doi:10.1006/bbrc.2000.3850 (2000).

234 Gupta, A. *et al.* Cell cycle- and cancer-associated gene networks activated by Dsg2: evidence of cystatin A deregulation and a potential role in cell-cell adhesion. *PLoS One* **10**, e0120091, doi:10.1371/journal.pone.0120091 (2015).

235 Luo, A. *et al.* Differentiation-associated genes regulated by c-Jun and decreased in the progression of esophageal squamous cell carcinoma. *PLoS One* **9**, e96610, doi:10.1371/journal.pone.0120091 (2014).

236 Sharma, S., Kelly, T. K. & Jones, P. A. Epigenetics in cancer. *Carcinogenesis* **31**, 27-36 (2010).

237 Chen, Q. W., Zhu, X. Y., Li, Y. Y. & Meng, Z. Q. Epigenetic regulation and cancer (review). *Oncol Rep* **31**, 523-532, doi:10.3892/or.2013.2913 (2014).

238 Kim, G. D., Ni, J., Kelesoglu, N., Roberts, R. J. & Pradhan, S. Co-operation and communication between the human maintenance and de novo DNA (cytosine-5) methyltransferases. *Embo J* **21**, 4183-4195, doi:10.1093/emboj/cdf401 (2002).

239 Larsen, F., Gundersen, G., Lopez, R. & Prydz, H. CpG islands as gene markers in the human genome. *Genomics* **13**, 1095-1107, doi:10.1016/0888-7543(92)90024-m (1992).

240 Tate, P. H. & Bird, A. P. Effects of DNA methylation on DNA-binding proteins and gene expression. *Curr Opin Genet Dev* **3**, 226-231, doi:10.1016/0959-437x(93)90027-m (1993).

241 Newell-Price, J., Clark, A. J. & King, P. DNA methylation and silencing of gene expression. *Trends Endocrinol Metab* **11**, 142-148 (2000).

242 Marques-Magalhães, Â., Graça, I., Henrique, R. & Jerónimo, C. Targeting DNA Methyltranferases in Urological Tumors. *Front Pharmacol* **9**, 366, doi:10.3389/fphar.2018.00366 (2018).

243 Feinberg, A. P. & Vogelstein, B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* **301**, 89-92, doi:10.1038/301089a0 (1983).

244 Esteller, M. Epigenetics in cancer. *N Engl J Med* **358**, 1148-1159, doi:10.1056/NEJMra072067 (2008).

245 Rivenbark, A. G., Jones, W. D. & Coleman, W. B. DNA methylation-dependent silencing of CST6 in human breast cancer cell lines. *Lab Invest* **86**, 1233-1242, doi:10.1038/labinvest.3700485 (2006).

246 Kim, T. Y., Zhong, S., Fields, C. R., Kim, J. H. & Robertson, K. D. Epigenomic profiling reveals novel and frequent targets of aberrant DNA methylation-mediated silencing in malignant glioma. *Cancer Res* **66**, 7490-7501, doi:10.1158/0008-5472.Can-05-4552 (2006).

247 Ma, Y., Chen, Y. & Petersen, I. Expression and epigenetic regulation of cystatin B in lung cancer and colorectal cancer. *Pathol Res Pract* **213**, 1568-1574, doi:10.1016/j.prp.2017.06.007 (2017).

248 Morgenstern, J. P. & Land, H. Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. *Nucleic Acids Res* **18**, 3587-3596 (1990).

249 Welshons, W. V., Wolf, M. F., Murphy, C. S. & Jordan, V. C. Estrogenic activity of phenol red. *Mol Cell Endocrinol* **57**, 169-178, doi:10.1016/0303-7207(88)90072-x (1988).

250 Brown, A. M., Jeltsch, J. M., Roberts, M. & Chambon, P. Activation of pS2 gene transcription is a primary response to estrogen in the human breast cancer cell line MCF-7. *Proc Natl Acad Sci U S A* **81**, 6344-6348, doi:10.1073/pnas.81.20.6344 (1984).

251 Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671-675 (2012).

252 Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-408, doi:10.1006/meth.2001.1262 (2001).

253 Karlsson, J. O., Ostwald, K., Kabjorn, C. & Andersson, M. A method for protein assay in Laemmli buffer. *Anal Biochem* **219**, 144-146, doi:10.1006/abio.1994.1243 (1994).

254 Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. Protein measurement with the Folin phenol reagent. *J Biol Chem* **193**, 265-275 (1951).

255 Shivaswamy, S. & Iyer, V. R. Genome-wide analysis of chromatin status using tiling microarrays. *Methods* **41**, 304-311, doi:10.1016/j.ymeth.2006.11.002 (2007).

256 Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* **44**, W3-W10, doi:10.1093/nar/gkw343 (2016).

257 Andrews, S. *FastQC A Quality Control tool for High Throughput Sequence Data*, <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2010).

258 Blankenberg, D. *et al.* Manipulation of FASTQ data with Galaxy. *Bioinformatics* **26**, 1783-1785, doi:10.1093/nar/gkw343 (2010).

259 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25, doi:10.1093/bioinformatics/btq281 (2009).

260 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1186/gb-2009-10-3-r25 (2009).

261 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1093/bioinformatics/btp352 (2008).

262 Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res* **12**, 996-1006, doi:10.1186/gb-2008-9-9-r137 (2002).

263 Gibson, D. G. Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol* **498**, 349-361, doi:10.1016/b978-0-12-385120-8.00015-2 (2011).

264 Gyorffy, B. *et al.* An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat* **123**, 725-731, doi:10.1007/s10549-009-0674-9 (2010).

265 Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61-70, doi:10.1038/nature11412 (2012).

266 Goldman, M., Craft, B., Brooks, A., Zhu, J. & Haussler, D. The UCSC Xena Platform for cancer genomics data visualization and interpretation. *BioRxiv*, 326470 (2018).

267 Hunter, K. W., Crawford, N. P. & Alsarraj, J. Mechanisms of metastasis. *Breast Cancer Res* **10 Suppl 1**, S2, doi:10.1186/bcr1988 (2008).

268 Kelly, T. *et al.* Proteolysis of extracellular matrix by invadopodia facilitates human breast cancer cell invasion and is mediated by matrix metalloproteinases. *Clin Exp Metastasis* **16**, 501-512, doi:10.1023/a:1006538200886 (1998).

269 Engbring, J. A. & Kleinman, H. K. The basement membrane matrix in malignancy. *J Pathol* **200**, 465-470, doi:10.1002/path.1396 (2003).

270 Strojan, P. *et al.* Stefin a and stefin B: markers for prognosis in operable squamous cell carcinoma of the head and neck. *Int J Radiat Oncol Biol Phys* **68**, 1335-1341, doi:10.1016/j.ijrobp.2007.02.004 (2007).

271 Shiba, D. *et al.* Clinicopathological significance of cystatin A expression in progression of esophageal squamous cell carcinoma. *Medicine (Baltimore)* **97**, e0357, doi:10.1097/md.0000000000010357 (2018).

272 Calkins, C. C. & Sloane, B. F. Mammalian cysteine protease inhibitors: biochemical properties and possible roles in tumor progression. *Biol Chem Hoppe Seyler* **376**, 71-80 (1995).

273 Ruffini, E., Weder, W., Filosso, P. & Girard, N. *IASLC Thoracic Oncology (Second Edition), 2018*. 569-589.e4 (2018).

274 Roy, D. M. & Walsh, L. A. Candidate prognostic markers in breast cancer: focus on extracellular proteases and their inhibitors. *Breast Cancer (Dove Med Press)* **6**, 81-91, doi:10.2147/bctt.S46020 (2014).

275 Cianfrocca, M. & Goldstein, L. J. Prognostic and predictive factors in early-stage breast cancer. *The oncologist* **9**, 606-616 (2004).

276 Wakefield, L. M. & Roberts, A. B. TGF-beta signaling: positive and negative effects on tumorigenesis. *Curr Opin Genet Dev* **12**, 22-29, doi:10.1016/s0959-437x(01)00259-3 (2002).

277 Chimge, N. O. & Frenkel, B. The RUNX family in breast cancer: relationships with estrogen signaling. *Oncogene* **32**, 2121-2130, doi:10.1038/onc.2012.328 (2013).

278 Browne, G. *et al.* Runx1 is associated with breast cancer progression in MMTV-PyMT transgenic mice and its depletion in vitro inhibits migration and invasion. *J Cell Physiol* **230**, 2522-2532, doi:10.1002/jcp.24989 (2015).

279 Mercado-Matos, J., Matthew-Onabanjo, A. N. & Shaw, L. M. RUNX1 and breast cancer. *Oncotarget* **8**, 36934-36935, doi:10.18632/oncotarget.17249 (2017).

280 Zhang, M. H., Man, H. T., Zhao, X. D., Dong, N. & Ma, S. L. Estrogen receptor-positive breast cancer molecular signatures and therapeutic potentials (Review). *Biomed Rep* **2**, 41-52, doi:10.3892/br.2013.187 (2014).

281 Yager, J. D. & Davidson, N. E. Estrogen carcinogenesis in breast cancer. *N Engl J Med* **354**, 270-282, doi:10.1056/NEJMra050776 (2006).

282 Deroo, B. J. & Korach, K. S. Estrogen receptors and human disease. *J Clin Invest* **116**, 561-570, doi:10.1172/jci27987 (2006).

283 Albini, A. *et al.* 17 beta-estradiol regulates and v-Ha-ras transfection constitutively enhances MCF7 breast cancer cell interactions with basement membrane. *Proc Natl Acad Sci U S A* **83**, 8182-8186, doi:10.1073/pnas.83.21.8182 (1986).

284 Thompson, E. W. *et al.* Differential regulation of growth and invasiveness of MCF-7 breast cancer cells by antiestrogens. *Cancer Res* **48**, 6764-6768 (1988).

285 Rochefort, H. *et al.* Estrogen receptor mediated inhibition of cancer cell invasion and motility: an overview. *J Steroid Biochem Mol Biol* **65**, 163-168, doi:10.1016/s0960-0760(98)00010-7 (1998).

286 Wingrove, C. S., Garr, E., Godsland, I. F. & Stevenson, J. C. 17beta-oestradiol enhances release of matrix metalloproteinase-2 from human vascular smooth muscle cells. *Biochim Biophys Acta* **1406**, 169-174, doi:10.1016/s0925-4439(97)00097-5 (1998).

287 Nilsson, U. W., Garvin, S. & Dabrosin, C. MMP-2 and MMP-9 activity is regulated by estradiol and tamoxifen in cultured human breast cancer cells. *Breast Cancer Res Treat* **102**, 253-261, doi:10.1007/s10549-006-9335-4 (2007).

288 Kousidou, O. *et al.* Estradiol-estrogen receptor: a key interplay of the expression of syndecan-2 and metalloproteinase-9 in breast cancer cells. *Mol Oncol* **2**, 223-232, doi:10.1016/j.molonc.2008.06.002 (2008).

289 Howlader, N. *et al.* SEER Cancer Statistics Review, 1975-2017, National Cancer Institute. Bethesda, MD, based on November 2019 SEER data submission, posted to the SEER web site, April 2020. (2019).

290 World Health Organization & Department of Information Evidence and Research. *Mortality database*, <https://www-dep.iarc.fr/WHOdb/WHOdb.htm> (2019).

291 Ferlay, J., Colombet, M. & Bray, F. *Cancer Incidence in Five Continents, CI5plus: IARC CancerBase No. 9 Lyon, France: International Agency for Research on Cancer*, <http://ci5.iarc.fr/> (2018).

292 Manjegowda, M. C., Deb, G., Kumar, N. & Limaye, A. M. Expression profiling of genes modulated by estrogen, EGCG or both in MCF-7 breast cancer cells. *Genomics data* **5**, 210-212, doi:10.1016/j.gdata.2015.05.040 (2015).

293 Sheeba, J. D. J. *et al.* Estrogen-regulated extracellular matrix remodeling genes in MCF-7 breast cancer cells. *Gene Reports* **3**, 14-21, doi:10.1016/j.genrep.2015.12.001 (2016).

294 Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W. W. & Lenhard, B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res* **32**, D91-94, doi:10.1093/nar/gkh012 (2004).

295 Kolar, Z., Jarvinen, M. & Negrini, R. Demonstration of proteinase inhibitors cystatin A, B and C in breast cancer and in cell lines MCF-7 and ZR-75-1. *Neoplasma* **36**, 185-189 (1989).

296 Palungwachira, P. *et al.* Immunohistochemical localization of cathepsin L and cystatin A in normal skin and skin tumors. *J Dermatol* **29**, 573-579, doi:10.1111/j.1346-8138.2002.tb00182.x (2002).

297 Tycko, B. Epigenetic gene silencing in cancer. *J Clin Invest* **105**, 401-407, doi:10.1172/jci9462 (2000).

298 Jones, P. A. & Baylin, S. B. The fundamental role of epigenetic events in cancer. *Nat Rev Genet* **3**, 415-428, doi:10.1038/nrg816 (2002).

299 Saied, M. H., Rady, A. S., El Naga, G. M. A. & Sharaki, O. A. Clinical Utility of promoter methylation of the tumor suppressor genes DKK3, and RASSF1A in breast cancer patients. *Egypt J Med Hum Genet* **19**, 87-90 (2018).

300 Manavathi, B. *et al.* Derailed estrogen signaling and breast cancer: an authentic couple. *Endocr Rev* **34**, 1-32, doi:10.1210/er.2011-1057 (2013).

301 Hervouet, E., Cartron, P. F., Jouvenot, M. & Delage-Mourroux, R. Epigenetic regulation of estrogen signaling in breast cancer. *Epigenetics* **8**, 237-245, doi:10.4161/epi.23790 (2013).

302 Putnik, M., Zhao, C., Gustafsson, J. A. & Dahlman-Wright, K. Global identification of genes regulated by estrogen signaling and demethylation in MCF-7 breast cancer cells. *Biochem Biophys Res Commun* **426**, 26-32, doi:10.1016/j.bbrc.2012.08.007 (2012).

303 Al-Dhaheri, M. H., Shah, Y. M., Basrur, V., Pind, S. & Rowan, B. G. Identification of novel proteins induced by estradiol, 4-hydroxytamoxifen and acolbifene in T47D breast cancer cells. *Steroids* **71**, 966-978, doi:10.1016/j.steroids.2006.07.006 (2006).

304 Wormke, M., Stoner, M., Saville, B. & Safe, S. Crosstalk between estrogen receptor alpha and the aryl hydrocarbon receptor in breast cancer cells involves unidirectional activation of proteasomes. *FEBS Lett* **478**, 109-112, doi:10.1016/s0014-5793(00)01830-5 (2000).

305 Pryzbylkowski, P., Obajimi, O. & Keen, J. C. Trichostatin A and 5 Aza-2' deoxycytidine decrease estrogen receptor mRNA stability in ER positive MCF7 cells through modulation of HuR. *Breast Cancer Res Treat* **111**, 15-25, doi:10.1007/s10549-007-9751-0 (2008).

306 Kazanets, A., Shorstova, T., Hilmi, K., Marques, M. & Witcher, M. Epigenetic silencing of tumor suppressor genes: Paradigms, puzzles, and potential. *Biochim Biophys Acta* **1865**, 275-288, doi:10.1016/j.bbcan.2016.04.001 (2016).

307 Ai, L. *et al.* Epigenetic silencing of the tumor suppressor cystatin M occurs during breast cancer progression. *Cancer Res* **66**, 7899-7909, doi:10.1158/0008-5472.Can-06-0576 (2006).

308 Rivenbark, A. G., Jones, W. D., Risher, J. D. & Coleman, W. B. DNA methylation-dependent epigenetic regulation of gene expression in MCF-7 breast cancer cells. *Epigenetics* **1**, 32-44, doi:10.4161/epi.1.1.2358 (2006).

309 Schagdarsurengin, U., Pfeifer, G. P. & Dammann, R. Frequent epigenetic inactivation of cystatin M in breast carcinoma. *Oncogene* **26**, 3089-3094, doi:10.1038/sj.onc.1210107 (2007).

310 Arapshian, A., Kuppumbatti, Y. S. & Mira-y-Lopez, R. Methylation of conserved CpG sites neighboring the beta retinoic acid response element may mediate retinoic acid receptor beta gene silencing in MCF-7 breast cancer cells. *Oncogene* **19**, 4066-4070, doi:10.1038/sj.onc.1203734 (2000).

311 Deng, G., Song, G. A., Pong, E., Sleisenger, M. & Kim, Y. S. Promoter methylation inhibits APC gene expression by causing changes in chromatin conformation and interfering with the binding of transcription factor CCAAT-binding factor. *Cancer Res* **64**, 2692-2698, doi:10.1158/0008-5472.can-03-3000 (2004).

312 Graff, J. R., Gabrielson, E., Fujii, H., Baylin, S. B. & Herman, J. G. Methylation patterns of the E-cadherin 5' CpG island are unstable and reflect the dynamic, heterogeneous loss of E-

cadherin expression during metastatic progression. *J Biol Chem* **275**, 2727-2732, doi:10.1074/jbc.275.4.2727 (2000).

313 Sathyanarayana, U. G. *et al.* Molecular detection of noninvasive and invasive bladder tumor tissues and exfoliated cells by aberrant promoter methylation of laminin-5 encoding genes. *Cancer Res* **64**, 1425-1430, doi:10.1158/0008-5472.can-03-0701 (2004).

314 Sathyanarayana, U. G. *et al.* Aberrant promoter methylation and silencing of laminin-5-encoding genes in breast carcinoma. *Clin Cancer Res* **9**, 6389-6394 (2003).

315 Sathyanarayana, U. G. *et al.* Aberrant promoter methylation of laminin-5-encoding genes in prostate cancers and its relationship to clinicopathological features. *Clin Cancer Res* **9**, 6395-6400 (2003).

316 Sathyanarayana, U. G. *et al.* Epigenetic inactivation of laminin-5-encoding genes in lung cancers. *Clin Cancer Res* **9**, 2665-2672 (2003).

317 Lapidus, R. G. *et al.* Mapping of ER gene CpG island methylation-specific polymerase chain reaction. *Cancer Res* **58**, 2515-2519 (1998).

318 Ottaviano, Y. L. *et al.* Methylation of the estrogen receptor gene CpG island marks loss of estrogen receptor expression in human breast cancer cells. *Cancer Res* **54**, 2552-2555 (1994).

319 Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T. & Henikoff, S. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet* **39**, 61-69, doi:10.1038/ng1929 (2007).

320 Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75-82, doi:10.1038/nature11232 (2012).

321 Comb, M. & Goodman, H. M. CpG methylation inhibits proenkephalin gene expression and binding of the transcription factor AP-2. *Nucleic Acids Res* **18**, 3975-3982, doi:10.1093/nar/18.13.3975 (1990).

322 Miranda, T. B. & Jones, P. A. DNA methylation: the nuts and bolts of repression. *J Cell Physiol* **213**, 384-390, doi:10.1002/jcp.21224 (2007).

323 Prendergast, G. C., Lawe, D. & Ziff, E. B. Association of Myn, the murine homolog of max, with c-Myc stimulates methylation-sensitive DNA binding and ras cotransformation. *Cell* **65**, 395-407, doi:10.1016/0092-8674(91)90457-a (1991).

324 Choy, M.-K. *et al.* Genome-wide conserved consensus transcription factor binding motifs are hyper-methylated. *BMC Genomics* **11**, 519 (2010).

325 Holler, M., Westin, G., Jiricny, J. & Schaffner, W. Sp1 transcription factor binds DNA and activates transcription even when the binding site is CpG methylated. *Genes Dev* **2**, 1127-1135, doi:10.1101/gad.2.9.1127 (1988).

326 Hendrich, B. & Bird, A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol* **18**, 6538-6547, doi:10.1128/mcb.18.11.6538 (1998).

327 Ung, M., Ma, X., Johnson, K. C., Christensen, B. C. & Cheng, C. Effect of estrogen receptor α binding on functional DNA methylation in breast cancer. *Epigenetics* **9**, 523-532 (2014).

328 Wittekind, C. & Neid, M. Cancer invasion and metastasis. *Oncology* **69 Suppl 1**, 14-16, doi:10.1159/000086626 (2005).

329 Iiizumi, M., Liu, W., Pai, S. K., Furuta, E. & Watabe, K. Drug development against metastasis-related genes and their pathways: a rationale for cancer therapy. *Biochim Biophys Acta* **1786**, 87-104, doi:10.1016/j.bbcan.2008.07.002 (2008).

330 Keppler, D. Towards novel anti-cancer strategies based on cystatin function. *Cancer Lett* **235**, 159-176, doi:10.1016/j.canlet.2005.04.001 (2006).

331 Sokol, J. P., Neil, J. R., Schiemann, B. J. & Schiemann, W. P. The use of cystatin C to inhibit epithelial-mesenchymal transition and morphological transformation stimulated by transforming growth factor-beta. *Breast Cancer Res* **7**, R844-853, doi:10.1186/bcr1312 (2005).

332 Alvarez-Diaz, S. *et al.* Cystatin D is a candidate tumor suppressor gene induced by vitamin D in human colon cancer cells. *J Clin Invest* **119**, 2343-2358, doi:10.1172/jci37205 (2009).

333 Dai, D. N. *et al.* Elevated expression of CST1 promotes breast cancer progression and predicts a poor prognosis. *J Mol Med (Berl)* **95**, 873-886, doi:10.1007/s00109-017-1537-1 (2017).

334 Angadi, P. & Kale, A. Epithelial-mesenchymal transition - A fundamental mechanism in cancer progression: An overview. *Indian J Health Sci Biomed Res* **8**, 77-84 (2015).

335 Sexton, P. S. & Cox, J. L. Inhibition of motility and invasion of B16 melanoma by the overexpression of cystatin C. *Melanoma Res* **7**, 97-101, doi:10.1097/00008390-199704000-00002 (1997).

336 Shridhar, R. *et al.* Cystatin M suppresses the malignant phenotype of human MDA-MB-435S cells. *Oncogene* **23**, 2206-2215, doi:10.1038/sj.onc.1207340 (2004).

337 Tian, M. & Schiemann, W. P. Preclinical efficacy of cystatin C to target the oncogenic activity of transforming growth factor Beta in breast cancer. *Transl Oncol* **2**, 174-183, doi:10.1593/tlo.09145 (2009).

338 Hunten, S. & Hermeking, H. p53 directly activates cystatin D/CST5 to mediate mesenchymal-epithelial transition: a possible link to tumor suppression by vitamin D3. *Oncotarget* **6**, 15842-15856, doi:10.18632/oncotarget.4683 (2015).

339 Ferrer-Mayorga, G. *et al.* Cystatin D locates in the nucleus at sites of active transcription and modulates gene and protein expression. *J Biol Chem* **290**, 26533-26548, doi:10.1074/jbc.M115.660175 (2015).

A

Appendix

Appendix I

List of primers used in this study.

Primer name	Sequence	Amplicon length (bp)	T_a (°C)	Purpose
CSTA-F	ATCTGAGGCCAAACCCGCC	275	60	Routine RT-PCR and qRT-PCR
CSTA-R	AGCCCGTCAGCTCGTCATC			
CycA-F	GGGCCGCGTCTCCTTTGAGC	158	60	Routine RT-PCR and qRT-PCR
CycA-R	GGCGTGTGAAGTCACCACCC			
ERα-F	GCCCTACTACCTGGAGAA	132	60	Routine RT-PCR
ERα-R	CCCTTGTCATTGGTACTGG			
pS2-F	AATGGCCACCATGGAGAACA	211	56	Routine RT-PCR
pS2-R	ATAGAAGCACCAGGGGACCC			
pS2-F	GGGTCCCCTGGTGCTTCTAT	141	60	qRT-PCR
pS2-R	CAGCCGAGCTCTGGGACTAA			
CSTA-Region 2-F	TCCTACTGGATCTCAGCCAC	369	60	ChIP
CSTA-Region 2-R	GCCCTGTTCTTAGAATAGTGC			
pS2-ChIP-F	CATTGCCTCCTCTCTGCTCC	423	60	ChIP
pS2-ChIP-R	ACTGTTGTCACGGCCAAGCC			
Region 1-F	TAATTTTGATATTATTAGTAAGTTTTGT	489	57	Bisulfite sequencing to amplify Region 1
Region 1-R	ATCAACAATCTCCTAAATTTCTAAAA			
Region 2-F	TTTTGTTATTAAATTTTTTTTATTGGATT	388	55	Bisulfite sequencing to amplify Region 2
Region 2-R	AAACCCTATTCTTAAAATAATACATAA			
CSTA-ORF-F	CACTCCTTCTCTAGGCGCCGGCCGGATCCGCCACCATGATACCTGGAGGCTTATCTGAGG	381	66	Cloning of CSTA ORF by Gibson assembly
CSTA-ORF-R	CTGACACACATTCCACAGGGTCGACCTACTTGTCGTCATCGTCTTTGTAGTCAAAGCCCGTCAGCTCGTCAT			

F and R indicate sense and anti-sense primers, respectively. T_a - Annealing temperature.

Appendix II

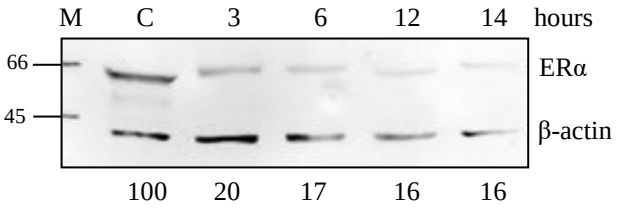

Figure A2.1. Western blotting analysis of ERα degradation in MCF-7 cells upon fulvestrant treatment. MCF-7 cells were treated with 100 nM fulvestrant for indicated periods of time. Total protein was subjected to western blotting analysis with antibodies specific for ERα and β-actin. The normalized integrated density of control was assigned the value of 100 and relative ERα levels were determined in fulvestrant-treated cells. Note that 80% reduction of ERα protein in MCF-7 cells treated with fulvestrant for 3 h with respect to control.

Appendix III

Quality assessment of custom generated polyclonal CSTA antibody.

Polyclonal CSTA antibody was custom generated and the specificity was tested by indirect ELISA. Among the three sets of immune sera collected from two rabbits (A and B), first (Bleed-1) and third set (Bleed-3) was used for indirect ELISA. Third set of immune sera of rabbit A showed high reactivity. Pre-immune sera were used as negative control. After affinity purification, purified antibody was validated by ELISA along with the immune sera and pre-immune sera and the reactivity was slightly increased after affinity purification.

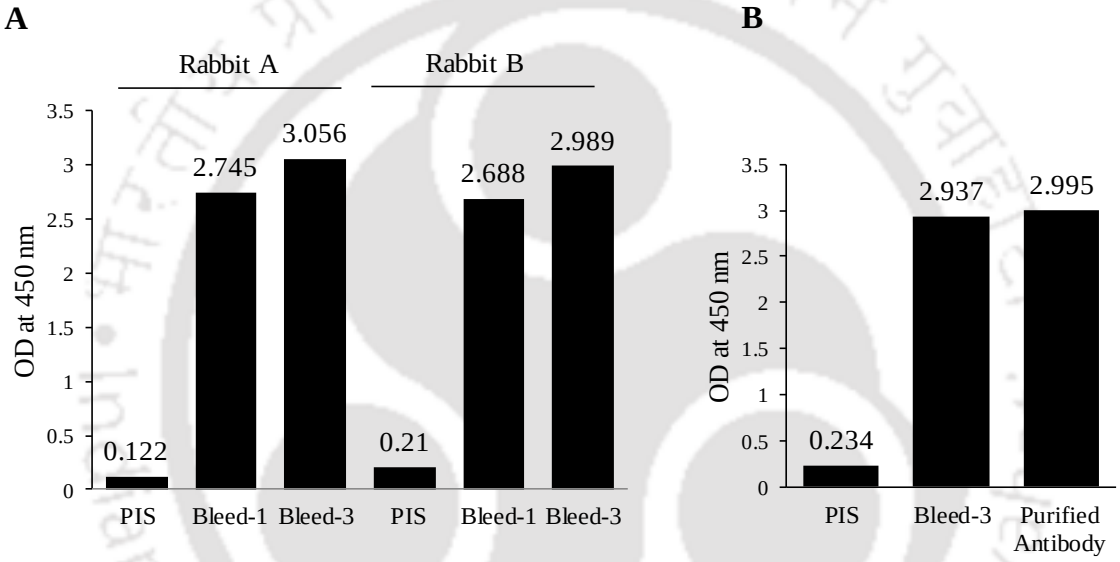

Figure A3.1 Indirect ELISA for testing the reactivity of hyper-immune sera and affinity-purified polyclonal CSTA antibody. A. Hyper-immune sera collected from first and third bleeds of two rabbits (A and B) were compared with their respective pre-immune sera (PIS) by indirect ELISA as described in the section 3.5 of chapter 3. B. Bleed-3 of rabbit A was affinity-purified using CSTA peptide and the reactivity was tested by indirect ELISA.

Validation of custom generated polyclonal CSTA antibody

The specificity of custom generated polyclonal CSTA antibody was validated by western blotting analysis using total protein from MCF-7 and MDA-MB-231 cells. At RNA level, MCF-7 cells express CSTA but not MDA-MB-231. Western blotting analysis showed a band of around 15 kDa when probed with immune sera, but not in pre-immune sera. Moreover, it was observed only in MCF-7, not in MDA-MB-231, which falls in line with the mRNA expression data. In addition to that, few non-specific bands were observed in both immune sera and pre-immune sera. Therefore, immune sera were affinity-purified in sulfo-link matrix using the peptide used for antibody generation.

Western blotting analysis with affinity-purified antibody showed a prominent 15 kDa band. Further to confirm the specificity, before probing the blot with affinity-purified antibody, it was pre-blocked with peptide. Antibody with peptide re-suspension buffer was used as positive control. When the total protein of MCF-7 cells was probed with peptide blocked-CSTA antibody, detection of 15 kDa band was blocked, which validates the specificity of the observed band. The specificity of the antibody was assessed with purified CSTA protein. Immune sera detected the purified CSTA protein but not pre-immune sera. Moreover, the reactivity was enhanced when it was probed with affinity-purified CSTA antibody. When the antibody was pre-blocked by CSTA peptide, purified CSTA protein was not detected. This further confirms the specificity of custom generated polyclonal antibody for the detection of CSTA protein.

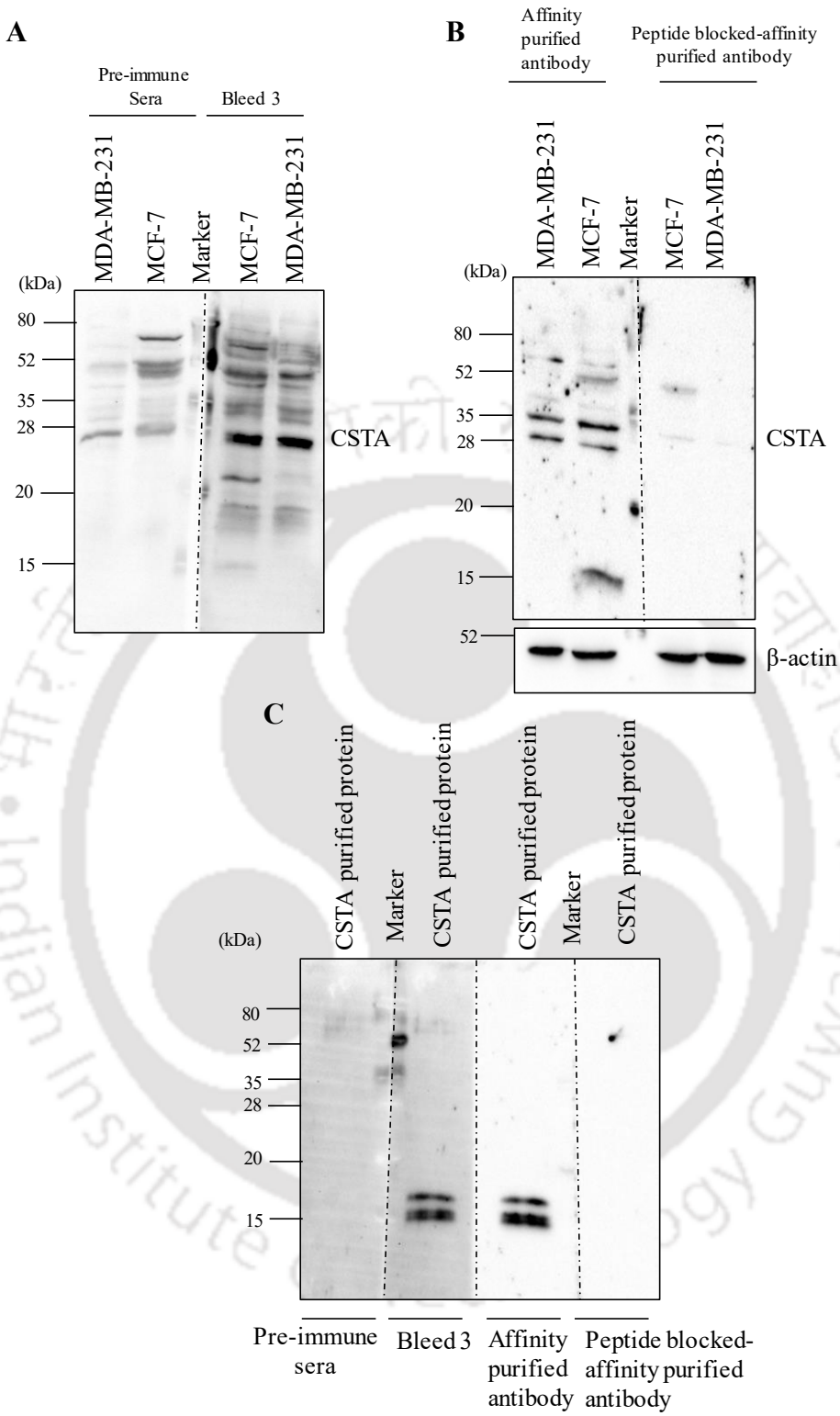

Figure A3.2. Validation of custom generated polyclonal CSTA antibody by western blotting. Total protein of MCF-7 and MDA-MB-231 (A &B) and CSTA purified protein (C) were probed with pre-immune sera, immune sera (Bleed 3) of rabbit A, affinity-purified antibody or peptide blocked-affinity-purified antibody. β-actin was used as an internal control.

Appendix IV

List of publications

1. **Dixcy Jaba Sheeba John Mary,** Girija Sikarwar, Ajay Kumar, Anil Mukund Limaye, Interplay of ERα binding and DNA methylation in the intron-2 determines the expression and estrogen regulation of cystatin A in breast cancer cells, Molecular and Cellular Endocrinology (2020) 504:110701 **(Portion of chapter 6 is published in this article).**

2. Ninad V. Puranik, Pratibha Srivastava, Gaurav Bhatt, **Dixcy Jaba Sheeba John Mary**, Anil M. Limaye & Jayanthi Sivaraman, Determination and analysis of agonist and antagonist potential of naturally occurring flavonoids for estrogen receptor (ERα) by various parameters and molecular modelling approach, Scientific reports (2019) 9(1), 7450.

3. Ajay Kumar, Mohan C Manjegowda, **Dixcy Jaba Sheeba John Mary**, Uttariya Pal, Sachin Kumar & Anil Mukund Limaye, Estrogen receptor-α is a determinant of protocadherin-8 expression in breast cancer cells, Gene Reports (2019) 14, 6-11.

4. **Dixcy Jaba Sheeba John Mary**, Mohan C Manjegowda, Ajay Kumar, Sarbajeet Dutta, Anil Mukund Limaye, The role of cystatin A in breast cancer and its functional link with ERα, Journal of Genetics and Genomics (2017) 44(12):593-597 **(Portions of chapter 4 and 5 are published in this article).**

5. **Dixcy Jaba Sheeba John Mary,** Mohan C Manjegowda, Hussain Marine, Gauri Deb, Neeraj Kumar, Anil Mukund Limaye, Estrogen-regulated extracellular matrix remodeling genes in MCF-7 breast cancer cells, Gene Reports (2016) 3, 14–21.

Appendix V

List of presentations

Platform presentations

1. **Dixcy Jaba Sheeba John Mary,** Girija Sikarwar, Mohan C Manjegowda, Ajay Kumar, Anil Mukund Limaye, Regulation of cystatin A by estrogen and DNA methylation in breast cancer cells in the annual conference of Indian Association of Cancer Research (IACR)-2020 (February 5th-7th, 2020) at Kovalam, Trivandrum (OP-5).

Poster presentations

1. **Dixcy Jaba Sheeba J M**, Mohan C Manjegowda, Ajay Kumar and Anil M Limaye, On the role of cystatin A in breast cancer and its functional link with ERα in International Congress of Cell Biology (January 27th-31st, 2018) organized by CCMB, Hyderabad. P342, Page no.-279
2. **Dixcy Jaba Sheeba J M,** Mohan C Manjegowda, Ajay Kumar and Anil M. Limaye, Study of regulation of steady-state mRNA levels of Cystatin A by estrogen in breast cancer cells in Research Conclave (March 18th-20th, 2016) at IIT Guwahati, Assam.
3. **Dixcy Jaba Sheeba J M**, Mohan C Manjegowda, Ajay Kumar and Anil M. Limaye, Regulation of Cystatin A by estrogen in breast cancer cells in International Conference of Cancer Research: New Horizons 2015 (November 19th-21th, 2015) at NCCS, Pune, P-23; Page no.-77.
4. **Dixcy Jaba Sheeba J M,** Mohan C Manjegowda, Marine Hussain, Gauri Deb, Neeraj Kumar, Anil M Limaye, Estrogen regulation of extracellular matrix remodeling genes in MCF-7 breast cancer cells in national Conference on Recent Advances in Cancer Biology and Therapeutics - 2014 (December 5th-7th, 2014) at IIT Guwahati, Assam.

Appendix VI

Copyrights and permission