Interplay of ER α binding and DNA methylation in the intron-2 determines the expression and estrogen regulation of Cystatin A in breast cancer cells.

Dixcy Jaba Sheeba John Mary, Girija Sikarwar, Ajay Kumar, Anil Mukund Limaye *

Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India

*Corresponding author

Anil Mukund Limaye

Associate Professor

Department of Biosciences and Bioengineering

Indian Institute of Technology Guwahati

Guwahati 781039, Assam, India

Tel.: +91-361-2582218

Fax: +91-361-2582249

E-mail address: amul@iitg.ac.in

Abstract

Despite advances in early detection and treatment, invasion and metastasis of breast tumors remains a major hurdle. Cystatin A (CSTA, also called stefin A), an estrogen-regulated gene in breast cancer cells, is an inhibitor of cysteine cathepsins, and a purported tumor suppressor. Loss of CSTA expression in breast tumors evidently shifts the balance in favor of cysteine cathepsins, thereby promoting extracellular matrix remodeling, tumor invasion and metastasis. However, the underlying mechanism behind the loss of CSTA expression in breast tumors is not known. Here, we have analyzed CSTA expression, and methylation of upstream and intron-2 CpG sites within the CSTA locus in human breast cancer cell lines and breast tumors of the TCGA cohort. Results showed an inverse relationship between expression and methylation. Sequence analysis revealed a potential estrogen response element (ERE) in the intron-2. Analysis of ChIP-seq data (ERP000380) and our own ChIP experiments showed that 17 β -estradiol (E2) enhanced ER α binding to this ERE in MCF-7 cells. This ERE was located amidst the differentially methylated intron-2 CpG sites, which provoked us to examine the possible conflict between estrogen-regulation of CSTA and DNA methylation in the intron-2. We analyzed the expression of CSTA and its regulation by estrogen in MDA-MB-231 and T47D cells subjected to global demethylation by 5-azacytidine (5-aza). 5-aza, not only enhanced CSTA expression in these cell lines but also restored estrogen-regulation of CSTA in these cells. Taken together,

our results indicate that DNA methylation- dependent silencing could play a significant role in the loss of CSTA expression in breast tumors. The potential of DNA methylation as an indicator of CSTA expression or as a marker of tumor progression can be explored in future investigations. Furthermore, our results indicate the convergence of ER α -mediated estrogen regulation and DNA methylation in the intron-2, thereby offering a novel context to understand the role of estrogen-ER α signaling axis in breast tumor invasion and metastasis.

Keywords: Cystatin A, Stefin A, Estrogen, Methylation, Bisulfite sequencing, Intron-2, Chromatin immunoprecipitation, TCGA

1. Introduction

Cysteine cathepsins are a major group of 11 lysosomal acid hydrolases in humans. The presence of cysteine in their active site distinguishes them from other lysosomal serine- or aspartic cathepsins (Khaket et al., 2019; Turk et al., 2012). Collectively, they hydrolyze proteins within and outside the cell (Turk et al., 2000). They are involved in various cellular processes, such as protein turnover, antigen presentation, apoptosis, and matrix remodeling (Fonović and Turk, 2014; Turk et al., 2000, 2012; Vizovišek et al., 2019). Overt or mislocalized expression of cysteine cathepsins is associated with diverse pathological conditions, including tumors and metastasis (Palermo and Joyce, 2008; Vidak et al., 2019). Cysteine cathepsins are regulated by compartmentalization, proteolytic cleavage (zymogen activation), pH and endogenous inhibitory proteins called cystatins (Turk et al., 2000, 2001; Verma et al., 2016). Cathepsins and cystatins are aberrantly expressed in a wide variety of tumors, including those of the breast (Pogorzelska et al., 2018). 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 (Kolwijck et al., 2010; Paraoan et al., 2009; Yano et al., 2001).

Cystatin A (CSTA, also known as Stefin A), is a member of the Stefin group within the Cystatin superfamily (Keppler, 2006). It is an 11 kDa protein. It inhibits the activity of cathepsins -B, -H and -L (Rivenbark and Coleman, 2009), which are implicated in tumor development, progression and survival (Pogorzelska et al., 2018). 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 (Duivenvoorden et al., 2017; Jones et al., 2004). Estrogen reduces CSTA expression in MCF-7 breast cancer cells, via a mechanism that depends on the estrogen receptor α (ER α). Tamoxifen (a selective estrogen receptor modulator), fulvestrant (a selective estrogen receptor degrader), and ER α knockdown increase the expression of CSTA in MCF-7

breast cancer cells (John Mary et al., 2017). However, the precise role of ER α in the expression and estrogen-mediated regulation of CSTA in breast cancer cells is not understood.

The relationship between CSTA expression, and breast cancer progression and prognosis is confusing. 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 (Kuopio et al., 1998). 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 (Lah et al., 1992). 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. However, they did not find a significant association between CSTA and breast cancer prognostic factors (Levicar et al., 2002). 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 (Parker et al., 2008). Recently we analyzed the expression of CSTA in breast tumors from the TCGA cohort and reported significantly higher expression of CSTA mRNA in normal breast tissues compared with the primary breast tumors. High CSTA expression was associated with reduced overall (OS), relapse-free (RFS), and distant metastasis-free survival (DMFS). However, survival analysis of each subtype of breast tumors yielded interesting results. CSTA expression did not have any effect on survival in HER2+ and basal tumors. CSTA expression correlated with reduced OS and RFS in luminal A, whereas it correlated with prolonged OS and DMFS in luminal B (John Mary et al., 2017). 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 (Parker et al., 2008). Using the same model, Withana et al showed that knockdown or selective inhibition of Cathepsin B can also reduce bone and lung metastasis (Withana et al., 2012). Taken together, these studies suggest that CSTA inhibits tumor invasion and metastasis by inhibiting Cathepsin B. Moreover, its tumor suppressor role is manifested in the tumor microenvironment. Loss of CSTA expression in primary tumors appears to be an important driving force towards metastatic progression. However, the mechanisms that lead to the loss of CSTA expression in breast tumors are not known.

Here, we have examined the relationship between CSTA expression and DNA methylation in specific regions within the CSTA locus in breast cancer cell lines, and breast tumors of the TCGA cohort. We show that CSTA expression has an inverse relationship with DNA methylation, suggesting that DNA methylation-dependent silencing could, at least in part, be responsible for the loss of CSTA expression in breast tumors. Furthermore, we found that estrogen regulation via ER α , and DNA methylation-dependent silencing converge in the intron-2 of CSTA.

2. Materials and methods

2.1. Plastic ware, chemicals and reagents

Cell culture plastic wares were purchased from Eppendorf (Hamburg, Germany). Cell culture media, fetal bovine serum (FBS), PowerUp SYBR Green PCR Master Mix, and High-Capacity cDNA Reverse Transcription Kit were from Invitrogen (CA, USA). Trypsin, penicillin, streptomycin, and charcoal-stripped FBS (csFBS) were from HiMedia (Mumbai, India). pMD20 vector was purchased from Clontech (CA, USA). Monoclonal β -Actin (AM4302) antibody and EpiJET Bisulfite Conversion Kit were purchased from Thermo Scientific (PA, USA). Monoclonal CSTA antibody (ab10442) was from Abcam (Cambridge, UK) and polyclonal ER α antibody (sc-543) was purchased from Santa Cruz Biotechnology (CA, USA). Progesterone Receptor (PR) antibody (8757S) was from Cell Signaling Technology (Massachusetts, USA). 5-azacytidine (5-aza) was procured from MP Biomedicals (Solon, USA). 17 β -estradiol (E2) were purchased from Sigma Aldrich (MO, USA). Routine laboratory buffers, solvents and salts were from Merck (Mumbai, India) or SRL (Mumbai, India).

2.2. Cell lines and cell culture

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 Rosewell Park Memorial Institute-1640 medium (RPMI-1640) with phenol red. These media for routine culture were supplemented with 10% (v/v) heat-inactivated FBS, 100 units/ml penicillin and 100 μ g/ml streptomycin (M1 medium) in a humidified 37 °C incubator with 5% CO₂. 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 μ g/ml streptomycin (M2 medium) were used.

2.3. Treatment protocols

2.3.1. Global demethylation using 5-aza

Cells were seeded in M1 medium. After 24 h, the cells were treated with fresh M1 medium containing 10 μ M 5-aza for 5 days. Fresh M1 medium with 5-aza was replenished every 48 h. The cells treated with DMSO (vehicle) served as control.

2.3.2. 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. After 24 h the cells were harvested for isolation of total RNA or protein.

2.3.3. 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.

2.4. Routine reverse transcription-polymerase chain reaction (RT-PCR) and quantitative RT-PCR (qRT-PCR)

Total RNA was isolated, reverse transcribed and the resultant cDNA was subjected to routine RT-PCR and qRT-PCR analyses with gene-specific primers (Supplementary table 1) for CSTA, ER α and cyclophilin A (CycA) as previously described (John Mary et al., 2017). CycA served as an internal control

2.5. Western blotting

Total protein was isolated from breast cancer cells with Laemmli sample buffer (Laemmli, 1970) or from organic phase of TRIzol lysates and then quantified by TCA method (Karlsson et al., 1994) or Lowry's method (Lowry et al., 1951) respectively. Protein samples (30 μ g) were resolved by 12% PAGE, transferred to 0.22 μ nitrocellulose membrane, and blocked with 1% gelatin in TBST for 2 h. Blots were probed with CSTA, ER α , PR, β -actin, or histone H3 antibodies, overnight, in TBST containing 0.1% gelatin. The blots were washed with TBST (6 $\times$ 5 min). Blots were then probed with anti-rabbit HRP-conjugated secondary antibody for 1 h, washed with TBST (6 $\times$ 5 min) and developed using Clarity Western ECL Substrate (Bio-Rad, California, US). Images were captured with ChemiDoc XRS+ system (Bio-Rad, California, US).

2.6. TCGA-BRCA data analysis

Methylation (generated with Illumina Infinium® Human Methylation 450K BeadChip array) and CSTA expression (RNAseq) data were retrieved from TCGA database (Koboldt et al., 2012) using the UCSC Xena Browser (Goldman et al., 2019). 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 ($\log_2\text{RPKM}+1$) 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 analysed by Welch two-sample *t*-test.

2.7. 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. 50 ng of converted gDNA was used for PCR with specific primers (Supplementary table 1) that amplified Region 1 and Region 2, 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, we ensured that 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 -13 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.

2.8. ChIP

Control and E2-treated MCF-7 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 pre-cleared with Protein G plus-Agarose beads precoated with 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 ER α antibody or rabbit IgG antibody for 2 h. Samples were immuno-precipitated by incubating with 20 μl of coated beads for 2 h, followed by centrifugation. Immunoprecipitates were washed with a series of wash buffers (Shivaswamy and Iyer, 2007) and eluted in 300 μl elution buffer containing proteinase-K for 2 h at 55°C, and overnight incubation at 65°C. Chromatin was column purified and ER α occupancy was assessed by PCR with primers specific to pS2 (positive control) or Region 2 of CSTA locus (Supplementary table 1).

2.9. *ChIP-Seq analysis*

Raw data of Chip-Seq experiments were retrieved from Sequence Read Archival (SRA) and analyzed using Galaxy, web-based platform (Afgan et al., 2016). 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 ER α was chosen for this study. IgG control ChIP-seq data was used as the negative control. Using FASTQC tool (Andrews, 2010), the quality of input reads were assessed. After converting the quality score to sanger quality type by FASTQ Groomer (Blankenberg et al., 2010), reads were mapped to reference human genome (hg19) using “Map with Bowtie for Illumina” tool (Langmead et al., 2009). Unmapped reads were discarded by “Filter SAM or BAM, output SAM or BAM” tool (Li et al., 2009). Genomic regions with enriched sequencing reads were identified by MACS (Model-based analysis of ChIP-Seq) tool (Zhang et al., 2008). Resultant Wig files were converted to bigWig files using “Wig/BedGraph-to-bigWig” tool and the peaks representing ER α occupancy were visualized using UCSC genome browser (Kent et al., 2002).

2.10. *Statistical analysis*

Bisulfite sequencing data were analyzed for statistical difference in proportions of methylated CpGs for every pair of cell lines. The total number of CpGs sampled for each cell lines was greater than 30. We assumed that methylation at each CpG site was independent of the adjacent CpGs. Under 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

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

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 as described in section 2.6. All other data were analyzed by ANOVA followed by Tukey’s HSD, or Welch two-sample t -test using R-statistical package.

3. Results

3.1. Differential expression of CSTA in breast cancer cell lines

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

3.2. 5-aza induces CSTA expression in MDA-MB-231 cells

Treatment with 5-aza, a DNA methyltransferase 1 (DNMT1) inhibitor, causes genome-wide demethylation. We tested the effect of 5-aza on CSTA expression in MDA-MB-231 cells. As shown in Fig. 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.

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

CSTA does not have CpG islands (Rivenbark and Coleman, 2009). Methylation data obtained from 450K bead chip arrays (Encode project) was analyzed to ascertain the status of methylation in five CpGs in the CSTA locus. The locations of these CpGs is indicated by the colored vertical lines in Fig. 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 (Fig. 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.

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

We isolated gDNA from the panel of breast cancer cells indicated in Fig. 1. They were subjected to bisulfite sequencing analysis of Region 1 and Region 2 (Fig. 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. 84 CpGs were interrogated across 12 independent TA clones in the Region 1. 39 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 Fig. 3B and 3C, respectively. We determined the proportion of methylated CpGs in each region in the cell lines. In the 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 the 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 (Fig. 3B and 3C). Generally, methylation in both regions appeared to be inversely related to CSTA expression in the breast cancer cell lines.

3.5. Methylation of the upstream CpGs correlates inversely with CSTA expression in the breast tumors

CSTA expression (RNASeq; $\log_2(\text{RPKM}+1)$) 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 (Goldman et al., 2019). Methylation data were available for 4 probes in the CSTA locus. These probes correspond to the four CpG sites in the Region 1 shown in Fig. 3A. We first performed a probe-wise analysis of the correlation between methylation (beta value) and CSTA expression in primary tumors. Methylation at each of the 4 CpG sites inversely correlated with CSTA expression (Table 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 (Fig. 4A-D). We generated a composite methylation score for each sample by averaging the beta values for all the probes. The composite methylation score correlated inversely with CSTA expression ($\rho = -0.582$, $p < 0.0001$, Fig. 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 (Fig. 4F).

3.6. Estrogen enhances ER α occupancy in the intron-2 region of CSTA in MCF-7 cells

Analysis of the CSTA locus using JASPAR (Sandelin et al., 2004) revealed a potential estrogen response element (ERE) in the intron-2 region. This ERE was located amidst the CpG sites analyzed in Region 2 (Fig. 5A). Chip-Seq data of chromatin samples from vehicle-, E2- or tamoxifen-treated MCF-7 cells, which were immunoprecipitated with ER α -specific antibody, were analyzed using Galaxy. The results were viewed in the UCSC genome browser. A robust peak of ER α occupancy was observed in the intron-2 region of CSTA in E2-treated MCF-7 cells (indicated by the red rectangle, Fig. 5B). This peak was diminished or negligible in tamoxifen or vehicle treated MCF-7 cells. Notably, the peak of ER α binding overlapped with ERE predicted by JASPAR. To validate these observations, we performed ChIP experiments on ethanol (vehicle) or E2-treated MCF-7 cells using an ER α specific antibody. Enrichment of the ERE containing sequence in the pS2 locus (positive control) following E2 treatment validated the ChIP protocol. As shown in Fig. 5C, the Region-2 was enriched in immunoprecipitated chromatin samples from E2-treated MCF-7 cells.

3.7. Global demethylation restores estrogen regulation of CSTA in MDA-MB231 and T47D cells

In MDA-MB-231 cells, DNA methylation-dependent silencing is the underlying mechanism for the loss of ER α (Lapidus et al., 1998) and CSTA (this study) expression. We examined ER α and CSTA expression in 5-aza-untreated or -pretreated MDA-MB-231 cells, which were stimulated with vehicle or 10 nM E2. As expected, in 5-aza-untreated cells, ER α protein expression was not detectable after vehicle or E2 treatment (Fig. 6A, lanes 1 and 2). There was no significant difference in CSTA mRNA expression (Fig. 6C, 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 ER α specific antibody (Fig. 6A, lanes 3 and 4). This immunoreactive protein had a higher molecular mass compared to the expected 66 kDa for ER α . Notwithstanding this discrepancy, induction of PR, and further enhancement of its expression with E2 confirmed the generation of a functional ER α in 5-aza pretreated cells (Fig. 6A, B, lanes 3 and 4). 5-aza significantly induced CSTA mRNA expression in MDA-MB-231 cells (Fig. 6C, 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 pre-treated cells with and without E2 treatment were significantly different when analyzed by the Welch two-sample *t*-test (Fig. 6C, bars 3 and 4, $p = 0.0098$). Western blots failed to demonstrate CSTA protein in MDA-MB-231 cells.

We applied the same experimental design to ER α positive T47D cells, which express a very low or undetectable level of CSTA. Without 5-aza pretreatment, T47D cells treated with E2 showed decreased levels of ER α protein and increased levels of PR, as expected (Fig. 6D, E, lanes 1 and 2) (Al-Dhaheri et al., 2006; Wormke et al., 2000). There was no observable effect on CSTA protein (Fig. 6D, lanes 1 and 2). However, we observed an increase in CSTA mRNA (Fig. 6F, bars 1 and 2; ANOVA followed by Tukey's HSD), although the increase was not statistically significant. 5-aza pre-treatment alone caused a decrease in ER α protein in T47D cells (Fig. 6D, lanes 1 and 3) in a manner similar to that reported in MCF-7 cells (Pryzbylowski et al., 2008). This also led to increased CSTA (Fig. 6D, lanes 1 and 3) and decreased PR protein expression (Fig. 6E, lanes 1 and 3). E2 induction of PR in 5-aza pre-treated T47D cells showed that ER α was functional (Fig. 6E, lanes 3 and 4). E2 not only enhanced CSTA protein expression (Fig. 6D, lanes 3 and 4), but also significantly enhanced CSTA mRNA in 5-aza pre-treated cells (Fig. 6F, bars 1, 3 and 4; ANOVA followed by Tukey's HSD).

4. Discussion

There were two motivating reasons behind our study of the relationship between CSTA gene expression and methylation in breast cancer. Firstly, CSTA is a proposed tumor suppressor (Calkins and Sloane, 1995; Duivenvoorden et al., 2017), and methylation-dependent silencing of tumor suppressors are well known (Kazanets et al., 2016). 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 (Ma et al., 2018). 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 (Ai et al., 2006; Rivenbark et al., 2006b, 2006a; Schagdarsurengin et al., 2007). Our analysis 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. There are examples of DNA methylation-dependent regulation of CpG island-less genes (Arapshian et al., 2000; Deng et al., 2004; Graff et al., 2000; Sathyanarayana et al., 2004, 2003a, 2003b, 2003c). Hence, our results are not surprising.

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 bead chip 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 (Fig. 4E). The ENCODE project data corresponding to MCF-7 and T47D cells, and our bisulfite sequencing results clearly demonstrated that CSTA expression, and methylation in the intron-2 CpGs, are inversely correlated. Collectively, these data provide compelling evidence 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 (Lah et al., 1997, 1992; Levicar et al., 2002). Ablation or inhibition of Cathepsin B also inhibits spine and lung metastasis in the animal model (Withana et al., 2012). We therefore, propose that DNA methylation-mediated silencing of CSTA in primary breast tumors tips the CTSE/CSTA balance in favor of CTSE, 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.

Our previous analysis of CSTA and ER α mRNA expression levels in breast tumors of the TCGA cohort had revealed an inverse correlation. This is consistent with the observed induction in CSTA mRNA in MCF-7 cells following ER α knockdown (John Mary et al., 2017). The inverse relationship between CSTA and ER α expression is reiterated in the ER α positive breast cancer cell lines used in this study. ZR-75-1 which has the highest expression of CSTA has the least ER α expression, whereas T47D, which has the least expression of CSTA has the highest expression of ER α . CSTA and ER α expression levels in MCF-7 cells are in between these two extremes. In the previous study, we also showed that estrogen suppressed CSTA expression in MCF-7 cells via ER α (John Mary et al., 2017). However, the mechanistic role of ER α was not known. The prediction of an ERE by JASPAR, the peak of ER α binding revealed by ChIPseq data, and our own validation of increased ER α 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 -ER α binding that leads 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 ER α 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. The study offers no insight into the reasons behind differential responses of the breast cancer cell lines to estrogen stimulation. However, it highlights the role of intron-2 methylation in estrogen regulation of CSTA.

Here we have analyzed two specific regions, namely Region 1 and Region 2, that encompass few of the upstream and intron-2 CpG sites, respectively. While it is worth analyzing methylation at every CpG

site in the CSTA locus, our study shows that intron-2 is the site of convergence of estrogen regulation and DNA methylation-dependent silencing. Interestingly, the ER α binding site is located amidst the intron-2 CpGs analyzed in this study. Our global demethylation experiments with MDA-MB-231 and T47D cells revealed the conflict between ER α binding and DNA methylation in region-2. Due to methylation-dependent silencing, MDA-MB-231 cells do not express ER α (Lapidus et al., 1998; Ottaviano et al., 1994) and CSTA (this study). Global demethylation in MDA-MB-231 cells established functional ER α (as revealed by PR induction upon estrogen stimulation), and CSTA mRNA expression. 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 ER α . The conflict is also supported by the results from T47D cells. It must be noted that T47D cells show significantly greater level of region 2 methylation than MCF-7 cells, which likely prevents ER α binding to the ERE. This is arguably the reason why despite detectable levels of CSTA and functional ER α 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 at this point of time 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 indicate 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 (Zilberman et al., 2007). 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” (Thurman et al., 2012). On the other hand, methylation of cytosine residues in CpG dinucleotides prevents binding of transcription factors to their cognate response elements on DNA (Choy et al., 2010; Comb and Goodman, 1990; Miranda and Jones, 2007; Prendergast et al., 1991) 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 (Höller et al., 1988). Alternatively, methylated CpG sites attract methyl-cytosine binding proteins (MCBPs) (Hendrich and Bird, 1998). 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 TF access. In the context of the interaction between transcriptional activity and transcription factor binding, ER α is not an

exception. Ung and co-workers have analyzed DNA methylation in relation to ER α expression and binding (Ung et al., 2014). They found an inverse correlation between ER α expression and CpG methylation within ER α binding sites. Methylation of these CpGs was therefore interpreted as being dependent on ER α activity, consistent with the passive model of DNA methylation; more the ER α binding, lesser the methylation. Except for T47D, the cell lines used in this study have patterns of ER α 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 ER α expression, cannot be stated with certainty. Parallely, results from T47D cells are consistent with the converse model, wherein despite ER α 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 can restore estrogen regulation at the CSTA locus. We believe that this is most likely due to the restoration of the ER α 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. This result may be exploited for predicting CSTA expression or metastatic progression of breast tumors. Furthermore, this study has uncovered an interesting interplay between ER α binding and transcriptional regulation in the CSTA locus. We propose a model in which CSTA expression in breast cancer cells is an integrated result of estrogen regulation and DNA methylation-dependent silencing converging on the intron-2. Our study provides the framework for epigenetic restoration of CSTA expression in breast tumors using pharmacological agents and understanding the role of estrogen in breast tumor invasion and metastasis.

Figure legends

Fig 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, ER α , 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, ER α , and β -actin. β -actin served as an internal control.

Fig. 2. 5-aza induces the expression of CSTA mRNA in MDA-MB-231 cells. MDA-MB-231 cells were treated with DMSO (vehicle control) or 5-aza for 5 days in M1 medium. Total RNA was isolated and reverse transcribed. The expression of CSTA was analysed 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.

Fig. 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 ($0.2 < \text{beta value} < 0.6$), bright blue = unmethylated ($0 < \text{beta value} \leq 0.2$). 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 significant difference as described in materials and methods. The indicated *p*-values are adjusted *p*-values obtained following Bonferroni correction, which are indicated only for the pair of cell lines with significant difference in the proportion of methylated CpGs.

Fig. 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 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 analysed by Welch two-sample *t*-tests.

Fig. 5. Estrogen enhances ER α occupancy in the intron-2 region of CSTA in MCF-7 cells. A. The location of ERE predicted by JASPAR in the Region 2. The underlined sequence is the ERE. The CpG sites analysed by bisulfite sequencing are indicated by bold letters. B. A snapshot of the UCSC genome browser displaying the results of ChIP-Seq analysis. Fastq files of ChIP-Seq data (Project ID: ERP000380) were retrieved from SRA repository and analysed in Galaxy. Note that E2 treatment increases the ER α occupancy in the intron-2 in a region that overlaps with Region 2 (red rectangle). C. The result of ChIP experiment. 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 ER α or IgG specific antibodies. Immunoprecipitated DNA was reverse cross-linked, purified and subjected to PCR analysis using primers flanking the intron-2 ERE. pS2, a known E2 induced gene served as a positive control. Data shown are representative of three independent experiments.

Fig. 6. Global demethylation restores estrogen regulation of CSTA. MDA-MB-231 (A-C) or T47D (D-F) cells were subjected to global demethylation using 5-aza. The cells were then stimulated with 10 nM E2 or ethanol (vehicle) for 24 h. CSTA expression was analyzed by western blotting (A, B, D, E) and RT-PCR (C) or qRT-PCR (F). CycA was used as internal control in RT-PCR and qRT-PCR analyses. Histone H3 served as an internal control in western blots. Relative CSTA mRNA expression data are represented as bar graphs. The expression in the control samples (without 5-aza and E2 treatments) were set to 1 and the expression in the other treatment groups were expressed relative to the control. Bars represents mean relative expression ($\pm$ S.D.) Data were analyzed by ANOVA followed by Tukey's HSD. $n = 3$ biological replicates for MDA-MB-231 and $n = 4$ biological replicates for T47D. Each biological replicate comprises of one dish each for control and treated cells. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Acknowledgements

This work was supported by a grant from the Department of Biotechnology (DBT), Govt. of India (Sanction order No. 6242-P90/RGCB/PMD/DBT/ANLM, to A.M.L.). We acknowledge the support from the DBT funded Bioinformatics Facility at IIT Guwahati, and infrastructural support from IIT Guwahati. We thank Mohan C. Manjegowda for his help in ChIP-Seq analysis.

References

499 Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements,
500 D., Coraor, N., Eberhard, C., Grüning, B., Guerler, A., Hillman-Jackson, J., Von Kuster, G., Rasche,
501 E., Soranzo, N., Turaga, N., Taylor, J., Nekrutenko, A., Goecks, J., 2016. The Galaxy platform for
502 accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* 44,
503 W3–W10. <https://doi.org/10.1093/nar/gkw343>

504 Ai, L., Kim, W.-J., Kim, T.-Y., Fields, C.R., Massoll, N.A., Robertson, K.D., Brown, K.D., 2006.
505 Epigenetic silencing of the tumor suppressor cystatin M occurs during breast cancer progression.
506 *Cancer Res.* 66, 7899–909. <https://doi.org/10.1158/0008-5472.CAN-06-0576>

507 Al-Dhaheri, M.H., Shah, Y.M., Basrur, V., Pind, S., Rowan, B.G., 2006. Identification of novel proteins
508 induced by estradiol, 4-hydroxytamoxifen and acolbifene in T47D breast cancer cells. *Steroids* 71,
509 966–978. <https://doi.org/10.1016/j.steroids.2006.07.006>

510 Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data. Babraham
511 Bioinformatics. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

512 Arapshian, A., Kuppumbatti, Y.S., Mira-y-Lopez, R., 2000. Methylation of conserved CpG sites
513 neighboring the beta retinoic acid response element may mediate retinoic acid receptor beta gene
514 silencing in MCF-7 breast cancer cells. *Oncogene* 19, 4066–70.
515 <https://doi.org/10.1038/sj.onc.1203734>

516 Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., Nekrutenko, A., Galaxy Team, the
517 G., 2010. Manipulation of FASTQ data with Galaxy. *Bioinformatics* 26, 1783–5.
518 <https://doi.org/10.1093/bioinformatics/btq281> (accessed on 12 February 2017)

519 Calkins, C.C., Sloane, B.F., 1995. Mammalian cysteine protease inhibitors: biochemical properties and
520 possible roles in tumor progression. *Biol. Chem. Hoppe. Seyler.* 376, 71–80.

521 Choy, M.-K., Movassagh, M., Goh, H.-G., Bennett, M.R., Down, T.A., Foo, R.S.Y., 2010. Genome-wide
522 conserved consensus transcription factor binding motifs are hyper-methylated. *BMC Genomics* 11,
523 519. <https://doi.org/10.1186/1471-2164-11-519>

524 Comb, M., Goodman, H.M., 1990. CpG methylation inhibits proenkephalin gene expression and binding
525 of the transcription factor AP-2. *Nucleic Acids Res.* 18, 3975–3982.
526 <https://doi.org/10.1093/nar/18.13.3975>

527 Deng, G., Song, G.-A., Pong, E., Sleisenger, M., Kim, Y.S., 2004. 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, 2692–8.

Duivenvoorden, H.M., Rautela, J., Edgington-Mitchell, L.E., Spurling, A., Greening, D.W., Nowell, C.J., Molloy, T.J., Robbins, E., Brockwell, N.K., Lee, C.S., Chen, M., Holliday, A., Selinger, C.I., Hu, M., Britt, K.L., Stroud, D.A., Bogyo, M., Möller, A., Polyak, K., Sloane, B.F., O’Toole, S.A., Parker, B.S., 2017. Myoepithelial cell-specific expression of stefin A as a suppressor of early breast cancer invasion. *J. Pathol.* 243, 496–509. <https://doi.org/10.1002/path.4990>

Fonović, M., Turk, B., 2014. Cysteine cathepsins and extracellular matrix degradation. *Biochim. Biophys. Acta* 1840, 2560–70. <https://doi.org/10.1016/j.bbagen.2014.03.017>

Goldman, M., Craft, B., Hastie, M., Repečka, K., Kamath, A., McDade, F., Rogers, D., Brooks, A.N., Zhu, J., Haussler, D., 2019. The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. *bioRxiv* 326470. <https://doi.org/10.1101/326470>

Graff, J.R., Gabrielson, E., Fujii, H., Baylin, S.B., Herman, J.G., 2000. 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, 2727–32. <https://doi.org/10.1074/JBC.275.4.2727>

Hendrich, B., Bird, A., 1998. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.* 18, 6538–47.

Höller, M., Westin, G., Jiricny, J., Schaffner, W., 1988. Sp1 transcription factor binds DNA and activates transcription even when the binding site is CpG methylated. *Genes Dev.* 2, 1127–35.

John Mary, D.J.S., Manjegowda, M.C., Kumar, A., Dutta, S., Limaye, A.M., 2017. The role of cystatin A in breast cancer and its functional link with ER α . *J. Genet. Genomics* 44, 593–597. <https://doi.org/10.1016/j.jgg.2017.10.001>

Jones, C., Mackay, A., Grigoriadis, A., Cossu, A., Reis-Filho, J.S., Fulford, L., Dexter, T., Davies, S., Bulmer, K., Ford, E., Parry, S., Budroni, M., Palmieri, G., Neville, A.M., O’Hare, M.J., Lakhani, S.R., 2004. Expression Profiling of Purified Normal Human Luminal and Myoepithelial Breast Cells. *Cancer Res.* 64, 3037–3045. <https://doi.org/10.1158/0008-5472.CAN-03-2028>

Karlsson, J.O., Ostwald, K., Kabjorn, C., Andersson, M., 1994. A Method for Protein Assay in Laemmli Buffer. *Anal. Biochem.* 219, 144–146. <https://doi.org/10.1006/abio.1994.1243>

557 Kazanets, A., Shorstova, T., Hilmi, K., Marques, M., Witcher, M., 2016. Epigenetic silencing of tumor
558 suppressor genes: Paradigms, puzzles, and potential. *Biochim. Biophys. Acta* 1865, 275–88.
559 <https://doi.org/10.1016/j.bbcan.2016.04.001>

560 Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., Haussler, a. D., 2002.
561 The Human Genome Browser at UCSC. *Genome Res.* 12, 996–1006.
562 <https://doi.org/10.1101/gr.229102>

563 Keppler, D., 2006. Towards novel anti-cancer strategies based on cystatin function. *Cancer Lett.* 235,
564 159–176. <https://doi.org/10.1016/j.canlet.2005.04.001>

565 Khaket, T.P., Kwon, T.K., Kang, S.C., 2019. Cathepsins: Potent regulators in carcinogenesis. *Pharmacol.*
566 *Ther.* <https://doi.org/10.1016/j.pharmthera.2019.02.003>

567 Koboldt, D.C., Fulton, R.S., McLellan, M.D., Schmidt, H., Kalicki-Veizer, J., McMichael, J.F., Fulton,
568 L.L., Dooling, D.J., Ding, L., Mardis, E.R., Wilson, R.K., Ally, A., Balasundaram, M., Butterfield,
569 Y.S.N., Carlsen, R., Carter, C., Chu, A., Chuah, E., Chun, H.-J.E., Coope, R.J.N., Dhalla, N., Guin,
570 R., Hirst, C., Hirst, M., Holt, R.A., Lee, D., Li, H.I., Mayo, M., Moore, R.A., Mungall, A.J.,
571 Pleasance, E., Gordon Robertson, A., Schein, J.E., Shafiei, A., Sipahimalani, P., Slobodan, J.R.,
572 Stoll, D., Tam, A., Thiessen, N., Varhol, R.J., Wye, N., Zeng, T., Zhao, Y., Birol, I., Jones, S.J.M.,
573 Marra, M.A., Cherniack, A.D., Saksena, G., Onofrio, R.C., Pho, N.H., Carter, S.L., Schumacher,
574 S.E., Tabak, B., Hernandez, B., Gentry, J., Nguyen, H., Crenshaw, A., Ardlie, K., Beroukhi, R.,
575 Winckler, W., Getz, G., Gabriel, S.B., Meyerson, M., Chin, L., Park, P.J., Kucherlapati, R.,
576 Hoadley, K.A., Todd Auman, J., Fan, C., Turman, Y.J., Shi, Y., Li, L., Topal, M.D., He, X., Chao,
577 H.-H., Prat, A., Silva, G.O., Iglesia, M.D., Zhao, W., Usary, J., Berg, J.S., Adams, M., Booker, J.,
578 Wu, J., Gulabani, A., Bodenheimer, T., Hoyle, A.P., Simons, J. V., Soloway, M.G., Mose, L.E.,
579 Jefferys, S.R., Balu, S., Parker, J.S., Neil Hayes, D., Perou, C.M., Malik, S., Mahurkar, S., Shen, H.,
580 Weisenberger, D.J., Triche Jr, T., Lai, P.H., Bootwalla, M.S., Maglinte, D.T., Berman, B.P., Van
581 Den Berg, D.J., Baylin, S.B., Laird, P.W., Creighton, C.J., Donehower, L.A., Getz, G., Noble, M.,
582 Voet, D., Saksena, G., Gehlenborg, N., DiCara, D., Zhang, J., Zhang, H., Wu, C.-J., Yingchun Liu,
583 S., Lawrence, M.S., Zou, L., Sivachenko, A., Lin, P., Stojanov, P., Jing, R., Cho, J., Sinha, R., Park,
584 R.W., Nazaire, M.-D., Robinson, J., Thorvaldsdottir, H., Mesirov, J., Park, P.J., Chin, L., Reynolds,
585 S., Kreisberg, R.B., Bernard, B., Bressler, R., Erkkila, T., Lin, J., Thorsson, V., Zhang, W.,
586 Shmulevich, I., Ciriello, G., Weinhold, N., Schultz, N., Gao, J., Cerami, E., Gross, B., Jacobsen, A.,
587 Sinha, R., Arman Aksoy, B., Antipin, Y., Reva, B., Shen, R., Taylor, B.S., Ladanyi, M., Sander, C.,
588 Anur, P., Spellman, P.T., Lu, Y., Liu, W., Verhaak, R.R.G., Mills, G.B., Akbani, R., Zhang, N.,

Broom, B.M., Casasent, T.D., Wakefield, C., Unruh, A.K., Baggerly, K., Coombes, K., Weinstein, J.N., Haussler, D., Benz, C.C., Stuart, J.M., Benz, S.C., Zhu, J., Szeto, C.C., Scott, G.K., Yau, C., Paull, E.O., Carlin, D., Wong, C., Sokolov, A., Thusberg, J., Mooney, S., Ng, S., Goldstein, T.C., Ellrott, K., Grifford, M., Wilks, C., Ma, S., Craft, B., Yan, C., Hu, Y., Meerzaman, D., Gastier-Foster, J.M., Bowen, J., Ramirez, N.C., Black, A.D., XPATH ERROR: unknown variable "tname"., R.E., White, P., Zmuda, E.J., Frick, J., Lichtenberg, T.M., Brookens, R., George, M.M., Gerken, M.A., Harper, H.A., Leraas, K.M., Wise, L.J., Tabler, T.R., McAllister, C., Barr, T., Hart-Kothari, M., Tarvin, K., Saller, C., Sandusky, G., Mitchell, C., Iacocca, M. V., Brown, J., Rabeno, B., Czerwinski, C., Petrelli, N., Dolzhansky, O., Abramov, M., Voronina, O., Potapova, O., Marks, J.R., Suchorska, W.M., Murawa, D., Kycler, W., Ibbs, M., Korski, K., Spychała, A., Murawa, P., Brzeziński, J.J., Perz, H., Łażniak, R., Teresiak, M., Tatka, H., Leporowska, E., Bogusz-Czerniewicz, M., Malicki, J., Mackiewicz, A., Wiznerowicz, M., Van Le, X., Kohl, B., Viet Tien, N., Thorp, R., Van Bang, N., Sussman, H., Duc Phu, B., Hajek, R., Phi Hung, N., Viet The Phuong, T., Quyet Thang, H., Zaki Khan, K., Penny, R., Mallery, D., Curley, E., Shelton, C., Yena, P., Ingle, J.N., Couch, F.J., Lingle, W.L., King, T.A., Maria Gonzalez-Angulo, A., Mills, G.B., Dyer, M.D., Liu, S., Meng, X., Patangan, M., Waldman, F., Stöppler, H., Kimryn Rathmell, W., Thorne, L., Huang, M., Boice, L., Hill, A., Morrison, C., Gaudioso, C., Bshara, W., Daily, K., Egea, S.C., Pegram, M.D., Gomez-Fernandez, C., Dhir, R., Bhargava, R., Brufsky, A., Shriver, C.D., Hooke, J.A., Leigh Campbell, J., Mural, R.J., Hu, H., Somiari, S., Larson, C., Deyarmin, B., Kvecher, L., Kovatich, A.J., Ellis, M.J., King, T.A., Hu, H., Couch, F.J., Mural, R.J., Stricker, T., White, K., Olopade, O., Ingle, J.N., Luo, C., Chen, Y., Marks, J.R., Waldman, F., Wiznerowicz, M., Bose, R., Chang, L.-W., Beck, A.H., Maria Gonzalez-Angulo, A., Pihl, T., Jensen, M., Sfeir, R., Kahn, A., Chu, A., Kothiyal, P., Wang, Z., Snyder, E., Pontius, J., Ayala, B., Backus, M., Walton, J., Baboud, J., Berton, D., Nicholls, M., Srinivasan, D., Raman, R., Girshik, S., Kigonya, P., Alonso, S., Sanbhadti, R., Barletta, S., Pot, D., Sheth, M., Demchok, J.A., Mills Shaw, K.R., Yang, L., Eley, G., Ferguson, M.L., Tarnuzzer, R.W., Zhang, J., Dillon, L.A.L., Buetow, K., Fielding, P., Ozenberger, B.A., Guyer, M.S., Sofia, H.J., Palchik, J.D., 2012. Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61–70. <https://doi.org/10.1038/nature11412>

Kolwijck, E., Kos, J., Obermajer, N., Span, P.N., Thomas, C.M.G., Massuger, L.F.A.G., Sweep, F.C.G.J., 2010. The balance between extracellular cathepsins and cystatin C is of importance for ovarian cancer. *Eur. J. Clin. Invest.* 40, 591–9. <https://doi.org/10.1111/j.1365-2362.2010.02305.x>

Kuopio, T., Kankaanranta, a, Jalava, P., Kronqvist, P., Kotkansalo, T., Weber, E., Collan, Y., 1998. Cysteine proteinase inhibitor cystatin A in breast cancer. *Cancer Res.* 58, 432–436.

- Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–5.
- Lah, T.T., Kokalj-Kunovar, M., Strukelj, B., Pungercar, J., Barlic-Maganja, D., Drobnic-Kosorok, M., Kastelic, L., Babnik, J., Golouh, R., Turk, V., 1992. Stefins and lysosomal cathepsins B, L and D in human breast carcinoma. *Int. J. cancer* 50, 36–44.
- Lah, T.T., Kos, J., Blejec, A., Frkovic-Georgio, S., Golouh, R., Vrhovec, I., Turk, V., 1997. The Expression of Lysosomal Proteinases and Their Inhibitors in Breast Cancer: Possible Relationship to Prognosis of the Disease. *Pathol. Oncol. Res.* 3, 89–99.
- Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25. <https://doi.org/10.1186/gb-2009-10-3-r25>
- Lapidus, R.G., Nass, S.J., Butash, K.A., Parl, F.F., Weitzman, S.A., Graff, J.G., Herman, J.G., Davidson, N.E., 1998. Mapping of ER gene CpG island methylation-specific polymerase chain reaction. *Cancer Res.* 58, 2515–9.
- Levicar, N., Kos, J., Blejec, A., Golouh, R., Vrhovec, I., Frkovic-Grazio, S., Lah, T.T., 2002. 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, 42–9.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 1000 Genome Project Data Processing, 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–9. <https://doi.org/10.1093/bioinformatics/btp352>
- Lowry, O.H., Rosebrough, N.J., Farr, A.L., Randall, R.J., 1951. Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* 193, 265–75.
- Ma, Y., Chen, Y., Li, Y., Grün, K., Berndt, A., Zhou, Z., Petersen, I., 2018. Cystatin A suppresses tumor cell growth through inhibiting epithelial to mesenchymal transition in human lung cancer. *Oncotarget* 9, 14084–14098. <https://doi.org/10.18632/oncotarget.23505>
- Miranda, T.B., Jones, P.A., 2007. DNA methylation: the nuts and bolts of repression. *J. Cell. Physiol.* 213, 384–90. <https://doi.org/10.1002/jcp.21224>

651 Ottaviano, Y.L., Issa, J.P., Parl, F.F., Smith, H.S., Baylin, S.B., Davidson, N.E., 1994. Methylation of the
652 estrogen receptor gene CpG island marks loss of estrogen receptor expression in human breast
653 cancer cells. *Cancer Res.* 54, 2552–5.

654 Palermo, C., Joyce, J.A., 2008. Cysteine cathepsin proteases as pharmacological targets in cancer. *Trends*
655 *Pharmacol. Sci.* 29, 22–8. <https://doi.org/10.1016/j.tips.2007.10.011>

656 Paraoan, L., Gray, D., Hiscott, P., Garcia-Finana, M., Lane, B., Damato, B., Grierson, I., 2009. Cathepsin
657 S and its inhibitor cystatin C: imbalance in uveal melanoma. *Front. Biosci. (Landmark Ed.* 14,
658 2504–13.

659 Parker, B., Ciocca, D., Bidwell, B., Gago, F., Fanelli, M., George, J., Slavin, J., Möller, A., Steel, R.,
660 Pouliot, N., Eckhardt, B., Henderson, M., Anderson, R., 2008. Primary tumour expression of the
661 cysteine cathepsin inhibitor Stefin A inhibits distant metastasis in breast cancer. *J. Pathol.* 214, 337–
662 346. <https://doi.org/10.1002/path.2265>

663 Pogorzelska, A., Żołnowska, B., Bartoszewski, R., 2018. Cysteine cathepsins as a prospective target for
664 anticancer therapies-current progress and prospects. *Biochimie* 151, 85–106.
665 <https://doi.org/10.1016/j.biochi.2018.05.023>

666 Prendergast, G.C., Lawe, D., Ziff, E.B., 1991. Association of Myc, the murine homolog of max, with c-
667 Myc stimulates methylation-sensitive DNA binding and ras cotransformation. *Cell* 65, 395–407.

668 Pryzbylowski, P., Obajimi, O., Keen, J.C., 2008. Trichostatin A and 5 Aza-2' deoxycytidine decrease
669 estrogen receptor mRNA stability in ER positive MCF7 cells through modulation of HuR. *Breast*
670 *Cancer Res. Treat.* 111, 15–25. <https://doi.org/10.1007/s10549-007-9751-0>

671 Rivenbark, A., Coleman, W., 2009. Epigenetic regulation of cystatins in cancer. *Front. Biosci. Volume,*
672 453. <https://doi.org/10.2741/3254>

673 Rivenbark, A.G., Jones, W.D., Coleman, W.B., 2006a. DNA methylation-dependent silencing of CST6 in
674 human breast cancer cell lines. *Lab. Invest.* 86, 1233–42. <https://doi.org/10.1038/labinvest.3700485>

675 Rivenbark, A.G., Jones, W.D., Risher, J.D., Coleman, W.B., 2006b. DNA methylation-dependent
676 epigenetic regulation of gene expression in MCF-7 breast cancer cells. *Epigenetics* 1, 32–44.

677 Sandelin, A., Alkema, W., Engström, P., Wasserman, W.W., Lenhard, B., 2004. JASPAR: an open-access
678 database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.* 32, 91D–94.
679 <https://doi.org/10.1093/nar/gkh012>

680 Sathyanarayana, U.G., Maruyama, R., Padar, A., Suzuki, M., Bondaruk, J., Sagalowsky, A., Minna, J.D.,
681 Frenkel, E.P., Grossman, H.B., Czerniak, B., Gazdar, A.F., 2004. Molecular detection of
682 noninvasive and invasive bladder tumor tissues and exfoliated cells by aberrant promoter
683 methylation of laminin-5 encoding genes. *Cancer Res.* 64, 1425–30.

684 Sathyanarayana, U.G., Padar, A., Huang, C.X., Suzuki, M., Shigematsu, H., Bekele, B.N., Gazdar, A.F.,
685 2003a. Aberrant promoter methylation and silencing of laminin-5-encoding genes in breast
686 carcinoma. *Clin. Cancer Res.* 9, 6389–94.

687 Sathyanarayana, U.G., Padar, A., Suzuki, M., Maruyama, R., Shigematsu, H., Hsieh, J.-T., Frenkel, E.P.,
688 Gazdar, A.F., 2003b. Aberrant promoter methylation of laminin-5-encoding genes in prostate
689 cancers and its relationship to clinicopathological features. *Clin. Cancer Res.* 9, 6395–400.

690 Sathyanarayana, U.G., Toyooka, S., Padar, A., Takahashi, T., Brambilla, E., Minna, J.D., Gazdar, A.F.,
691 2003c. Epigenetic inactivation of laminin-5-encoding genes in lung cancers. *Clin. Cancer Res.* 9,
692 2665–72.

693 Schagdarsurengin, U., Pfeifer, G.P., Dammann, R., 2007. Frequent epigenetic inactivation of cystatin M
694 in breast carcinoma. *Oncogene* 26, 3089–94. <https://doi.org/10.1038/sj.onc.1210107>

695 Shivaswamy, S., Iyer, V., 2007. Genome-wide analysis of chromatin status using tiling microarrays.
696 *Methods* 41, 304–311. <https://doi.org/10.1016/j.ymeth.2006.11.002>

697 Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C.,
698 Stergachis, A.B., Wang, H., Vernot, B., Garg, K., John, S., Sandstrom, R., Bates, D., Boatman, L.,
699 Canfield, T.K., Diegel, M., Dunn, D., Ebersol, A.K., Frum, T., Giste, E., Johnson, A.K., Johnson,
700 E.M., Kutayavin, T., Lajoie, B., Lee, B.-K., Lee, K., London, D., Lotakis, D., Neph, S., Neri, F.,
701 Nguyen, E.D., Qu, H., Reynolds, A.P., Roach, V., Safi, A., Sanchez, M.E., Sanyal, A., Shafer, A.,
702 Simon, J.M., Song, L., Vong, S., Weaver, M., Yan, Y., Zhang, Z., Zhang, Z., Lenhard, B., Tewari,
703 M., Dorschner, M.O., Hansen, R.S., Navas, P.A., Stamatoyannopoulos, G., Iyer, V.R., Lieb, J.D.,
704 Sunyaev, S.R., Akey, J.M., Sabo, P.J., Kaul, R., Furey, T.S., Dekker, J., Crawford, G.E.,
705 Stamatoyannopoulos, J.A., 2012. The accessible chromatin landscape of the human genome. *Nature*
706 489, 75–82. <https://doi.org/10.1038/nature11232>

707 Turk, B., Turk, D., Turk, V., 2000. Lysosomal cysteine proteases: more than scavengers. *Biochim.*
708 *Biophys. Acta* 1477, 98–111.

709 Turk, V., Stoka, V., Vasiljeva, O., Renko, M., Sun, T., Turk, B., Turk, D., 2012. Cysteine cathepsins:

710 From structure, function and regulation to new frontiers. *Biochim. Biophys. Acta - Proteins*
711 *Proteomics* 1824, 68–88. <https://doi.org/10.1016/j.bbapap.2011.10.002>

712 Turk, V., Turk, B., Turk, D., 2001. Lysosomal cysteine proteases: facts and opportunities. *EMBO J.* 20,
713 4629–33. <https://doi.org/10.1093/emboj/20.17.4629>

714 Ung, M., Ma, X., Johnson, K.C., Christensen, B.C., Cheng, C., 2014. Effect of estrogen receptor α
715 binding on functional DNA methylation in breast cancer. *Epigenetics* 9, 523–532.
716 <https://doi.org/10.4161/epi.27688>

717 Verma, S., Dixit, R., Pandey, K.C., 2016. Cysteine Proteases: Modes of Activation and Future Prospects
718 as Pharmacological Targets. *Front. Pharmacol.* 7, 107. <https://doi.org/10.3389/fphar.2016.00107>

719 Vidak, E., Javoršek, U., Vizovišek, M., Turk, B., 2019. Cysteine Cathepsins and their Extracellular Roles:
720 Shaping the Microenvironment. *Cells* 8, 264. <https://doi.org/10.3390/cells8030264>

721 Vizovišek, M., Fonović, M., Turk, B., 2019. Cysteine cathepsins in extracellular matrix remodeling:
722 Extracellular matrix degradation and beyond. *Matrix Biol.* 75–76, 141–159.
723 <https://doi.org/10.1016/j.matbio.2018.01.024>

724 Withana, N.P., Blum, G., Sameni, M., Slaney, C., Anbalagan, A., Olive, M.B., Bidwell, B.N., Edgington,
725 L., Wang, L., Moin, K., Sloane, B.F., Anderson, R.L., Bogoy, M.S., Parker, B.S., 2012. Cathepsin B
726 inhibition limits bone metastasis in breast cancer. *Cancer Res.* 72, 1199–209.
727 <https://doi.org/10.1158/0008-5472.CAN-11-2759>

728 Wormke, M., Stoner, M., Saville, B., Safe, S., 2000. Crosstalk between estrogen receptor α and the
729 aryl hydrocarbon receptor in breast cancer cells involves unidirectional activation of proteasomes.
730 *FEBS Lett.* 478, 109–12.

731 Yano, M., Hirai, K., Naito, Z., Yokoyama, M., Ishiwata, T., Shiraki, Y., Inokuchi, M., Asano, G., 2001.
732 Expression of cathepsin B and cystatin C in human breast cancer. *Surg. Today* 31, 385–9.
733 <https://doi.org/10.1007/s005950170126>

734 Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers,
735 R.M., Brown, M., Li, W., Liu, X.S., 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome*
736 *Biol.* 9, R137. <https://doi.org/10.1186/gb-2008-9-9-r137>

737 Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T., Henikoff, S., 2007. Genome-wide analysis of
738 *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and

739 transcription. Nat. Genet. 39, 61–9. <https://doi.org/10.1038/ng1929>

740

Table 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

Figure 1

A

B

Figure 2

Figure 3

A

B

C

Figure 4

Figure 5

A

TTCTGCTATCAAACCTTTTCCTACTGGATCTCAGCCAC**CG**ATCCCAGTTCCCTTTTACTTC
CTGGTAGTCTGGCTGTTGATCCCTTTGCTCTGAGGCACTCTAGATTTAAGGTCTTGCCAG
TGATGTGACCTTCTCTATGTATTTCAAGTACCTATCAAGAGGTAGGTGGTAGAATGGAAG
GACCACAAGCTTAGGTGCAGAGTGTCCTGGGTTTGAACCCTTGTTCAATTTGTTCTATG
GGAAGCTCCTCCTCCTCTCTGAGCCTTCATTCCTTATCTGCACAATGAGGGTAATAATC
TACTTT**CG**CAG**CG**TGTTGTGAGGAATAAATAAGCTGGAAATTTATTGAGCACTTATAATTC
ACTATGCACTATTCTAAGAACAGGGCTT

B

C

Figure 6

A

B

C

D

E

F