Dai et al. BMC Cancer (2021) 21:945
https://doi.org/10.1186/s12885-021-08645-3

BMC Cancer

RESEARCH ARTICLE Open Access

Gene-associated methylation status of *ST14* as a predictor of survival and hormone receptor positivity in breast Cancer

Yang-Hong Dai[1], Ying-Fu Wang[1], Po-Chien Shen[1], Cheng-Hsiang Lo[1], Jen-Fu Yang[1], Chun-Shu Lin[1], Hsing-Lung Chao[1,2] and Wen-Yen Huang[1,3*]

Abstract

Background: Genomic profiles of specific gene sets have been established to guide personalized treatment and prognosis for patients with breast cancer (BC). However, epigenomic information has not yet been applied in a clinical setting. *ST14* encodes matriptase, a proteinase that is widely expressed in BC with reported prognostic value.

Methods: In this present study, we evaluated the effect of *ST14* DNA methylation (DNAm) on overall survival (OS) of patients with BC as a representative example to promote the use of the epigenome in clinical decisions. We analyzed publicly available genomic and epigenomic data from 1361 BC patients. Methylation was characterized by the β-value from CpG probes based on sequencing with the Illumina Human 450 K platform.

Results: A high mean DNAm ($\beta > 0.6779$) across 34 CpG probes for *ST14*, as the gene-associated methylation (GAM) pattern, was associated with a longer OS after adjusting age, stage, histology and molecular features in Cox model (p value < 0.001). A high GAM status was also associated with a higher *XBP1* expression level and higher proportion of hormone-positive BC (p value < 0.001). Pathway analysis revealed that altered GAM was related to matrisome-associated pathway.

Conclusions: Here we show the potential role of *ST14* DNAm in BC prognosis and warrant further study.

Keywords: ST14, Matriptase, DNA methylation, Breast Cancer

Background

Progress in high-throughput gene expression profiling has now allowed for many types of genomic tests to be adopted in clinical practice for improving breast cancer (BC) classification and prognosis, such as the commercially available platforms PAM50, Mammaprint (Agenda, Huntington Beach, CA, USA), and Oncotype DX (Genomic Health, Redwood City, CA, USA) [1–3]. These tests were designed with the goal of promoting personalized treatment and management for patients with BC to ultimately improve their prognosis and survival. Although such gene-based profiling has achieved great success, recent studies have highlighted that combining gene signatures with DNA methylation (DNAm) patterns could further help to refine the molecular classification of BC [4–7].

DNAm is an epigenetic mark involving the addition of a methyl group to the cytosine pyrimidine in CpG dinucleotides [8]. Without changing the genetic sequence, DNAm alters transcription through integration of genome structures [9], and could thus serve as a useful molecular marker for gene silencing memory [10]. Indeed, BC subtypes have been associated with different methylation

* Correspondence: hwyyi@yahoo.com.tw
[1]Department of Radiation Oncology, Tri-Service General Hospital, National Defense Medical Center, No. 325, Chengong Rd., Sec. 2, Neihu, Taipei 114, Taiwan
[3]Institute of Clinical Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan
Full list of author information is available at the end of the article

© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

profiles [11]; thus, combining methylation profiling and gene expression data could help to elucidate the factors and mechanisms underlying the observed pathological and clinical heterogeneity among BC types [12]. Previous studies on epigenetic signatures of BC have mainly focused on promoter hypermethylation that leads to the silencing of tumor suppressor genes [5–7]. For example, Zhu et al. showed that methylation of the BRCA1 promoter was closely associated with decreased overall survival (OS) and disease-free survival in patients with basal-like BC [7]. Some studies revealed a high frequency of abnormal CDH1 promoter methylation in ductal breast tumors and its association with carcinogenesis [6]. However, in contrast to the wide use of the genomic tests mentioned above, methylation in only a few genes has been shown to have prognostic predictive power in BC [13]. Therefore, identification of the prognostic value of additional genes along with their methylation status is required for establishment of a complete DNAm panel.

One such candidate gene is suppressor of tumorigenicity 14 (*ST14*), which encodes matriptase, a type 2 transmembrane serine protease that plays crucial roles in physiology and cancer biology [14–16]. This protease was first discovered in BC cell lines and has been subsequently identified in various other cancer types, including ovarian, prostate, and colon cancers [17]. Overexpression of matriptase or *ST14* has also been associated with epithelial-mesenchymal transition (EMT), which in turn contributes to cancer metastasis or progression [18, 19]. In a study by Kim et al., they showed that high *ST14* expression was associated with poor survival in estrogen receptor negative patients and concluded that *ST14* is an emerging therapeutic target [20]. Accordingly, overexpression of matriptase has been shown to correlate with BC progression and a poor prognosis [21–26], whereas reduced matriptase levels could abrogate tumor progression, proliferation, and invasion in both a mouse model and BC cell lines [27, 28].

In a pan-cancer analysis, CpG methylation patterns in the promoter and gene body showed distinct correlations with gene expression levels, and different methylation patterns also influenced gene expression [29]. To date, differential expression of *ST14* according to different methylation patterns has not been investigated yet; and there were few studies addressing the role of *ST14* methylation in cancer. Existing evidence was limited to pancreatic adenocarcinoma where *ST14* is aberrantly methylated [30]. As *ST14* gene expression in BC prognosis has been reported, possible role of *ST14* methylation should be studied. Therefore, in contrast to focusing only on the promoter regions, we sought to explore the extended methylation status, including the promoter, gene body, and 3′ untranslated regions (3′ UTRs), in the DNA. We hypothesized that these gene-associated methylation (GAM) sites could cover the entire methylation information associated with a specific gene. As an example of this approach, we here highlight the impact of *ST14* methylation on OS in patients with BC, and investigate its potential as a novel prognostic biomarker.

Methods

Data and study design

Publicly available bioinformatics data of BC were retrieved from the Cancer Genome Atlas (TCGA), which was used as the primary dataset for obtaining the methylation profile and associated gene expression analyses. Genes associated with matriptase and EMT in BC were used to develop a classifier for the GAM status in *ST14*, which was then validated with additional datasets, GSE5364 and GSE22820. GSE75067 was used to confirm the impact of the GAM status on OS. The flowchart of the study are shown in Additional file 1: Fig. S1.

The TCGA breast carcinoma (BRCA) cohort was retrieved from Xena Browser (https://xenabrowser.net/). Initially, 1239 tumors were identified in the dataset. After excluding tumors lacking *ST14* methylation information, the final dataset included 858 tumors. Raw methylation data were based on Human Methylation 450 K (HM450K) bead arrays, and were downloaded from the Genomic Data Commons hub (https://gdc.xenahubs.net/). Following the analysis pipeline proposed by Tian et al., the raw data in IDAT format were first imported and processed with the 'ChAMP' package in R (version 2.22) [31]. Initially, 485,512 probes were identified. Those with a detection p value > 0.01 and with < 3 beads in at least 5% of samples were filtered out. Non-CpG, single nucleotide polymorphism-related probes, and multi-hit probes were further removed. Probes located in chromosomes X and Y were also excluded. The distribution of type II probes was normalized using the BMIQ function [32]. Singular value decomposition (SVD) analysis was then used to correlate the principal components with biological and technical factors. If the result of SVD analysis showed substantial technical variation, the ComBat function was used to remove the source of this variation [33]. After processing, 34 probes annotated with *ST14* remained, and each probe was characterized by its normalized β-value. We used β-value in the subsequent analyses as it corresponds approximately to the percentage of a methylated site, which is of biological effect for interpretation [34]. These probes corresponded to regions that covered the promoter, enhancer, 5′-UTR, gene body, and 3′-UTR in *ST14* (Additional file 2: Table S1). Since the mean β-value across CpG probes was shown to be associated with BC in an epigenome-wide association study [35, 36], we defined the GAM by averaging the β-values across the 34 probes. Within the dataset, the median averaged β-value was set as the cut-off screening point to

determine the GAM status. The median was found to be 0.6779 (i.e., 67.79% methylation) and tumors were stratified into GAM-High and GAM-Low. Differentially methylated probes (DMPs) were identified according to the different GAM status using the limma package in R (version 3.48). Using a linear model, the CpG probes of GAM-High and GAM-Low were compared and the outputs including the average expression, logFold Change (FC), *P*-value, and t-statistic were summarized; probes with an adjusted *p*-value < 0.05 were considered to be significantly different between the groups.

Level-3, normalized RNA sequencing (RNAseq) data were downloaded from Xena Browser (https://xenabrowser.net/). Gene expression levels were quantified as reads per kilobase per million. Pearson's correlation coefficient was used to evaluate the correlation between gene and methylation levels. The coefficients between each CpG probe and all of the matriptase- or EMT-associated genes were summed for evaluation of the overall correlation. Unsupervised hierarchical clustering was used to visualize the data clustering.

The GSE75067 dataset was based on a study assessing the association of methylation patterns with subtypes in BC [37]. The 450 K methylation data were obtained from GenomeStudio (Illumina) and converted to β-values. CpG probes with a detection *p*-value > 0.05 or number of beads < 3 were considered as missing measurements and excluded from the analysis. The bias between type I and II probes was adjusted using a peak normalization algorithm. The β-values were further smoothened by the Epanechnikov kernel function to estimate the unmethylated and methylated peaks for each chemical assay. Linear scaling was used to adjust the peaks toward 0 and 1 for the unmethylated and methylated probes, respectively. The CpG probes in *ST14* were identified and their methylation levels were adjusted with methylation median centering. This dataset included additional information about OS and HS, allowing for validation of the results derived from the TCGA cohort. Tumors with complete profiles of methylation, OS, and HS were retained for analysis, leading to a total of 144 samples.

The GSE5364 and GSE22820 datasets were used for gene expression analyses. The gene expression data were derived from the GPL96 and GPL6480 platforms, respectively. In brief, the data from GSE5364 were processed with the MAS5 algorithm and median-centering, while the data from GSE22820 were normalized with default procedures in GeneSpring 7.3.1. For genes with multiple probes, the expression of a gene was determined by geometrically averaging the probe intensities. Overall, 359 breast tumor samples were available for subsequent analyses.

Establishment of a classifier for GAM status in *ST14*

To develop a classifier for the GAM status, genes capable of distinguishing high and low methylation levels were first included as a training set. A previous study indicated that TWIST-induced EMT triggered chromatin accessibility and alterations of DNAm in human mammary epithelial cells [38]. Moreover, EMT was shown to cause widespread genome hypo-methylation and promoter hyper-methylation variations in response to extracellular signaling [39]. Therefore, we hypothesized that genes encoding proteins associated with *ST14* and EMT might reflect the alteration of methylation levels in *ST14*. Using the keywords "ST14", "matriptase", "EMT", and "epithelial-mesenchymal transition", the related genes were searched via Kyoto Encyclopedia of Genes and Genomes (KEGG, Release 97.0), Gene Set Enrichment Analysis, and Ingenuity Pathway Analysis (Qiagen, Hilden, Germany). The literature was also searched for EMT-linked genes in BC (Additional file 3: Table S2). The combined search results were filtered and genes lacking complete profiles in the TCGA database were removed. Finally, 41 genes, inclusive of 25 *ST14*- and 16 EMT-associated genes were selected, respectively.

Least absolute shrinkage and selection operator (LASSO) was then used to reduce feature dimensionality and to select gene signatures for training. This was accomplished using the 'glmnet' package in R (version 4.1-1). The optimal lambda (penalty) was identified as coefficients for each shrunk gene, and only genes with non-zero coefficients were selected for training. Tumor samples were randomly assigned as training and testing datasets at a 7:3 ratio, and each sample was labeled according to GAM-High or GAM-Low. The following six algorithms were employed to assess the accuracy of the classifiers: Logistic regression (LR), K-nearest neighbor, support vector classifier 1 (SVC1, using a linear kernel), SVC2 (using a radial basis function kernel), Gaussian naive Bayes, decision tree (DT), and random forest. The performance of the classifiers was evaluated by the Receiver operating characteristic (ROC) analysis with the area under curve (AUC) value and 10-fold cross validation. The classifier with the best accuracy and performance was chosen for GAM status prediction in the GSE5364 and GSE22820 datasets, which was partially validated by the *ST14* expression in each dataset. The preparation of training, testing datasets, preprocessing, model evaluation, and prediction was conducted using the 'scikit-learn' package (version 0.23.0) in Python (version 3.7.4).

Differential gene expression analysis

BioJupies, which applies limma and a geometrical approach (characteristic direction), was used to identify differentially expressed genes (DEGs) [40–42]. The preprocessed expression data for the 41 genes from the TCGA, GSE5364, and GSE22820 datasets were uploaded. Individual samples in each dataset were divided into GAM-High and GAM-Low.

Significant DEGs with an adjusted p-value < 0.05 were identified and visualized with a Volcano plot.

Gene ontology pathway analysis and network construction

In TCGA and GSE75067, DEGs were also derived for GAM status across the whole transcriptome. DEGs were pooled and Metascape (http://www.metascape.org) was used to assess the overexpression of Gene Ontology categories in biological networks and KEGG pathways.

Association of hormone status (HS) positivity and GAM status

Several studies have reported a correlation between CpG methylation and HS in BC [43, 44]. Therefore, we assessed this hypothesis to determine whether HS positivity is associated with the GAM status and CpG methylation in *ST14* using Chi-square test. Tumors with complete information on HS, i.e., estrogen receptor (ER) or progesterone receptor (PR) expression, from TCGA and GSE75067 were observed for this association, including a total of 889 eligible samples (TCGA, $n = 745$; GSE75067, $n = 144$). CpG probes with potential to classify the HS were identified with a feature selection method combining LASSO and the recursive feature elimination (RFE) algorithm ('caret' package in R, version 6.0–88). For the potential probes, their median β-values were set as the cut points for division into high and low methylation level to observe the distribution of HS.

Statistical analysis

Chi-Square test and the Wilcoxon rank-sum test were used for analyses of categorical and continuous variables, respectively; $p < 0.05$ was considered statistically significant. Kaplan-Meier (KM) analysis with the log-rank test was used to evaluate the distinguishing effect of GAM status on OS. Multivariate Cox model analysis was used to evaluate the impact of GAM on survival. All statistical analyses were conducted in R software (version 3.6.1).

Results

Defining the GAM status of *ST14* and developing a classifier for GAM status

In the KM analysis in Xena browser, there was no association between *ST14* expression and OS ($p = 0.3852$, Fig. 1a), whereas patients with a higher methylation status (average β-value > 0.6605) in *ST14* had significantly better OS ($p = 0.04014$). In addition, a higher *ST14* expression level was identified in breast tumors as compared with that in normal tissues (Wilcoxon's p value < 0.001, Fig. 1b) and in tumors with a lower methylation status (Wilcoxon's p value < 0.001). Taken together, these results suggested that the methylation status of *ST14* was related to its expression and might provide improved survival prediction than gene expression itself.

To confirm this hypothesis, GSE75067 dataset, based on a study assessing the association of methylation patterns with subtypes in BC was used. After removing missing data, 144 samples were included in our analysis. After preprocessing, 34 probes remained and the distribution of β-values across the 34 probes in the TCGA (primary tumors and normal tissues) and GSE75067 datasets showed extremely similar patterns, except for some CpG sites in the gene body (mean differences > 0.125, Wilcoxon's p value < 0.001)(Fig. 1c). β-value is the estimate of methylation level by using the ratio of intensities between methylated and unmethylated probes, with 0 being unmethylated and 1 being fully methylated. Using a median averaged β-value (0.6779) across the 34 probes, GAM was derived and stratified into GAM-High and GAM-Low. We then evaluated the predictive potential of the 41 *ST14*- and EMT-associated genes for GAM status. Using Pearson's correlation, we first observed that XBP1 had the strongest positive correlation with GAM in the TCGA cohort (Fig. 1d, Pearson's coefficient = 0.458, $p = 2.58e\text{-}5$). This higher correlation was also observed in the hierarchical clustering, in which the *XBP1* expression level showed strong clustering with GAM status (Fig. 1e).

Using LASSO, genes with non-zero coefficients were identified (Fig. 1f; Additional file 4: Table S3), which were *ACTA2*, *APC*, *AXL*, *IGFBP7*, *MMP14*, *PLAU*, *PLAUR*, *SNAI1*, *SNAI2*, *SPINT2*, *XBP1*, and *ZEB2*. These 12 genes were then used as the gene signature for training in the TCGA dataset. Among the algorithms applied, LR had the highest accuracy in classification (accuracy = 91.31%), followed by SVC1 (accuracy = 90.25%) and SVC2 (accuracy = 90.25%), and DT had the lowest accuracy (accuracy = 70.65%) (Fig. 1g). ROC analysis showed a high AUC (0.94, 95% confidence interval = 0.806–0.956) for LR. Therefore, LR was chosen as the classifier for GAM status in our study, which was then used for GAM status prediction in the GSE5364 and GSE22820 datasets. These two datasets, based on GPL96 and GPL6480 platforms respectively, contained genomic data for BC and were used for validation. Both validating datasets showed significantly lower *ST14* expression levels in GAM-High, similar to that observed in the TCGA cohort (Fig. 1h), suggesting the feasibility of our classifier.

DMPs between the GAM-High and GAM-Low were identified and are summarized in Additional file 5, 6: Table S4–5. Probes identified to be the DMPs were similar between the TCGA and GSE75067 datasets (Fig. 2a and b). Moreover, 23 and 20 DMPs were found in the gene body in the two datasets, accounting for 88.5 and 83.3% of the overall DMPs, respectively (Fig. 2c). The CpG probe with the greatest difference between groups was identified as cg01497747 in both datasets, with a logFC of 0.161 and 0.235 for TCGA and GSE75067, respectively.

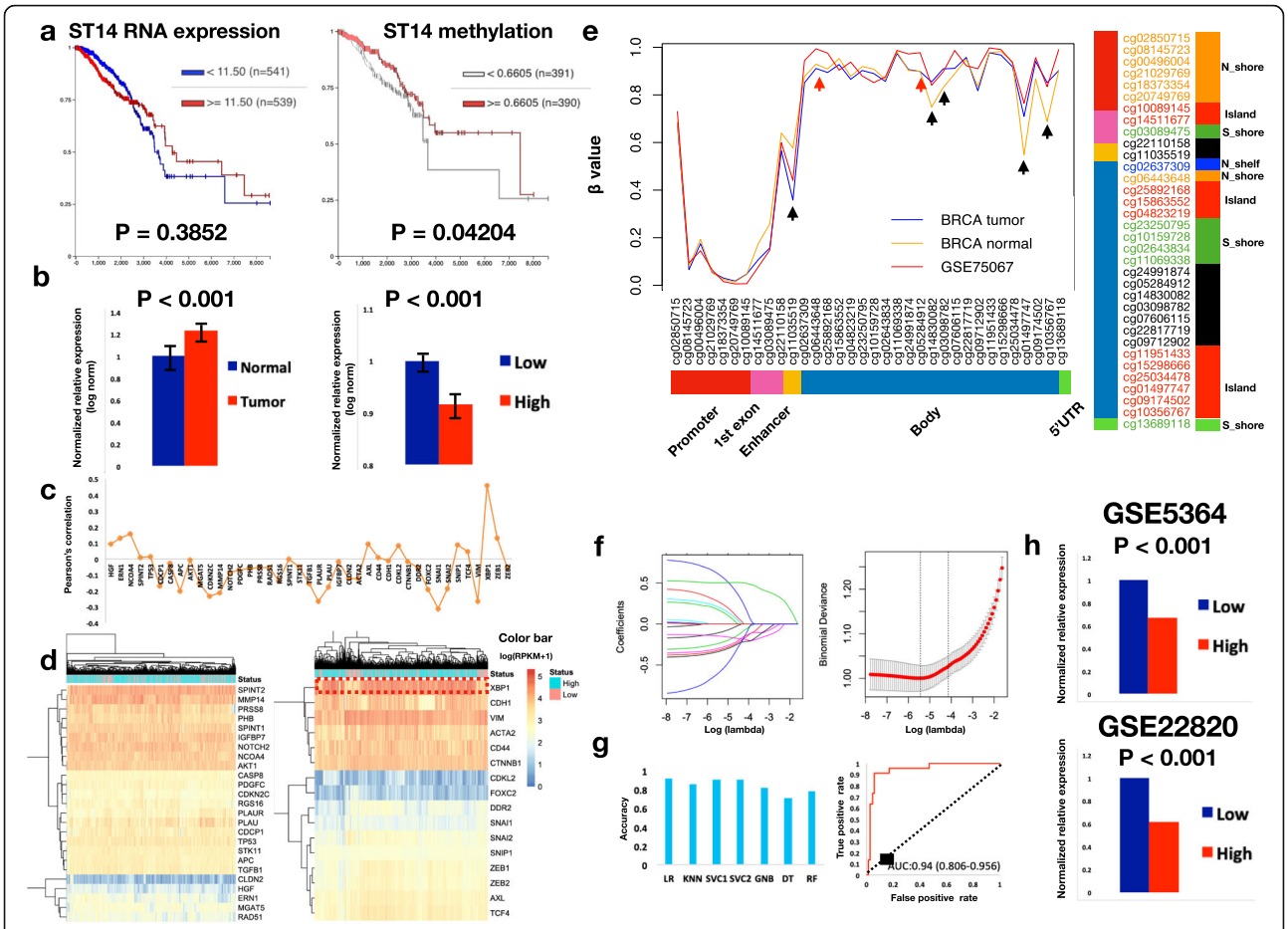

Fig. 1 Characterization of *ST14* methylation and establishment of a classifier for gene-associated methylation (GAM) status. **a** Kaplan-Meier curve for overall survival divided into two groups based on the median *ST14* expression level (left panel, log-rank $p = 0.382$) and methylation level (right panel, log-rank $p = 0.042$) in the TCGA BRCA cohort. The median split for gene expression was based on the RNAseq data (defined as log2(RPKM+ 1)) (11.5). Average β-values across the 40 CpG probes was calculated and their median value across all samples is 0.6605. RNAseq, RNA sequencing; RPKM, reads per kilobase per million mapped reads. **b** Difference in *ST14* gene expression levels between normal and primary tumors in the TCGA BRCA cohort (left panel, Wilcoxon's $p < 0.001$). The difference of *ST14* expression level between high and low methylation status groups (defined by the median methylation value) is shown on the right panel (Wilcoxon's $p < 0.001$). Error bars show the standard deviation. **c** Methylation profile (34 CpG probes) of *ST14* for TCGA tumor (blue) TCGA normal tissue (orange) and GSE75067 (red). The corresponding cgi probes and features are shown in the lower and right panels. Red arrows indicate regions with large differences (Average β-values > 0.125) between TCGA and GSE75067 tumors; black arrows indicate the regions with large differences between TCGA tumor and normal samples (**d**) Pearson's correlation coefficients between genes involved in matriptase-associated or epithelial mesenchymal transition (EMT)-associated pathways and the GAM in TCGA cohort. The highest positive correlation was noted for *XBP1*. **e** Unsupervised hierarchical clustering for matriptase-associated (left panel) and EMT-associated genes (right panel) in the TCGA cohort with GAM status. The gene expression level is based on log2 (RPKM+ 1) transformation of RNAseq data, with the color bar shown in the upper-right corner: high GAM status is in cyan, and low GAM status is in pink. **f** Identification of the optimal lambda value for the least absolute shrinkage and selection operator (LASSO). The left panel depicts the shrinkage of coefficients and the right panel shows the binomial deviance during shrinkage. The optimal lambda was 0.003970773. **g** Assessment of classifier accuracy (left panel). LR, logistic regression; KNN, K-nearest neighbor; SVC1, support vector classifier 1 (using a linear kernel); SVC2, support vector classifier 2 (using a radial basis function kernel); GNB, Gaussian naive Bayes; DT, decision tree; RF, random forest. The highest accuracy was obtained with LR (accuracy = 91.31%). Receiver operating curve for LR with the area under curve value and 95% confidence interval shown in the lower part. **h** Normalized gene expression levels of *ST14* between high and low GAM status groups for GSE5364 (upper panel, $p < 0.001$) and GSE22820 (lower panel, $p < 0.001$). The GAM status was predicted using the classifier constructed by LR

Target gene for the GAM status of *ST14*

Focusing on the aforementioned 41 genes, differential gene expression analysis between GAM-High and -Low identified 18, 18, and 22 DEGs with significantly altered expression in the TCGA, GSE5364, and GSE22820 datasets, respectively (Additional file 7, 8 and 9: Table S6-8S and Fig. 3a). In these three datasets, *XBP1* was found to be up-regulated in the GAM-High, with the largest FC of 1.66 in the GSE5364 dataset. In the TCGA cohort, *PLAUR* was down-regulated in the GAM-High

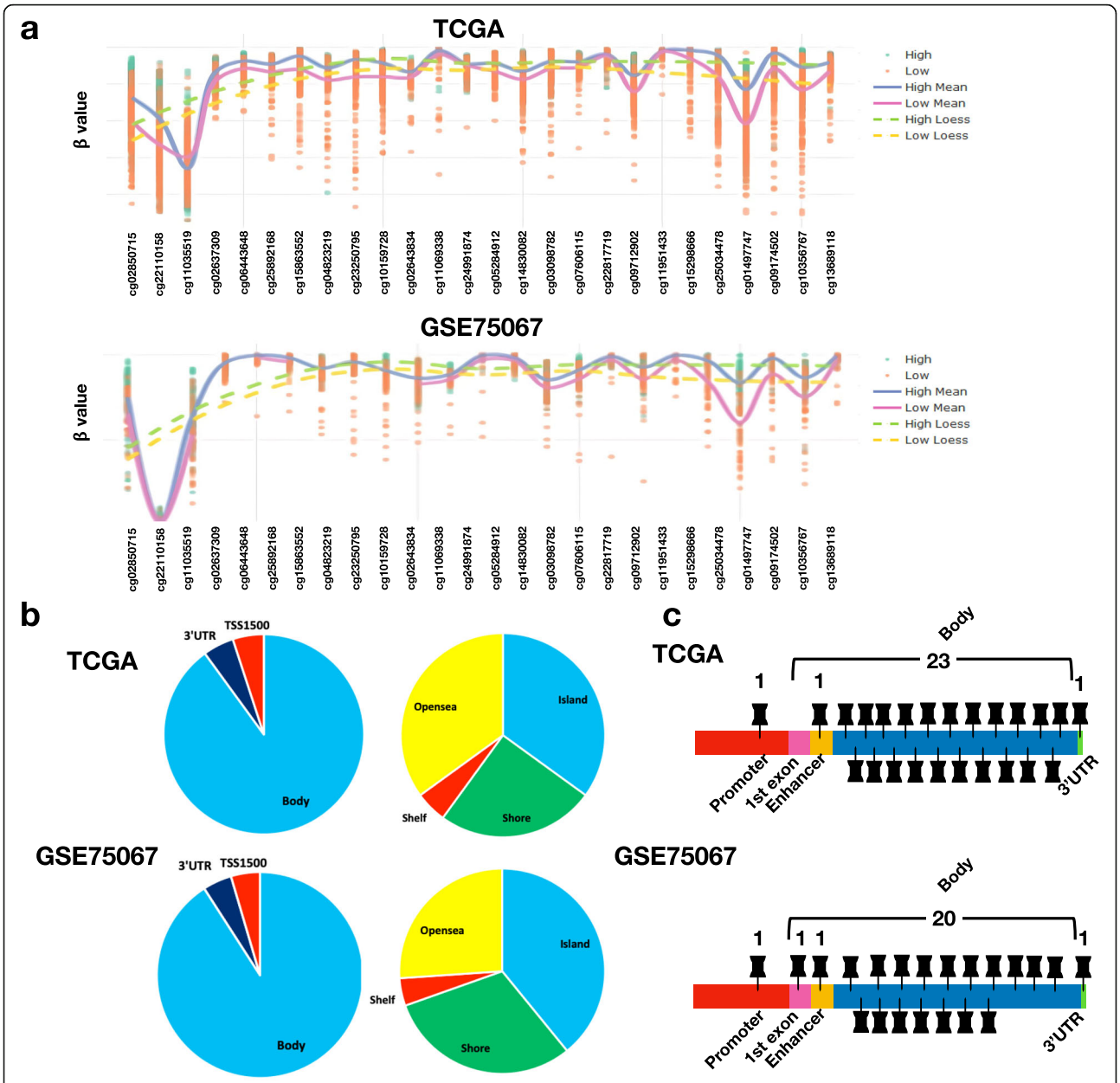

Fig. 2 Differential methylation analysis between tumor samples with high and low gene-associated methylation (GAM) status. **a** Distribution of mean β-values for the differentially methylated probes (DMPs) between high and low GAM status groups in the TCGA and GSE75067 databases (solid line). The dashed line indicates the local regression line fitted for high or low GAM status groups. The purple and blue line connects the mean β-value for each probe in the high and low GAM status group, respectively. All probes shown here are significantly different between the high and low GAM status groups, based on a linear model for microarray data. **b** Pie charts and diagrams for visualization of the DMPs annotated with genetic features and CpG features for TCGA (upper panel) and GSE75067 (lower panel) datasets. **c** Locations of DMPs in *ST14* are indicated

(logFC = – 0.747, –log10p = 21.61), which was also observed in GSE22820. The relative expression levels of these DEGs between the two GAM status groups are shown in Fig. 3b, supporting that the *XBP1* expression level was significantly higher in GAM-High (Wilcoxon's *p* value < 0.001).

Among the correlations between the CpG probes and the expression levels of the 41 genes, the highest abstract correlation was observed between cg11035519 and *XBP1* in the TCGA cohort (Pearson's coefficient = – 0.5919; Fig. 3c). Of note, cg11035519 was annotated as an enhancer in *ST14*. Furthermore, we identified that cg11035519 and cg14830082 had the highest cumulative correlations with EMT-associated genes (overall Pearson's coefficient = 4.2, Fig. 3d), implying the presence of a regulatory site between the enhancer and EMT-associated

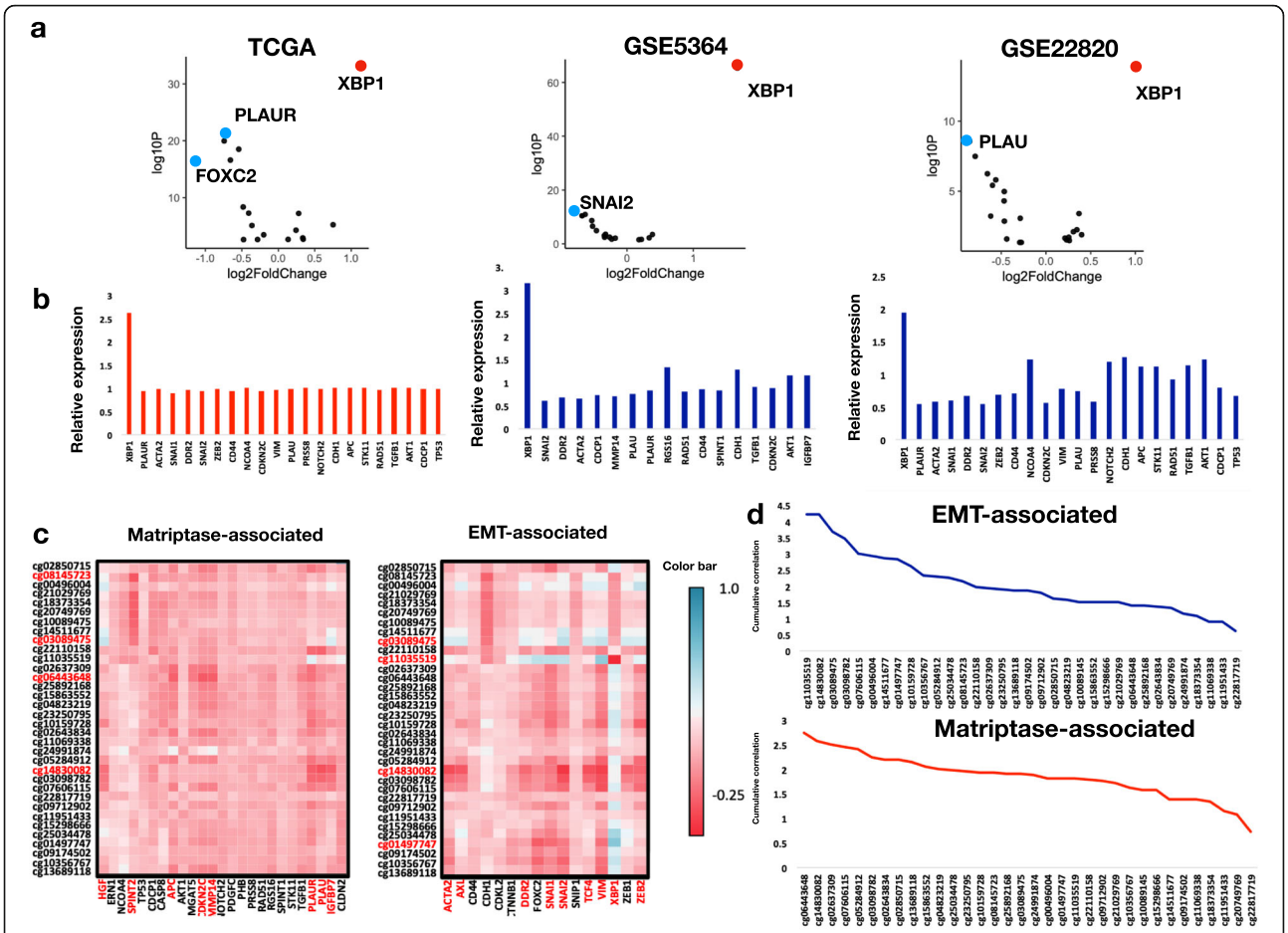

Fig. 3 Differential gene expression and correlation analyses. **a** Volcano plot analyses for genes involved in matriptase-associated and epithelial-mesenchymal transition (EMT)-associated pathways in TCGA, GSE5364, and GSE22820 datasets. In these three datasets, *XBP1* was significantly up-regulated with a high -log10P value in the high GAM status group (red). Conversely, *PLAUR*, *FOXC2*, and *SNAI2* were significantly down-regulated with a high -log10P (blue). **b** Relative expression levels of the differentially expressed genes in the three datasets, with low GAM status as the referencing condition. **c** Correlation heatmap between CpG probes, matriptase-associated genes (left panel), and EMT-associated genes (right panel). The color bar for Pearson's correlation coefficients is shown on the right. Genes and probes with higher correlation (abstract coefficient > 3) are colored in red. **d** Overall correlation between CpG probes, matriptase-associated genes (left panel), and EMT-associated genes (right panel)

genes. By contrast, there was a relatively lower correlation observed between the CpG probes and the matriptase-associated genes, with cg06443648 showing the highest cumulative correlation (overall Pearson's coefficient = 2.7). Further correlation analysis in the TCGA dataset showed a moderately positive correlation between cg11035519 methylation and expression levels of *FOXC1*, *FAM171A*, *RGMA*, *SFRP1*, *CHST3*, and *PTX3* (all coefficients > 0.6, $P < 0.05$, Additional file 10: Table S9).

Association of *ST14* GAM status with molecular features in BC

In the TCGA and GSE75067 datasets, the distribution of PAM50 molecular features was significantly different between GAM-High and GAM-Low (Table 1). Furthermore, hormone receptor status (HS) was also significantly different between the two statuses (Table 2), in which a GAM-High was associated with a positive HS. Moreover, GAM was also significantly higher for the positive HS samples (Fig. 4a). Unsupervised hierarchical clustering showed higher clustering of cg11035519 with the HS in the TCGA and GSE75067 cohorts, with higher methylation corresponding to negative HS (Fig. 4b). Using LASSO to select potential CpG probes, we observed that cg10089145 had the most negative coefficient (− 16.7349), followed by cg11035519 (− 4.8834) (Fig. 4c) in TCGA. The negative influence of cg11035519 was confirmed in the GSE75067 dataset (coefficient = − 9.77947). Probes with a non-zero coefficient were further eliminated with RFE to find the most predictive probes for HS (Fig. 4d). Combining LASSO and RFE, seven CpG probes were identified, which were cg03089475, cg11035519, cg02637309, cg25892168, cg01497747, and cg13689118. These seven probes

Table 1 Association of GAM status with PAM50 molecular feature in TCGA and GSE75067

PAM50	Overall	Low (β ≤ 0.6779)	High (β > 0.6779)	P
TCGA				
Normal	17	12	5	< 0.001
Luminal A	277	36	241	
Luminal B	126	23	103	
Her2	31	20	11	
Basal	86	60	26	
GSE75067				
Normal	11	5	6	0.002
Luminal A	34	10	24	
Luminal B	26	10	16	
Her2	26	15	11	
Basal	38	20	18	

TCGA the Cancer Genome Atlas

corresponded to the enhancer, gene body, and the 3′-UTR in *ST14*. Except for the probe annotated to the enhancer (cg11035519), the other CpG probes with a high β-value harbored higher portion of samples with positive HS in both TCGA and GSE75067 ($p < 0.001$, Fig. 4e).

Survival analysis

The KM plot demonstrated a significant survival difference between the GAM-High and GAM-Low in both TCGA (log-rank test $p = 0.016$) and GSE75067 (log-rank test $p = 0.018$) (Fig. 5). With longer follow-up in the TCGA cohort, GAM-High patients had a median survival (MS) of 122.3 months, compared with 115.7 months in the low GAM status group. Intriguingly, a drop in survival occurred within 3 months for GAM-Low patients in GSE75067, and these patients showed a 15.8-month shorter MS than those with GAM-High (MS: 5.3 vs 21.1 months). This pattern was not observed in the TCGA dataset. These findings suggested that GAM could help risk stratification in BC. Further, GAM status remained the only significant factor after adjusting age, pathological stage, histologic type and PAM50 subtypes in the multivariate Cox model (Table 3).

Table 2 Distribution of HS of TCGA and GSE breast cancer patients grouped by GAM status of *ST14*

	Overall	GAM status		P
		Low (β ≤ 0.6779)	High (β > 0.6779)	
TCGA				
Hormone receptor		$n = 260$	$n = 485$	
Positive	584	144	440	< 0.001
Negative	161	116	45	
GSE75067				
Hormone receptor		$n = 40$	$n = 104$	
Positive	85	8	77	< 0.001
Negative	59	32	27	

HS Hormone status, *GAM* Gene-associated methylation

Gene ontology analysis of differential genes for GAM status

The genes differentially expressed between different GAM statuses were analyzed using Gene Ontology and KEGG analysis. Two thousand nine hundred ninety-eight DEGs were identified (Additional file 11). We found the most enriched pathways are 'NABA MATRISOME ASSOCIATED' and 'extracellular structure organization' (−log10p > 20) (Fig. 6). This finding suggested genes involved in regulation of extracellular microenvironment could alter the GAM status in *ST14*.

Discussion

Identification of new biomarkers for BC can help to optimize the treatment of patients with varying disease spectra. Since cancer phenotypes are determined by both epigenetic modifications and genetic aberrations, investigation of epigenetic variations and integration of these data can improve risk stratification and cancer prognosis. Here, we found that the GAM status in *ST14* was associated with OS in patients with BC. This association was also observed for progression-free survival in the TCGA BRCA primary tumor cohort (Additional file 12: Fig. S2). These observations suggest GAM can potentially serve as a biomarker for OS in patients with BC. Several studies reported that a lower average methylation level across the epigenome in blood DNA was associated with a higher risk of BC [35, 36]. Alternatively, in the present study, lower tissue *ST14*-associated methylation (i.e., GAM status) was associated with a higher risk of death, providing a surrogate for risk evaluation in BC. Furthermore, a higher *ST14* expression level was significantly associated with a low GAM status in the training TCGA and the two validation cohorts. This implies the GAM status could alter the transcription efficiency and that a higher methylation level inhibits gene expression, suggesting that DNAm is more generalized than gene expression.

According to a recent review from Smith et al., the paradigm of promoter DNA methylation as a transcriptional silencing mechanism does not always hold true [45], and hypermethylation-induced transcriptional activation has been documented in a range of cellular changes, including development, malignancy, and metastatic disease. Under some circumstances, methylation of gene promoters facilitates context-dependent transcriptional activation across a range of biological settings. Therefore, gene methylation is regulated differently from its expression and could serve as biomarkers for certain

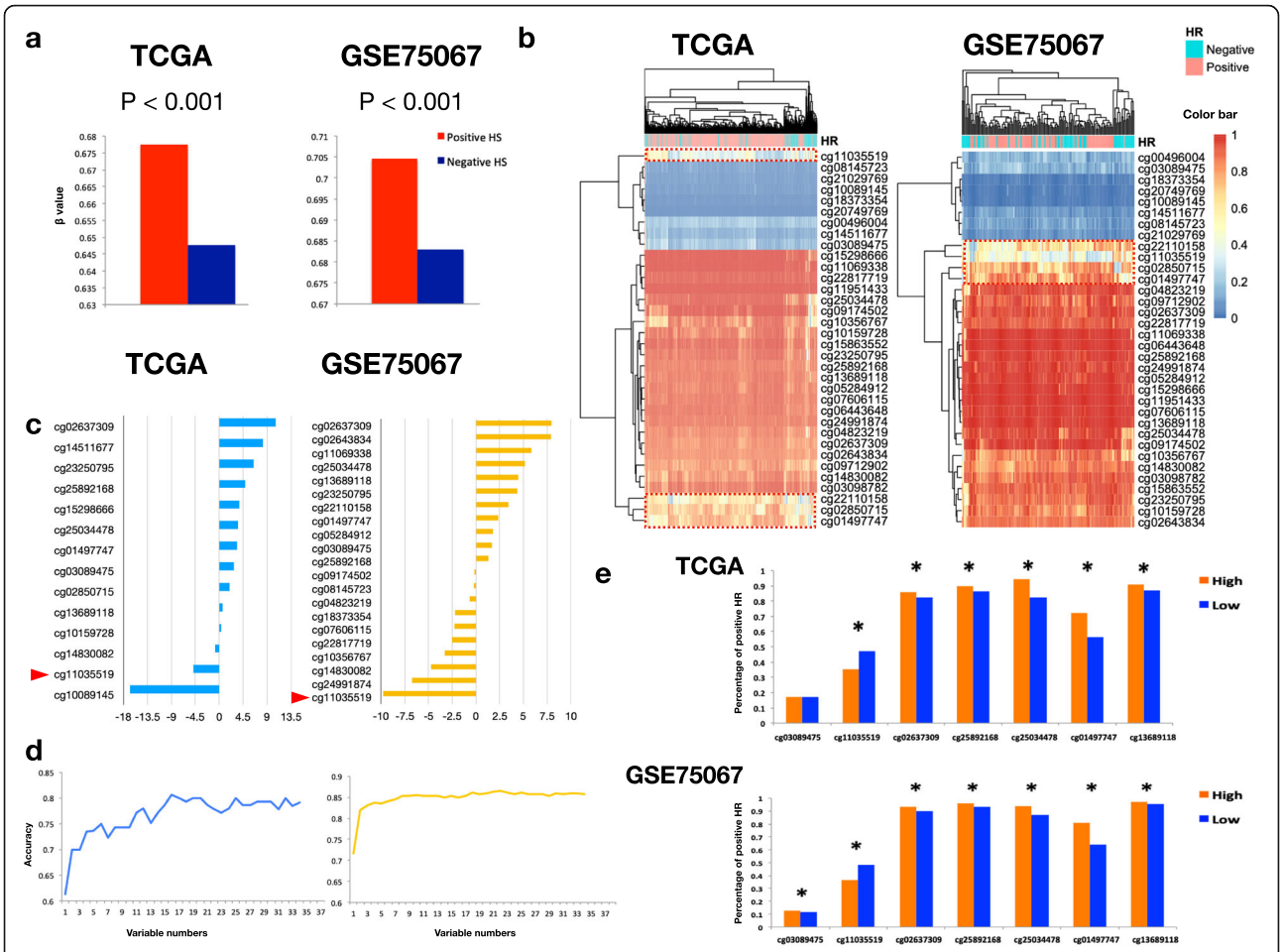

Fig. 4 Association of *ST14* methylation with hormone receptor status (HS) positivity in TCGA and GSE75067. **a** Mean GAM between positive and negative HS. Both datasets showed a significant difference of GAM, with a higher GAM observed in HS-positive breast cancers ($p < 0.001$). **b** Unsupervised hierarchical clustering for CpG probes and the HS. The β-value is indicated by the color bar in the upper-right corner; a positive HS is in pink and negative HS is in cyan. **c** Vertical bar plots for CpG probes with non-zero least absolute shrinkage and selection operator (LASSO) coefficients for TCGA (left panel) and GSE75067 (right panel). **d** Plot of recursive feature elimination (RFE) classification accuracy for the HS with probe numbers for TCGA (left panel) and GSE75067 (right panel). The selected probe set showed the highest accuracy. In TCGA, 16 CpG probes were selected, whereas 22 probes were selected in GSE75067. **e** CpG probes selected by LASSO and RFE in both datasets. The methylation status is defined by the median split for each probe. The percentage of positive HS is plotted against each selected probe with high and low methylation status. Significance of the difference is indicated by a star (Chi-square test $p < 0.05$)

endpoints. One well known example is the promoter methylation of MGMT, not its gene expression correlates well to the efficacy of Temozolomide in glioblastoma multiforme [46]. Although promoter methylation is a major mechanism of gene silencing, additional factors may affect the correlation between MGMT methylation, expression, and patient outcome. Alternative mechanisms, such as post-transcriptional modulation of MGMT by miRNAs or the association of MGMT methylation with IDH mutation or the glioma CpG island methylator phenotype, may explain these inconsistent correlations and different outcomes. Moreover, using DNA methylation as a predictor has a few advantages over other biomarkers. For instance, DNA methylation had a higher stability both in vivo and ex vivo, the requirement of a smaller amount of specimens to obtain enough DNA for analyzing methylation, and higher accuracy. Some study suggested that combinations of DNA methylation as predictors may yield higher sensitivity and specificity than individual DNA methylation. As shown in Fig. 1b in our manuscript, hypermethylation of *ST14* is associated with its lower expression. Therefore, the direct impact could be partly explained by the paradigm repressive role on its gene expression. Since different regulatory mechanisms exist and therefore varying observations in gene expression and methylation detected, survival impact could be different between gene expression and methylation assays.

Yang et al. reported a positive relationship between DNAm in the gene body and gene expression [47].

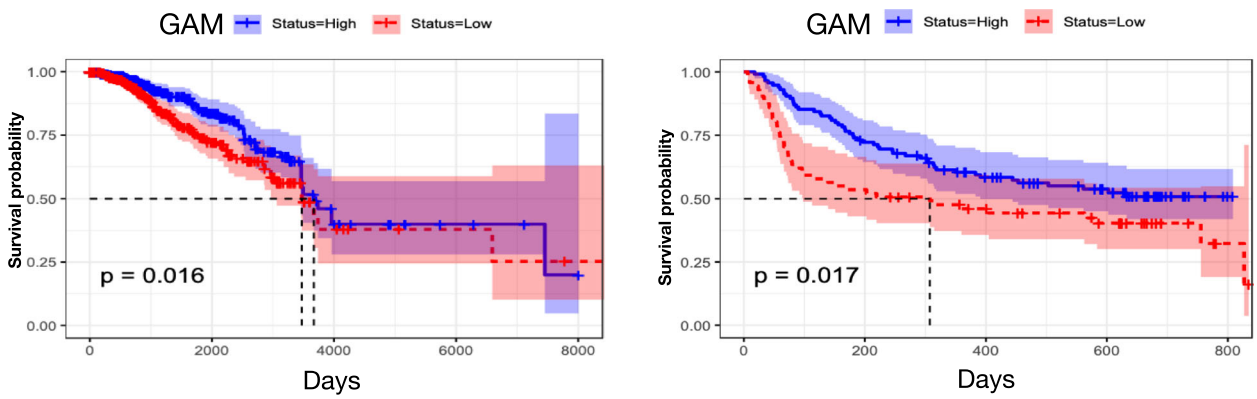

Fig. 5 Kaplan-Meier (KM) survival analysis for overall survival (OS) in TCGA and GSE75067. The samples in the two datasets were divided into high and low gene-associated methylation (GAM) status groups based on the study-defined β-value (0.6779). The KM plots also show areas with 95% confidence intervals

Another study in hepatocellular carcinoma supported this idea and identified that CpG hypermethylation in the gene body region was linked to the upregulated expression of several oncogenes [48]. Similarly, in the present study, hypermethylation in the gene body region was correlated with a higher *ST14* gene expression level in tumor tissues compared with that in normal tissues in the TCGA BRCA cohort (Additional file 13, 14: Fig. S3–4). By contrast, within tumor samples, DMPs with higher methylation in the gene body were found to be linked to a high GAM status, which was in turn associated with a low *ST14* expression level. This finding suggests a threshold methylation level in tumor samples that could further modify the transcription efficiency. *XBP1* was previously identified as a candidate oncogene that is induced in various cancer types. After being cleaved by inositol-requiring enzyme 1α (IRE1), a functional spliced variant of *XBP1* mRNA (*XBP1S*) is formed, which plays a role in the unfolded protein response and autophagy [49–52]. Accumulating evidence supports a direct role of *XBP1* in tumor invasion and metastasis [53, 54]. Moreover, *XBP1* was reported to be activated in triple-negative BC, contributing to tumorigenesis through the HIF-1α pathway [55]. In addition, the level of *XBP1S* expression was higher in basal-like BC cell lines, supporting the adverse effect of this variant [56]. By contrast, levels of unspliced *XBP1* (*XBP1U*) mRNA correlated with *ESR1* mRNA levels in luminal-type BC, and higher levels of *XBP1U* mRNA were associated with better survival in patients with BC [57]. Furthermore, higher ratios of *XBP1S/XBP1U* mRNA level were associated with poor survival. Collectively, these findings suggest that *XBP1* isoforms have distinct functions and that their expression is cell type-specific. Most of the tumor samples in the TGCA cohort were of the ER-positive luminal subtypes (Additional file 15: Fig. S5), implying a higher portion of *XBP1U* in this cohort. The high correlation coefficient between *XBP1* and *ESR1* supports this hypothesis (Pearson's correlation = 0.7754, $p = 1.58e\text{-}10$). In the differential expression analysis, *XBP1* was significantly upregulated in the GAM-High, which corresponded to a more positive HS and better OS. As such, higher *ST14* methylation might be associated with higher *XBP1U* expression, suggesting a novel link between *XBP1* and DNAm.

Table 3 Multivariate analysis of covariates for survival in breast cancer

Covariates	HR	95% CI	*P*
TCGA			
GAM status (ref: GAM-Low)	0.16	0.07–0.37	< 0.001
Age (ref: <= 50)	2.19	1.11–4.35	0.023
Pathological stage (ref: stage I)	2.79	0.86–6.05	0.09
Histologic type (ref: IDC)	0.52	1.11–4.35	0.13
PAM50 (ref: normal)			
Luminal A	0.51	0.10–2.56	0.41
Luminal B	0.74	0.15–3.71	0.71
Basal	0.28	0.05–1.39	0.11
Her2	0.47	0.07–3.06	0.42
GSE75067			
GAM status (ref: GAM-Low)	0.39	0.21–0.74	0.003
Age (ref: <= 50)	2.03	1.25–3.30	0.004
Histologic type (ref: IDC)	0.61	0.29–1.26	0.18
PAM50 (ref: normal)			
Luminal A	1.99	0.43–9.27	0.37
Luminal B	4.73	1.07–20.96	0.04
Basal	4.64	1.05–20.59	0.04
Her2	7.75	1.77–34.1	0.006

HR hazard ratio, *CI* confidence interval, *GAM* gene associated methylation, *IDC* invasive ductal carcinoma

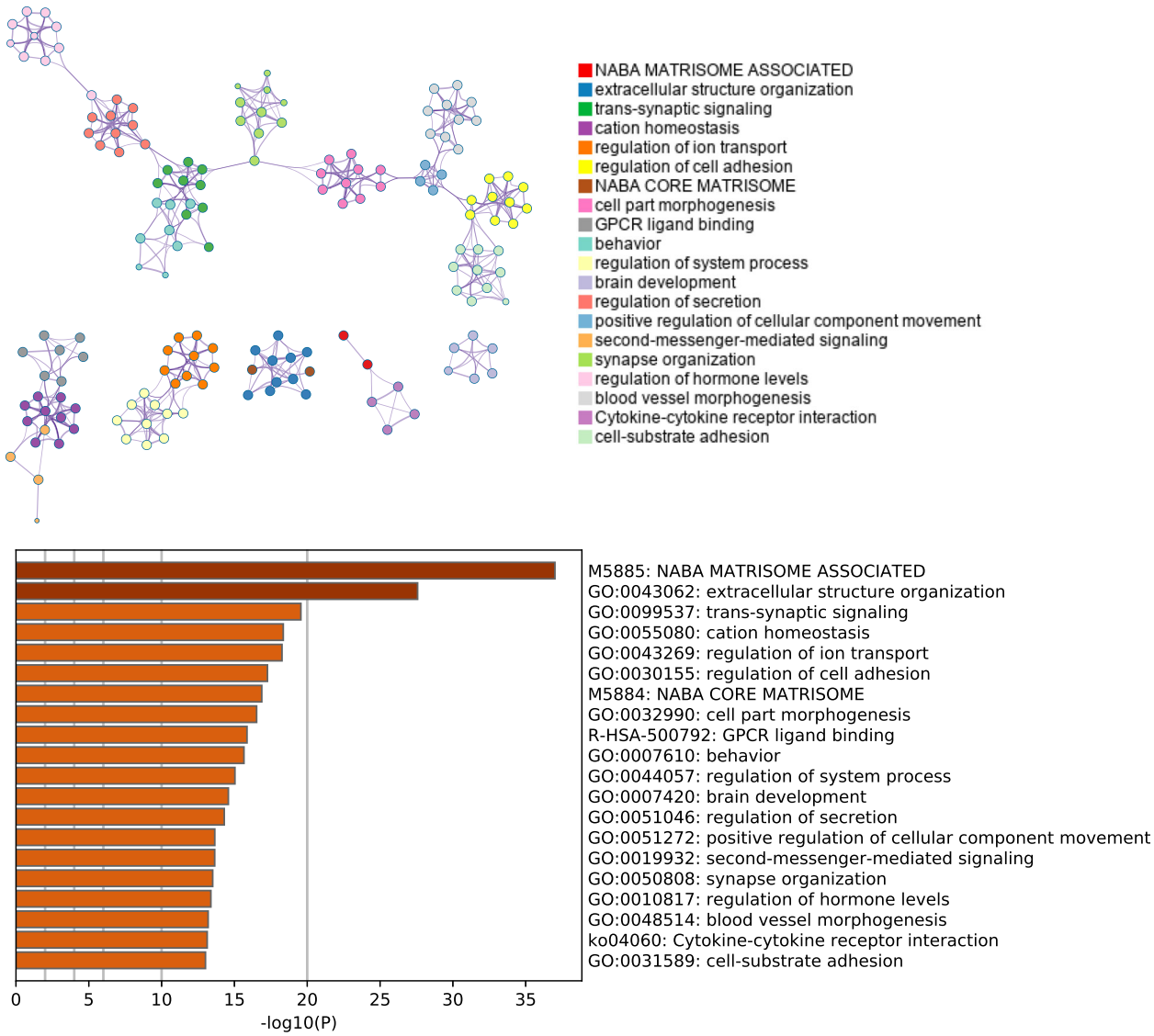

Fig. 6 Gene Ontology and pathway enrichment analysis of genes involved in methylation of *ST14*

Among the DMPs identified, cg11035519 was found to be the most negatively correlated with *XBP1* expression and had the highest overall correlation with EMT-associated genes. cg11035519 was annotated to an enhancer in *ST14*, and is located in the local minimum after the first exon. In the reference genome hg19, this probe is situated in ch11:130,051457–130,051459, which is the intron region of *ST14*. These findings suggest the possible existence of enhancer RNA that could affect the transcription of downstream oncogenes [58]. Of note, genes with mRNA levels positively correlated with the levels of cg11035519 methylation were reported to be induced in basal-like BC [59, 60], supporting a role in downstream activation, and a negative correlation between *XBP1U* and the basal-like histology. Moreover, we found a negative correlation between cg11035519 and the HS, which was not found for the other predictive CpG probes. Taken together, these results suggest that cg11035519 might be independently regulated in *ST14*.

Previous studies have shown that ER-positive BC with a primary origin displayed more hypermethylated loci and a higher overall DNAm than ER-negative or secondary ER-positive breast tumors [61–64]. Moreover, Fackler and colleagues used the HM27 platform to analyze samples from 103 patients with BC, which identified a classifier for HS with 40 CpG probes [62]. Another study demonstrated that hypermethylation of the promoters of several genes could strongly predict the HS in BC [44]. Taken together, these findings suggest that DNAm in a specific set of genes could reflect the HS in BC and

might be used to predict a clinical benefit with hormone therapy. Although DNAm in *ST14* has never been reported to correlate with the HS, several of the CpG probes identified in our study did show strong clustering with the HS and high classifying power. This suggests that the HS is also an epigenomic event, which in turn supports the finding of a different overall DNAm profile between ER-positive and ER-negative tumors. Indeed, in our study, GAM-High tumors were characterized by a higher percentage of positive HS. From a clinical viewpoint, ER/PR-positive BC is associated with a better outcome, supporting our survival findings.

Our study had several limitations. First, the GAM defined in our study could be further optimized via advanced bioinformatic methods, accounting only for probes that are of biological meaning and predictive power. Second, only DNAm in *ST14* was investigated in this study. There might be other potential genes with prognostic power in terms of GAM. Further study could be done to elucidate the interaction network of GAM in BC oncogenes. Third, the validation datasets are limited. Future establishment of validation datasets to confirm the role of GAM is required.

Conclusions

Our study identified that BC with a low GAM status in *ST14* was associated with poorer survival and a negative HS. In the era of personalized medicine, these findings could help refine the epigenomic panel for BC classification and prognostic evaluation. Before clinical use, its epigenetic modulation, relationship with EMT-associated genes, especially *XBP1*, and feasibility as a clinical biomarker require further confirmation. Nevertheless, these results still provide a new perspective on the role of matriptase from the point of epigenetic regulation in BC.

Abbreviations

AUC: Area under curve; BC: Breast cancer; BRCA: Breast carcinoma; DEG: Differentially expressed gene; DMP: Differentially methylated probe; DNAm: DNA methylation; EMT: Epithelial-mesenchymal transition; ER: Estrogen receptor; FC: Fold change; GAM: Gene-associated methylation; HM450K: Human Methylation 450 K; HS: Hormone status; IRE1: Inositol-requiring enzyme 1α; KEGG: Kyoto Encyclopedia of Genes and Genomes; KM: Kaplan-Meier; LASSO: Least absolute shrinkage and selection operator; LR: Logistic regression; MS: Median survival; OS: Overall survival; 3′ UTR: 3′ untranscribed region; PR: Progesterone receptor; RFE: Recursive feature elimination; RNAseq: RNA sequencing; ROC: Receiver operation characteristic; ST14: Suppressor of tumorigenicity 14; SVC: Support vector classifier; SVD: Singular value decomposition; TCGA: The Cancer Genome Atlas; *XBP1S*: Spliced variant of *XBP1* mRNA; *XBP1U*: Unspliced *XBP1*

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-021-08645-3.

Additional file 1: Figure S1. Study flow chart.

Additional file 2: Table S1. Detailed information of 34 probes annotated with *ST14*.

Additional file 3: Table S2. EMT biomarkers in breast cancer.

Additional file 4: Table S3. Coefficient of genes derived from LASSO.

Additional file 5: Table S4. DMPs for TCGA between high and low GAM groups.

Additional file 6: Table S5. DMPs for GSE75067 between high and low GAM groups.

Additional file 7: Table S6. Differential gene expression results for TCGA BRCA cohort.

Additional file 8: Table S7. Differential gene expression results for GSE5364.

Additional file 9: Table S8. Differential gene expression results for GSE22820.

Additional file 10: Table S9. Correlation between cg11035519 and gene expression of *FOXC1*, *FAM171A*, *RGMA*, *SFRP1*, *CHST3*, and *PTX3*.

Additional file 11. DEGs between high and low GAM groups.

Additional file 12: Figure S2. Progression-free survival in BRCA cohort.

Additional file 13: Figure S3. Snapshot of CpG beta value distribution between normal and tumor samples from Xena browser.

Additional file 14: Figure S4. DMPs between normal and tumor tissues in TCGA BRCA cohort.

Additional file 15: Figure S5. Snapshot of the percentage of hormone status for TCGA BRCA cohort.

Acknowledgments

Not applicable.

Authors' contributions

DYH designed the study and performed most of the data mining and processing. WYF, SPC, YJF, and LCH helped with data mining and clinical data interpretation. LCS and CHL helped with manuscript revisions. HWY directed the project. All authors read and approved the final version of the manuscript.

Funding

This study was supported by study project of Tri-Service General Hospital (TSGH-D390 109063) and Ministry of Science and Technology of Taiwan(110-2314-B-016-066-).

Availability of data and materials

The data that supports the findings of this study are available from TCGA (https://xenabrowser.net/) and gene expression omnibus (Accession number: GSE5364, GSE22820 and GSE75067). Further details and other data are available from the corresponding author upon request.

Declarations

Ethics approval and consent to participate

The project was reviewed and approved by the institutional Ethics Committee of Tri-Service General Hospital (IRB number: 1–107–05-016).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interest.

Author details

[1]Department of Radiation Oncology, Tri-Service General Hospital, National Defense Medical Center, No. 325, Chengong Rd., Sec. 2, Neihu, Taipei 114, Taiwan. [2]Department of Radiation Oncology, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan. [3]Institute of Clinical Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan.

Received: 4 March 2021 Accepted: 20 July 2021
Published online: 21 August 2021

References

1. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med. 2002;347(25):1999–2009. https://doi.org/10.1056/NEJMoa021967.
2. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med. 2004;351(27):2817–26. https://doi.org/10.1056/NEJMoa041588.
3. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 2009;27(8):1160–7. https://doi.org/10.1200/JCO.2008.18.1370.
4. Loeb DM, Evron E, Patel CB, Sharma PM, Niranjan B, Buluwela L, et al. Wilms' tumor suppressor gene (WT1) is expressed in primary breast tumors despite tumor-specific promoter methylation. Cancer Res. 2001;61(3):921–5.
5. Mirza S, Sharma G, Prasad CP, Parshad R, Srivastava A, Gupta SD, et al. Promoter hypermethylation of TMS1, BRCA1, ERalpha and PRB in serum and tumor DNA of invasive ductal breast carcinoma patients. Life Sci. 2007;81(4):280–7. https://doi.org/10.1016/j.lfs.2007.05.012.
6. Shargh SA, Sakizli M, Khalaj V, Movafagh A, Yazdi H, Hagigatjou E, et al. Downregulation of E-cadherin expression in breast cancer by promoter hypermethylation and its relation with progression and prognosis of tumor. Med Oncol. 2014;31(11):250. https://doi.org/10.1007/s12032-014-0250-y.
7. Zhu X, Shan L, Wang F, Wang J, Wang F, Shen G, et al. Hypermethylation of BRCA1 gene: implication for prognostic biomarker and therapeutic target in sporadic primary triple-negative breast cancer. Breast Cancer Res Treat. 2015;150(3):479–86. https://doi.org/10.1007/s10549-015-3338-y.
8. Rodriguez-Paredes M, Esteller M. Cancer epigenetics reaches mainstream oncology. Nat Med. 2011;17(3):330–9. https://doi.org/10.1038/nm.2305.
9. Tirado-Magallanes R, Rebbani K, Lim R, Pradhan S, Benoukraf T. Whole genome DNA methylation: beyond genes silencing. Oncotarget. 2017;8(3):5629–37. https://doi.org/10.18632/oncotarget.13562.
10. Raynal NJ, Si J, Taby RF, Gharibyan V, Ahmed S, Jelinek J, et al. DNA methylation does not stably lock gene expression but instead serves as a molecular mark for gene silencing memory. Cancer Res. 2012;72(5):1170–81. https://doi.org/10.1158/0008-5472.CAN-11-3248.
11. Holm K, Hegardt C, Staaf J, Vallon-Christersson J, Jonsson G, Olsson H, et al. Molecular subtypes of breast cancer are associated with characteristic DNA methylation patterns. Breast Cancer Res. 2010;12(3):R36. https://doi.org/10.1186/bcr2590.
12. Stefansson OA, Moran S, Gomez A, Sayols S, Arribas-Jorba C, Sandoval J, et al. A DNA methylation-based definition of biologically distinct breast cancer subtypes. Mol Oncol. 2015;9(3):555–68. https://doi.org/10.1016/j.molonc.2014.10.012.
13. de Almeida BP, Apolonio JD, Binnie A, Castelo-Branco P. Roadmap of DNA methylation in breast cancer identifies novel prognostic biomarkers. BMC Cancer. 2019;19(1):219. https://doi.org/10.1186/s12885-019-5403-0.
14. Tsai CH, Teng CH, Tu YT, Cheng TS, Wu SR, Ko CJ, et al. HAI-2 suppresses the invasive growth and metastasis of prostate cancer through regulation of matriptase. Oncogene. 2014;33(38):4643–52. https://doi.org/10.1038/onc.2013.412.
15. Ko CJ, Huang CC, Lin HY, Juan CP, Lan SW, Shyu HY, et al. Androgen-induced TMPRSS2 activates Matriptase and promotes extracellular matrix degradation, prostate Cancer cell invasion, tumor growth, and metastasis. Cancer Res. 2015;75(14):2949–60. https://doi.org/10.1158/0008-5472.CAN-14-3297.
16. Mukai S, Yorita K, Kawagoe Y, Katayama Y, Nakahara K, Kamibeppu T, et al. Matriptase and MET are prominently expressed at the site of bone metastasis in renal cell carcinoma: immunohistochemical analysis. Hum Cell. 2015;28(1):44–50. https://doi.org/10.1007/s13577-014-0101-3.
17. Uhland K. Matriptase and its putative role in cancer. Cell Mol Life Sci. 2006;63(24):2968–78. https://doi.org/10.1007/s00018-006-6298-x.
18. Cheng H, Fukushima T, Takahashi N, Tanaka H, Kataoka H. Hepatocyte growth factor activator inhibitor type 1 regulates epithelial to mesenchymal transition through membrane-bound serine proteinases. Cancer Res. 2009;69(5):1828–35. https://doi.org/10.1158/0008-5472.CAN-08-3728.
19. Lee HS, Kim C, Kim SB, Kim MG, Park D. Epithin, a target of transforming growth factor-beta signaling, mediates epithelial-mesenchymal transition. Biochem Biophys Res Commun. 2010;395(4):553–9. https://doi.org/10.1016/j.bbrc.2010.04.065.
20. Kim S, Yang JW, Kim C, Kim MG. Impact of suppression of tumorigenicity 14 (ST14)/serine protease 14 (Prss14) expression analysis on the prognosis and management of estrogen receptor negative breast cancer. Oncotarget. 2016;7(23):34643–63. https://doi.org/10.18632/oncotarget.9155.
21. Kang JY, Dolled-Filhart M, Ocal IT, Singh B, Lin CY, Dickson RB, et al. Tissue microarray analysis of hepatocyte growth factor/met pathway components reveals a role for met, matriptase, and hepatocyte growth factor activator inhibitor 1 in the progression of node-negative breast cancer. Cancer Res. 2003;63(5):1101–5.
22. Lee JW, Yong Song S, Choi JJ, Lee SJ, Kim BG, Park CS, et al. Increased expression of matriptase is associated with histopathologic grades of cervical neoplasia. Hum Pathol. 2005;36(6):626–33. https://doi.org/10.1016/j.humpath.2005.03.003.
23. Riddick AC, Shukla CJ, Pennington CJ, Bass R, Nuttall RK, Hogan A, et al. Identification of degradome components associated with prostate cancer progression by expression analysis of human prostatic tissues. Br J Cancer. 2005;92(12):2171–80. https://doi.org/10.1038/sj.bjc.6602630.
24. Tanimoto H, Shigemasa K, Tian X, Gu L, Beard JB, Sawasaki T, et al. Transmembrane serine protease TADG-15 (ST14/Matriptase/MT-SP1): expression and prognostic value in ovarian cancer. Br J Cancer. 2005;92(2):278–83. https://doi.org/10.1038/sj.bjc.6602320.
25. Saleem M, Adhami VM, Zhong W, Longley BJ, Lin CY, Dickson RB, et al. A novel biomarker for staging human prostate adenocarcinoma: overexpression of matriptase with concomitant loss of its inhibitor, hepatocyte growth factor activator inhibitor-1. Cancer Epidemiol Biomark Prev. 2006;15(2):217–27. https://doi.org/10.1158/1055-9965.EPI-05-0737.
26. Nakamura K, Hongo A, Kodama J, Abarzua F, Nasu Y, Kumon H, et al. Expression of matriptase and clinical outcome of human endometrial cancer. Anticancer Res. 2009;29(5):1685–90.
27. Zoratti GL, Tanabe LM, Varela FA, Murray AS, Bergum C, Colombo E, et al. Targeting matriptase in breast cancer abrogates tumour progression via impairment of stromal-epithelial growth factor signalling. Nat Commun. 2015;6(1):6776. https://doi.org/10.1038/ncomms7776.
28. Zoratti GL, Tanabe LM, Hyland TE, Duhaime MJ, Colombo E, Leduc R, et al. Matriptase regulates c-met mediated proliferation and invasion in inflammatory breast cancer. Oncotarget. 2016;7(36):58162–73. https://doi.org/10.18632/oncotarget.11262.
29. Volkow ND, Koroshetz WJ. The role of neurologists in tackling the opioid epidemic. Nat Rev Neurol. 2019;15(5):301–5. https://doi.org/10.1038/s41582-019-0146-8.
30. Sato N, Fukushima N, Hruban RH, Goggins M. CpG island methylation profile of pancreatic intraepithelial neoplasia. Mod Pathol. 2008;21(3):238–44. https://doi.org/10.1038/modpathol.3800991.
31. Tian Y, Morris TJ, Webster AP, Yang Z, Beck S, Feber A, et al. ChAMP: updated methylation analysis pipeline for Illumina BeadChips. Bioinformatics. 2017;33(24):3982–4. https://doi.org/10.1093/bioinformatics/btx513.
32. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics. 2013;29(2):189–96. https://doi.org/10.1093/bioinformatics/bts680.
33. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Gayther SA, Apostolidou S, et al. An epigenetic signature in peripheral blood predicts active ovarian cancer. PLoS One. 2009;4(12):e8274. https://doi.org/10.1371/journal.pone.0008274.
34. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics. 2010;11(1):587. https://doi.org/10.1186/1471-2105-11-587.
35. Severi G, Southey MC, English DR, Jung CH, Lonie A, McLean C, et al. Epigenome-wide methylation in DNA from peripheral blood as a marker of risk for breast cancer. Breast Cancer Res Treat. 2014;148(3):665–73. https://doi.org/10.1007/s10549-014-3209-y.
36. van Veldhoven K, Polidoro S, Baglietto L, Severi G, Sacerdote C, Panico S, et al. Epigenome-wide association study reveals decreased average methylation levels years before breast cancer diagnosis. Clin Epigenetics. 2015;7(1):67. https://doi.org/10.1186/s13148-015-0104-2.
37. Holm K, Staaf J, Lauss M, Aine M, Lindgren D, Bendahl PO, et al. An integrated genomics analysis of epigenetic subtypes in human breast tumors links DNA methylation patterns to chromatin states in normal mammary cells. Breast Cancer Res. 2016;18(1):27. https://doi.org/10.1186/s13058-016-0685-5.

38. Choi SK, Pandiyan K, Eun JW, Yang X, Hong SH, Nam SW, et al. Epigenetic landscape change analysis during human EMT sheds light on a key EMT mediator TRIM29. Oncotarget. 2017;8(58):98322–35. https://doi.org/10.18632/oncotarget.21681.
39. Pistore C, Giannoni E, Colangelo T, Rizzo F, Magnani E, Muccillo L, et al. DNA methylation variations are required for epithelial-to-mesenchymal transition induced by cancer-associated fibroblasts in prostate cancer cells. Oncogene. 2017;36(40):5551–66. https://doi.org/10.1038/onc.2017.159.
40. Clark NR, Hu KS, Feldmann AS, Kou Y, Chen EY, Duan Q, et al. The characteristic direction: a geometrical approach to identify differentially expressed genes. BMC Bioinformatics. 2014;15(1):79. https://doi.org/10.1186/1471-2105-15-79.
41. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W. Smyth GK: limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47. https://doi.org/10.1093/nar/gkv007.
42. Torre D, Lachmann A, Ma'ayan A. BioJupies: automated generation of interactive notebooks for RNA-Seq data analysis in the cloud. Cell Syst. 2018;7(5):556–561 e553. https://doi.org/10.1016/j.cels.2018.10.007.
43. Kajabova V, Smolkova B, Zmetakova I, Sebova K, Krivulcik T, Bella V, et al. RASSF1A promoter methylation levels positively correlate with estrogen receptor expression in breast Cancer patients. Transl Oncol. 2013;6(3):297–304. https://doi.org/10.1593/tlo.13244.
44. Benevolenskaya EV, Islam AB, Ahsan H, Kibriya MG, Jasmine F, Wolff B, et al. DNA methylation and hormone receptor status in breast cancer. Clin Epigenetics. 2016;8(1):17. https://doi.org/10.1186/s13148-016-0184-7.
45. Smith J, Sen S, Weeks RJ, Eccles MR, Chatterjee A. Promoter DNA Hypermethylation and paradoxical gene activation. Trends Cancer. 2020;6(5):392–406. https://doi.org/10.1016/j.trecan.2020.02.007.
46. Butler M, Pongor L, Su YT, Xi L, Raffeld M, Quezado M, et al. MGMT status as a clinical biomarker in glioblastoma. Trends Cancer. 2020;6(5):380–91. https://doi.org/10.1016/j.trecan.2020.02.010.
47. Yang X, Han H, De Carvalho DD, Lay FD, Jones PA, Liang G. Gene body methylation can alter gene expression and is a therapeutic target in cancer. Cancer Cell. 2014;26(4):577–90. https://doi.org/10.1016/j.ccr.2014.07.028.
48. Arechederra M, Daian F, Yim A, Bazai SK, Richelme S, Dono R, et al. Hypermethylation of gene body CpG islands predicts high dosage of functional oncogenes in liver cancer. Nat Commun. 2018;9(1):3164. https://doi.org/10.1038/s41467-018-05550-5.
49. Yoshida H, Matsui T, Yamamoto A, Okada T, Mori K. XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. Cell. 2001;107(7):881–91. https://doi.org/10.1016/S0092-8674(01)00611-0.
50. Vidal RL, Hetz C. Unspliced XBP1 controls autophagy through FoxO1. Cell Res. 2013;23(4):463–4. https://doi.org/10.1038/cr.2013.9.
51. Bright MD, Itzhak DN, Wardell CP, Morgan GJ, Davies FE. Cleavage of BLOC1S1 mRNA by IRE1 is sequence specific, temporally separate from XBP1 splicing, and dispensable for cell viability under acute endoplasmic reticulum stress. Mol Cell Biol. 2015;35(12):2186–202. https://doi.org/10.1128/MCB.00013-15.
52. Wang Y, Xing P, Cui W, Wang W, Cui Y, Ying G, et al. Acute endoplasmic reticulum stress-independent unconventional splicing of XBP1 mRNA in the nucleus of mammalian cells. Int J Mol Sci. 2015;16(6):13302–21. https://doi.org/10.3390/ijms160613302.
53. Mhaidat NM, Alzoubi KH, Abushbak A. X-box binding protein 1 (XBP-1) enhances colorectal cancer cell invasion. J Chemother. 2015;27(3):167–73. https://doi.org/10.1179/1973947815Y.0000000006.
54. Sun Y, Jiang F, Pan Y, Chen X, Chen J, Wang Y, et al. XBP1 promotes tumor invasion and is associated with poor prognosis in oral squamous cell carcinoma. Oncol Rep. 2018;40(2):988–98. https://doi.org/10.3892/or.2018.6498.
55. Chen X, Iliopoulos D, Zhang Q, Tang Q, Greenblatt MB, Hatziapostolou M, et al. XBP1 promotes triple-negative breast cancer by controlling the HIF1alpha pathway. Nature. 2014;508(7494):103–7. https://doi.org/10.1038/nature13119.
56. McCarthy N. Breast cancer: hypoxia and XBP1S. Nat Rev Cancer. 2014;14(5):295. https://doi.org/10.1038/nrc3733.
57. Davies MP, Barraclough DL, Stewart C, Joyce KA, Eccles RM, Barraclough R, et al. Expression and splicing of the unfolded protein response gene XBP-1 are significantly associated with clinical outcome of endocrine-treated breast cancer. Int J Cancer. 2008;123(1):85–8. https://doi.org/10.1002/ijc.23479.
58. Melo CA, Drost J, Wijchers PJ, van de Werken H, de Wit E, Oude Vrielink JA, et al. eRNAs are required for p53-dependent enhancer activity and gene transcription. Mol Cell. 2013;49(3):524–35. https://doi.org/10.1016/j.molcel.2012.11.021.
59. Han B, Bhowmick N, Qu Y, Chung S, Giuliano AE, Cui X. FOXC1: an emerging marker and therapeutic target for cancer. Oncogene. 2017;36(28):3957–63. https://doi.org/10.1038/onc.2017.48.
60. Sanawar R, Mohan Dan V, Santhoshkumar TR, Kumar R, Pillai MR. Estrogen receptor-alpha regulation of microRNA-590 targets FAM171A1-a modifier of breast cancer invasiveness. Oncogenesis. 2019;8(1):5. https://doi.org/10.1038/s41389-018-0113-z.
61. Bediaga NG, Acha-Sagredo A, Guerra I, Viguri A, Albaina C, Ruiz Diaz I, et al. DNA methylation epigenotypes in breast cancer molecular subtypes. Breast Cancer Res. 2010;12(5):R77. https://doi.org/10.1186/bcr2721.
62. Fackler MJ, Umbricht CB, Williams D, Argani P, Cruz LA, Merino VF, et al. Genome-wide methylation analysis identifies genes specific to breast cancer hormone receptor status and risk of recurrence. Cancer Res. 2011;71(19):6195–207. https://doi.org/10.1158/0008-5472.CAN-11-1630.
63. Figueroa JD, Yang H, Garcia-Closas M, Davis S, Meltzer P, Lissowska J, et al. Integrated analysis of DNA methylation, immunohistochemistry and mRNA expression, data identifies a methylation expression index (MEI) robustly associated with survival of ER-positive breast cancer patients. Breast Cancer Res Treat. 2015;150(2):457–66. https://doi.org/10.1007/s10549-015-3314-6.
64. Williams KE, Jawale RM, Schneider SS, Otis CN, Pentecost BT, Arcaro KF. DNA methylation in breast cancers: differences based on estrogen receptor status and recurrence. J Cell Biochem. 2019;120(1):738–55. https://doi.org/10.1002/jcb.27431.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC