World Cancer Research Journal WCRJ 2022; 9: e2244

A PREDICTED TWO TO TANGO INTERACTION BETWEEN MUC1 AND SOX2: A UNIQUE OPPORTUNITY TO TARGET CANCER STEM CELLS

A. DUTTA, S. PAUL

Division of Molecular Medicine, Bose Institute, Kolkata, India

Abstract – *Objective:* *Cancer stem cell (CSC) genes are emerging as cancer therapeutic targets. MUC1 is a cell surface-associated glycoprotein that is aberrantly over-expressed in >60% of human cancers including breast cancer. Recent findings have also indicated significant upregulation of MUC1 in cancer cells. Differential glycosylation of MUC1 in cancer cells as compared to normal cells makes it a lucrative target to precisely identify and attack cancer cells. Many MUC1 targeting antibodies have been developed by various groups to target MUC1-overexpressing cells.*

Materials and Methods: *We analyzed several publicly available datasets to recognize the factors controlling MUC1 expression in CSCs and also analyze the possible regulation of MUC1 expression in CSCs.*

Results: *In this meta-analysis, we explore the positive correlation between CSC signature genes and MUC1 and further extrapolate the findings to speculate about the differential methylation pattern of the MUC1 core promoter region. Recent findings showed that SOX2 binding sites might undergo passive demethylation and SOX2 binding motifs can attract demethylases. We also predicted a SOX2 binding region on the MUC1 core promoter which further strengthened our analysis to explore new insights on MUC1 and SOX2 interplay in CSCs.*

Conclusions: *The need of the hour is to unveil the detailed mechanistic crosstalk between MUC1 and SOX2 which might give us a new direction to target the root cause of cancer relapse by disarming the CSCs.*

KEYWORDS: *MUC1, SOX2, Cancer stem cells, Targeting cancer stem cells.*

ABBREVIATIONS: *CSCs- Cancer stem cells; TCGA-Tumor Breast Invasive Carcinoma; MUC1-Mucin 1; VNTR- Variable number tandem repeat.*

INTRODUCTION

Cancer stem cells (CSCs) are a unique subpopulation of cells located within the tumor bulk and they closely resemble the tissue-resident adult stem cell repertoire. It is suspected that CSCs are the cause for tumor initiation, progression, and relapse. CSCs can be distinguished from other tumor cells by the asymmetry of their cell division and high expression level of stemness-associate and drug-resistance-related genes[1]. CSCs are also known to have a slower rate of cell cycle and are termed as "quiescent cells". This unique property also protects them from insults induced by cytotoxic drugs thus making them resistant from most chemotherapeutic drugs[2]. Mucin 1

Corresponding Author: Abhishek Dutta, PhD; e-mail: adutta953@yahoo.in

(MUC1; also known as episialin, PEM, H23Ag, EMA, CA15-3, and MCA) has been demonstrated to participate in the maintenance, tumorigenicity, glycosylation, pluripotency and metastasis of CSCs[3]. MUC1 is a single-pass type I transmembrane protein which is heavily glycosylated in the extracellular domain that extends up to 200–500 nm from the cell surface[3,4]. The extracellular region of MUC1 contains VNTR regions that are heavily glycosylated and provide protection to the underlying epithelia in healthy tissues, creating a physical barrier and imparting an anti-adhesive property, limiting accessibility and preventing pathogenic colonization[5]. Aberrantly glycosylated MUC1 is overexpressed in most human epithelial cancers and has gained remarkable attention as an oncogenic molecule. MUC1 contributes to multiple hallmarks of the cancer stem cell, including induction of epithelial-mesenchymal transition (EMT) and stemness[3,6,7]. Studies have demonstrated that MUC1 induces CSC markers (BMI1, ALDH1, FOXA1, LIN28B) and the OCT4, SOX2, and NANOG pluripotency factors[8,9]. Given the multifaceted functions of MUC1 in CSCs, it is imperative to determine which factors in CSCs drive MUC1 expression. Our *in-silico* analysis has indicated a high expression of MUC1 in tumor patients which was also found to be closely associated with the stage of the disease. Further analysis also revealed a close association of MUC1 with the CSC signature gene and interestingly we also discovered that the MUC1 promoter region is hypomethylated in CSCs suggesting a possible explanation of higher MUC1 expression in CSCs. The present work demonstrates the close association between MUC1 and CSCs and the significant role of MUC1 in maintaining the unique nature of CSCs and which can be exploited in the future to efficiently target CSCs. SOX2 is known as the undifferentiated cell marker and is known to be involved in the maintenance of cancer stem cells (CSCs). Mechanistically, SOX2 like MUC1 promotes essential hallmarks of CSCs like survival, proliferation, invasion/metastasis, cancer stemness, and drug resistance. The fact that MUC1 and SOX2 regulate similar functional properties in CSCs, lead us to investigate any possible interaction between the two. We predicted a previously unknown binding of SOX2 on MUC1 promoter which may drive MUC1 expression in CSCs. The interaction between MUC1 and SOX2 needs to be further studied via *in-vitro* and *in-vivo* experiments. Understanding the crosstalk between MUC1 and SOX2 might give us a new direction to target the root cause of cancer relapse by disarming the CSCs.

MATERIALS AND METHODS

Search Strategy

The Web of Science, PubMed, Embase, were searched for studies on MUC1 expression and prognosis of breast cancer through June 2020. The pooled hazard ratios (HRs) with 95% confidence intervals (95% CIs) were calculated to evaluate the prognostic and clinicopathological value of MUC1 expression in breast cancer. Studies were selected using following key terms: (MUC OR MUC-1 OR MUC1) AND (Breast cancer OR Breast cancer stem cells) AND (prognosis OR prognostic OR outcome OR mortality OR survival). The references of manuscript were also examined to confirm potential studies. AD and SP conducted the search and assessed the eligibility of studies independently.

Inclusion and exclusion criteria

Inclusion criteria:

1. The patients were diagnosed with breast cancer according to pathological findings;
2. Clinicopathological parameters, MUC1 expression, and survival rate were investigated;
3. Values of hazard ratio (HR) with 95% confidence interval (CI) were calculated; and
4. Publication in English.

Exclusion criteria:

1. Repeated publication of data or poor-quality data, lacking raw data, or presenting incomplete information; and
2. Review articles, case reports, conference abstracts, commentary, or letters to editors. The most recent paper was selected when several studies were published on the same trial.

R2: Gene Selector - Genomic Analysis and Visualization Platform

Tumor Breast Invasive Carcinoma - TCGA - 528 - custom - tcgaovag1 dataset was selected using R2 gene selector (https://hgserver1.amc.nl/cgi-bin/r2/main.cgi) to correlate between MUC1 (GeneID: 4582) and KLF4 (GeneID: 9314) and found out their correlation coefficient, i.e, r=0.367, MUC1 (GeneID: 4582) and ALDH1A1 (GeneID: 216) with r=0.28, MUC1 (GeneID: 4582) and SOX2 (GeneID: 6657) with r=0.33, MUC1 (GeneID: 4582) and NANOG (GeneID: 79923) with r=0.29.

GEPIA (Gene Expression Profiling Interactive Analysis)

The bar plot in GEPIA (http://gepia.cancer-pku.cn/) depicts the gene expression profile across all tumor samples and paired normal tissues[10]. The median expression of certain tumor types or normal tissue is represented by the height of the bar. MUC1 (Ensembl ID: ENSG00000185499.16) is found to be overexpressed in BRCA mutated breast tumor as compared to their normal counterpart. The median gene expression in BRCA tumor is 419.67 and in normal is 76.37. The dot plot represents transcripts per million across all normal and paired tumor tissues, for BRCA mutated category, the number of the normal tissue sample, n=291, and n=1085 for tumor samples.

UALCAN-TCGA analysis

The expression of MUC1 was analyzed in http://ualcan.path.uab.edu/analysis.html[11] BRCA based sample types for the sample size of normal tissues=114 and tumor tissues=1097. The expression of MUC1 was analyzed in BRCA based on the patient's age divided into subgroups of 21-40 years, 41-60 years, 61-80 years, and 81-100 years. The expression of MUC1 was analyzed in BRCA based on individual cancer stages including stage 1 (183 patients), stage 2 (615 patients), stage 3 (247 patients), and stage 4 (20 patients). The promoter methylation level of MUC1 in BRCA1 is analyzed by plotting the beta value with the TCGA samples in various stages of cancer progression. The number of TCGA samples in case of normal is 97, stage 1 is 127, stage 2 is 442, stage 3 is 200, and stage 4 is 11.

R2: Kaplan-Meier Scanner

Using https://hgserver1.amc.nl/cgi-bin/r2/main.cgi[12], tumor breast Zhang-136- MAS5.0-u133a dataset for MUC1 (213693_s_at) with GeneID:4582, we analyzed the relapse-free survival (RFS) probability in y-axis with follow up in months in the x-axis. High MUC1 expression in patients has low RFS than low MUC1 expression patients. Using Tumor breast (MDC) Bertucci-266-MAS5.0-u133p2 dataset for MUC1 (211695_x_at) with GeneID:4582, we analyzed the disease-free survival (DFS). Using the Tumor Breast Bergh-159-MAS5.0-u133a dataset for MUC1 (211695_x_at) with GeneID:4582, we analyzed the overall survival probability.

Cancer Sea (Cancer Single-Cell Atlas)

Using http://biocc.hrbmu.edu.cn/CancerSEA/home.jsp[13], the correlation plot for the input or query gene name MUC1 is given to find out its functional relevance in the BRCA subtype. Utilizing datasets GSE77308, GSE75688, and GSE75367 we found out the correlation coefficient r, between MUC1 and stemness is r=0.44, between MUC1 and metastasis, is r=0.33 and MUC1 and invasion is r=0.41, between MUC1 and proliferation is r=-0.38.

Promoter scanning and methylation status detection analysis

Using EPD (The Eukaryotic promoter database (https://epd.epfl.ch//index.php) human MUC1 promoter sequence is obtained (-499bp to +100bp)[14]. Using the SMART app (http://www.bioinfo-zs.com/smartapp/)[15] analysis tool, the CpG-aggregated methylation value across all samples was analyzed. Using the JASPAR database (http://jaspar.genereg.net/)[16] the putative binding sites and scores (6.26 and 6.13) for SOX2 (Matrix ID: MA0143.4) on MUC1 promoter are detected.

STRING analysis

STRING (https://string-db.org/) is a database of known and predicted protein-protein interactions. We analyzed the interacting partners of MUC1 via STRING analysis.

Modelling the SOX2-MUC1 complex

The molecular models of the SOX2-MUC1 promoter DNA complexes were prepared using HDOCK web server available at http://hdock.phys.hust.edu.cn/ (https://academic.oup.com/nar/article/45/W1/W365/3829194)[17]. For the convenience of the modelling, the core promoter of MUC1 was modelled. For the binding region, 5'GGATAATGAGT3' strand was used. Then the SOX2 protein was docked separately into the promoter binding regions. The docking server yielded various possible models and their relative reliability is expressed with a score for each of them; the final models are further grouped into clusters based on a fraction of common intermolecular contacts. Those clusters are further ranked based on an average score of the top 4 models of each cluster. In this work, the top-ranked clusters were considered for further use.

University of California, Santa Cruz (UCSC) Cancer Genomics Browser analysis

The UCSC Cancer Genomics Browser (http://xena.ucsc.edu/) was searched and dataset TCGA Breast Cancer was used to verify and analyze the heatmap of MUC1 expression and the expression of ALDH1A1[18].

Statistical Analysis

The relationship between MUC1 expression and features of tumor progression was analyzed using CancerSea. GSVA and Spearman's Rank Correlations were used to compute the activities of functional states and the correlations between the activities and gene expressions, respectively. For analysis using GEPIA the method for differential analysis is one-way ANOVA, using disease state (Tumor or Normal) as variable for calculating differential expression. R2 determines *p*-values for the differential expression of genes by performing either a one-way Anova (default setting). Kaplan–Meier survival curves were plotted, and the log-rank test was performed. Univariate and multivariate analyses were performed using the Cox proportional hazards model. The $p < 0.05$ was considered statistically significant. Table 1 contains the median expression values and the corresponding statistical significance values. Table 2 contains the multivariate Cox analysis of prognostic significance of MUC1 gene expression in patients with breast cancer (OS, RFS and DFS).

RESULTS

MUC1 is highly expressed in most solid tumors including breast carcinoma

Gene Expression Profiling Interactive Analysis (GEPIA) (http://gepia.cancer-pku.cn/)[10] revealed that MUC1 is highly expressed in all solid tumors including breast carcinoma where it was overexpressed more than ~6 fold as compared to its paired normal tissue (Figure 1A and 1B). Transcripts per million (TPM) analysis for 1085 breast tumor tissues also showed a similar high expression profile of MUC1 which was also non-overlapping with its paired normal tissues (n=291) ***(Figure S1A)***. Data from TCGA were analyzed via UALCAN (http://ualcan.path.uab.edu/) interestingly showed that expression of MUC1 significantly ($p<0.0001$) increased with an increase in patient age and stage of disease progression (Figure 1C and 1D)[11]. Next, we also analyzed the level of protein expression in normal and tumor tissue using UALCAN CPTAC analysis. Results revealed a significantly higher expression of MUC1 in the tumor tissue (n=125) as compared to the normal tissue (n=18) (Figure 1E). Collectively, these data corroborate previous studies[19-21] which demonstrated that MUC1 expression was upregulated in tumor tissues.

MUC1 overexpression was associated with poor patient survival in breast cancer patients

To determine the prognostic value of MUC1 expression in breast cancers, we analyzed the data via the Kaplan–Meier plot. We analyzed the association of high expression of MUC1 protein with patient survival using R2 (https://hgserver1.amc.nl/cgi-bin/r2/main.cgi). High expression of MUC1 significantly decreased overall survival ($p=0.014$) including disease-free survival ($p<0.0001$) and recurrence-free survival ($p=0.052$) (Figure 2A, 2B, 2C). The mean survival of patients with low expression of MUC1 was 82.8 ± 7.0 months while it was 45.5 ± 5.2 months for those patients with high MUC1 expression. We also analyzed the correlation of patient survival with MUC1 protein expression via Kaplan-Meier Plotter breast cancer protein analysis (https://kmplot.com/analysis/index.php?p=service)[12]. Similar to mRNA expression MUC1 protein expression was also associated with poor prognosis with a survival of only 17 months in case of high expression as compared to 41 months in case of low expression. Our data-mining approach corroborated previous reports[21] showing high MUC1 expression poorly correlates with patient overall survival.

Positive Correlation between MUC1 and hallmarks of CSCs

A recent study confirmed the expression of MUC1 on human embryonic pluripotent stem cells which function as a receptor for a metastasis-associated protein NM23-H1[22]. Several studies have indicated that MUC1 plays a critical role in the transcriptional regulation of genes associated with tumor invasion, metastasis, angiogenesis, proliferation, drug resistance, inflammation, and immune regulation all of which are hallmark properties of CSCs[3]. Our data mining approach using cancer single-cell functional state atlas, CancerSea database (http://biocc.hrbmu.edu.cn/CancerSEA/home.jsp) and utilizing dataset GSE77308, GSE75688, and GSE75367 revealed a moderate positive correlation between MUC1 and CSC

TABLE 1. Median expression values and statistical analyses of MUC1 expression on tumor cells.

Variable	*Control*			*Patient*			*p-value*
MUC1 expression (Transcript per million analysis)	***Median expression***	***Upper quartile***	***Lower quartile***	***Median expression***	***Upper quartile***	***Lower quartile***	***Normal v/s Variable (one-way Anova)***
Normal	162.765	269.127	113.292				
Tumor				419.67	1469.397	276.431	<0.05
MUC1 expression (Patient age)							**Normal v/s Variable (Student's t-test)**
Normal	162.765	269.127	113.292				
21-40 BRCA				441.106	946.006	206.774	2.56820000044122E-07
41-60 BRCA				669.346	1461.625	270.406	<1E-12
61-80 BRCA				766.105	1544.14	303.02	<1E-12
81-100 BRCA				888.112	2097.268	393.065	1.47490000212969E-08
MUC1 expression (Patient stage)							
Normal	162.765	269.127	113.292				
Stage 1				775.395	1499.275	344.713	<1E-12
Stage 2				632.573	1399.778	232.467	1.62436730732907E-12
Stage 3				756.866	1551.649	312.77	1.62436730732907E-12
Stage 4				890.455	1861.857	281.251	5.220000E-04
MUC1 expression (Patient's race)							
Normal	162.765	269.127	113.292				
Caucasian				772.986	1593.52	323.608	<1E-12
African-american				386.389	824.145	165.192	5.03963537568097E-12
Asian				516.016	1218.413	196.32	1.89600000000123E-06
MUC1 expression (protein)							
Normal	-0.62	-0.331	-0.851				
Primary Tumor				0.002	0.524	-0.733	1.146679E-04
MUC1 promoter methylation level							
Normal	0.216	0.255	0.203				
Tumor				0.118	0.141	0.1	<1E-12
MUC1 promoter methylation level (Patient stage)							
Normal	0.216	0.225	0.203				
Stage 1				0.121	0.147	0.107	1.62447832963153E-12
Stage 2				0.113	0.136	0.095	<1E-12
Stage 3				0.126	0.146	0.105	1.62447832963153E-12
Stage 4				0.12	0.135	0.097	<1E-12

TABLE 2. Multivariate Cox regression analysis of prognostic significance of MUC1 gene expression in patients with breast cancer.

Survival	*Median survival in months (High expression cohort)*	*Median survival in months (Low expression cohort)*	*Hazard Ratio*	*95% CI*	*p-value*
Overall Survival	33	86	1.21	[1.09 -1.34]	0.039
Disease-free Survival	NA	NA	2.5	[1.19-5.26]	0.013
Relapse-free Survival	24.07	43	1.67	[1.73-2.71]	0.035

OS: overall survival; RFS: relapse-free survival; DFS: disease-free survival; HR: hazard ratio.

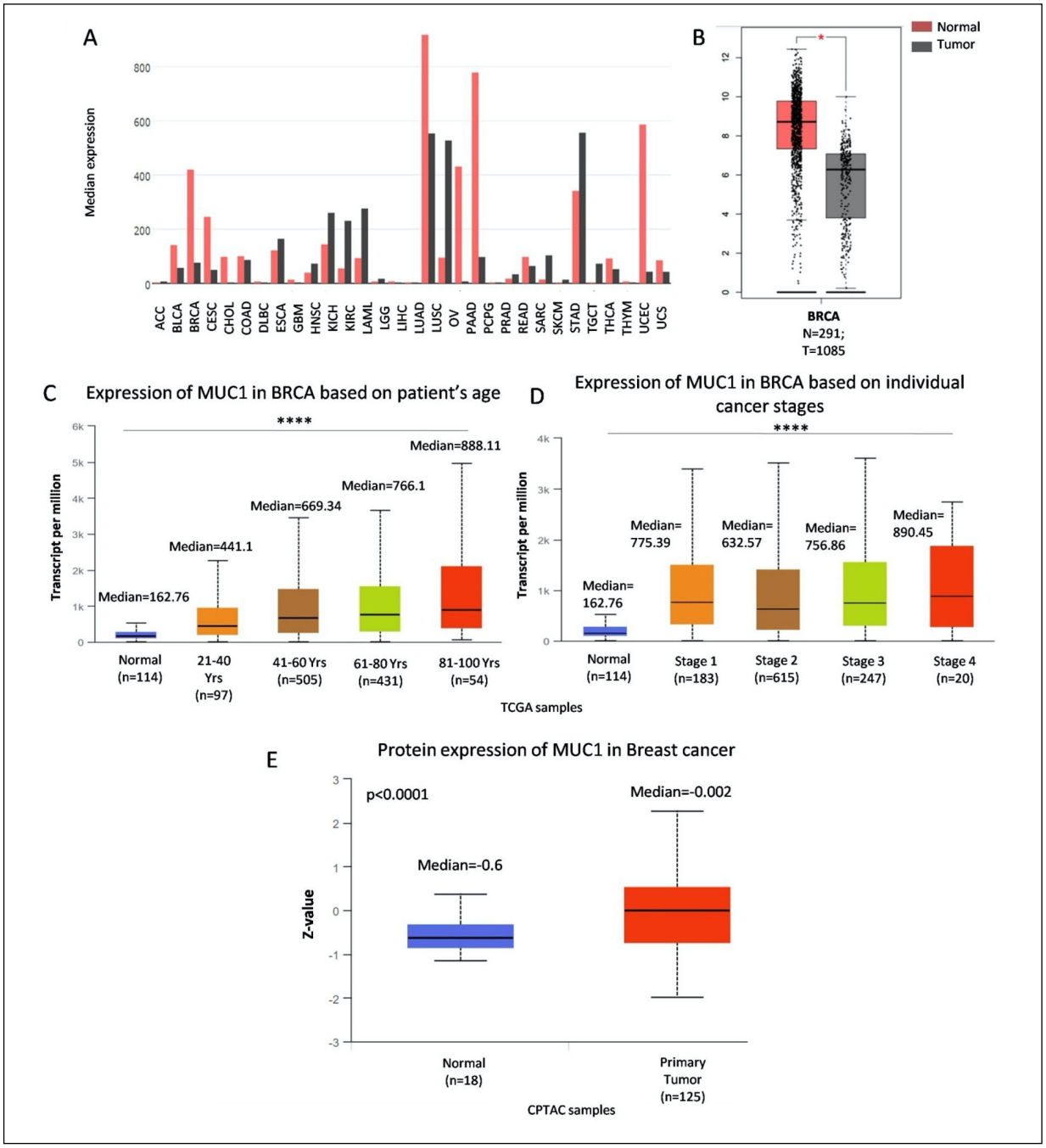

Figure 1. *A*, Bar diagram representing the expression pattern of MUC1 in normal and tumor tissues in various types of cancer. *B*, Box plot showing high expression of MUC1 in tumor tissue as compared to its paired normal. *C*, Box plot showing differential expression of MUC1 in tumor patients according to patient age. *D*, Box plot showing differential expression of MUC1 in tumor patients according to cancer stage. *E*, Box plot showing the expression of MUC1 protein in normal and tumor tissue.

hallmark functional states i.e stemness (r=0.44, $p<0.05$), metastasis (r=0.33, $p<0.01$) and invasion (r=0.41, $p<0.05$), and a moderate negative correlation with proliferation (r=-0.38, $p<0.05$) confirming the role of MUC1 is regulating the functional state of CSCs (Figure 3A and 3B) corroborating previous reports by Yuan et al[13].

Positive Correlation between MUC1 and CSC signature genes

Having confirmed the role of MUC1 in regulating the functional state of CSCs via our *in-silico* analysis, we next investigated its possible role in regulating CSC signature genes. We conducted a

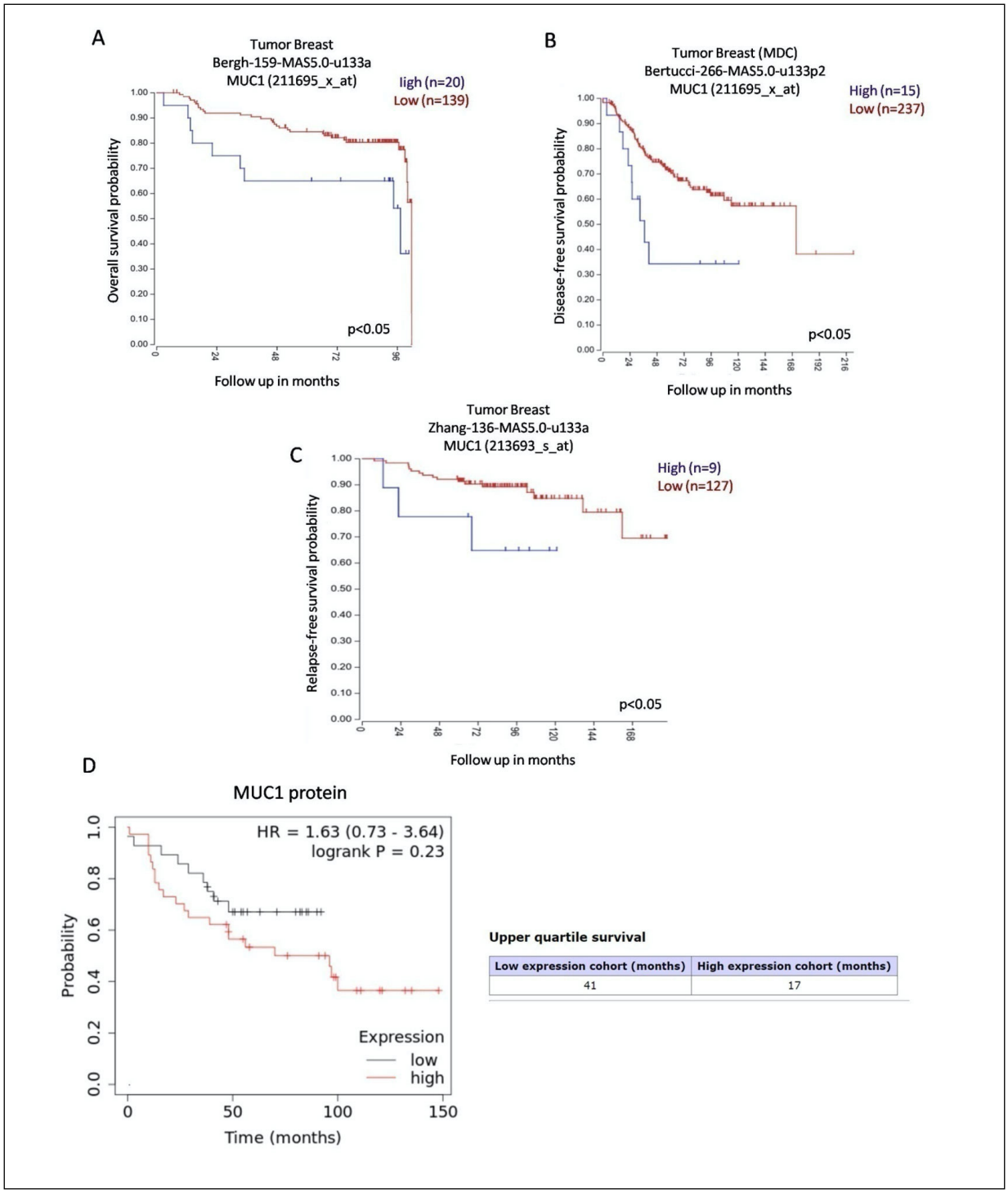

Figure 2. KM plot showing decrease in *(A)* overall survival *(B)* disease-free survival *(C)* relapse-free survival of patients with an increase in MUC1 mRNA expression. *D*, KM plot showing overall survival in case of high and low MUC1 protein expression.

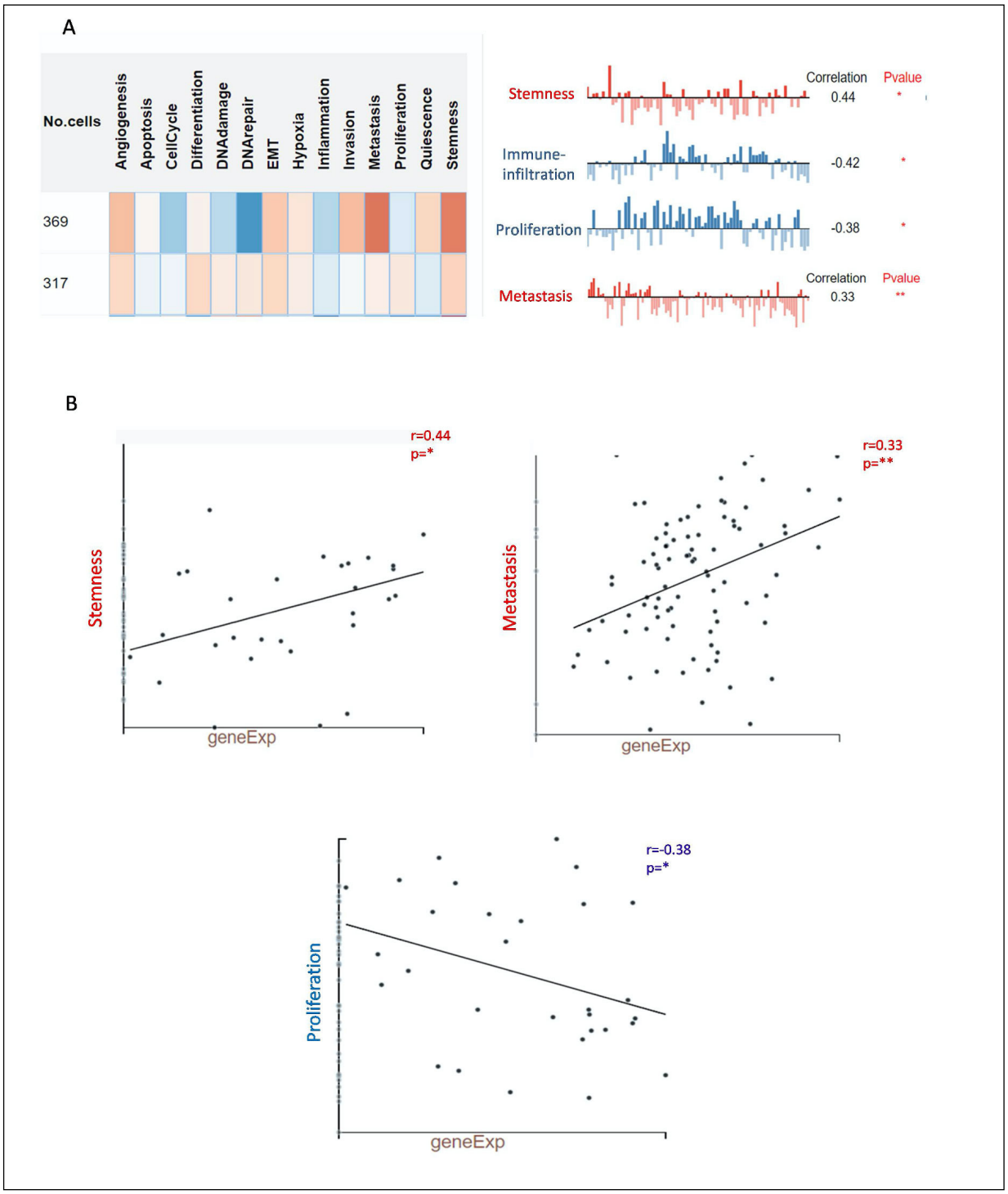

Figure 3. *A* and *B*, Single-cell functional state analysis showing positive correlation of MUC1 expression with hallmarks of CSCs including metastasis and stemness. A negative correlation between MUC1 expression and proliferation.

TCGA analysis using genomics analysis and visualization application R2 (https://hgserver1.amc.nl/cgi-bin/r2/main.cgi). Our analysis revealed a moderate positive correlation between MUC1 and cancer stem cell signature gene KLF4 (r=0.367; $p<0.0001$)[23], Nanog (r=0.29; $p<0.0001$)[21], SOX2 (r=0.33; $p<0.0001$)[24], and ALDH1A1 (r=0.28; $p<0.0001$)[25]. Whereas we observed a weak positive correlation between OCT4 (r=0.26; $p<0.0001$)[8], and MYC (r=0.266; $p<0.0001$)[26] (Figure 4A). We also analyzed the correlation of MUC1 and CSC signature gene ALDH1 using mRNA expression heat-map generated using TCGA BRCA dataset, it was confirmed that MUC1 expression gradually

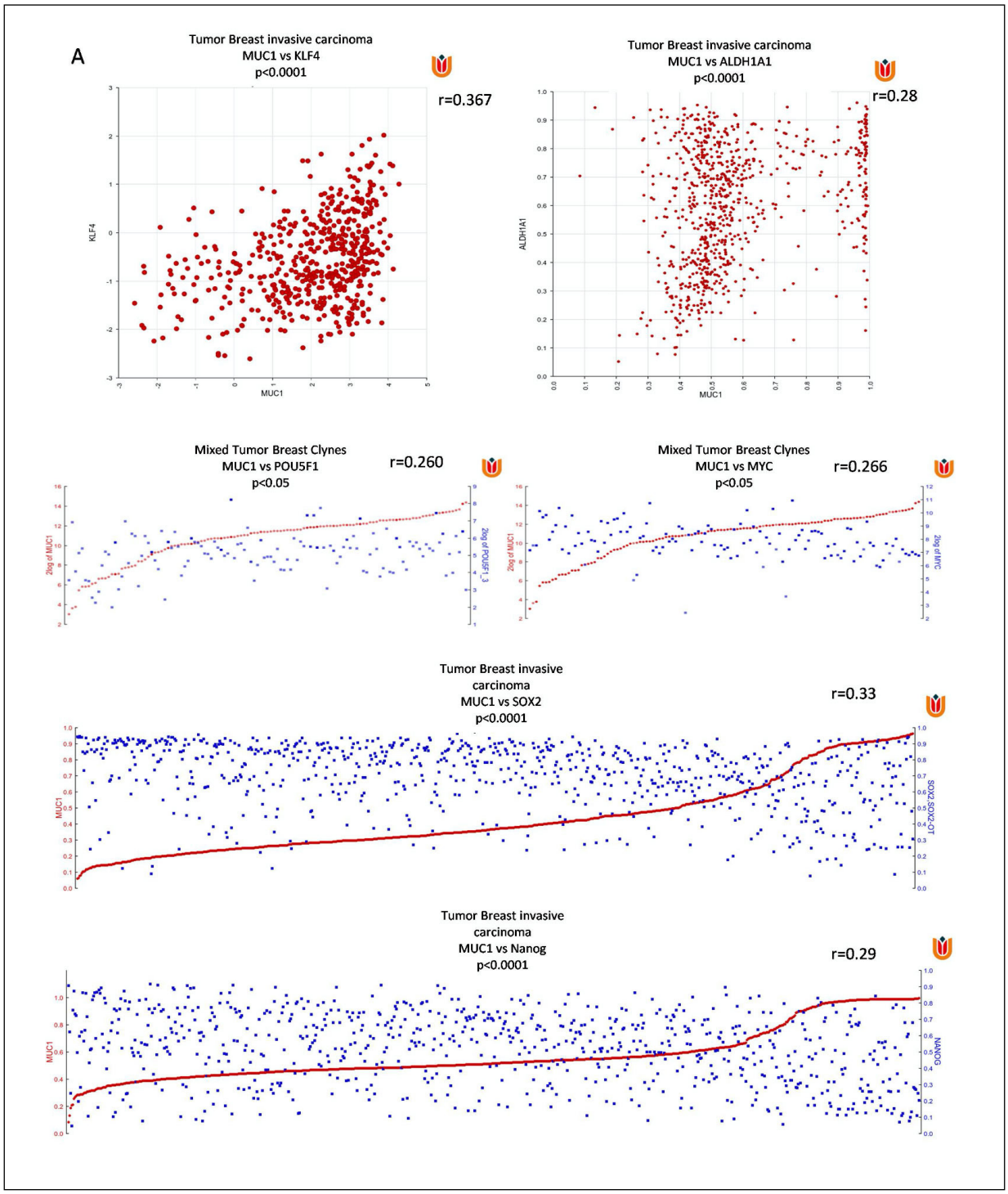

Figure 4. A, Positive correlation between CSC signature gene KLF4 (R=0.367, *p*=2.95*10-18), ALDH1 (R=0.1, *p*=0.019), SOX2 (R=0.155, *p*=0.089), Nanog (R=0.157, *p*=0.086), OCT4 (R=0.26, *p*=3.988*10-3, MYC (R=0.266, *p*=3.17*10-3).

increased with increasing ALDH1 expression, we also found a strong correlation between MUC1 and ALDH1 among different races of the population using UALCAN (http://ualcan.path.uab.edu/) ***(Figure S2A and S2B)***. Collectively, our results suggest a strong positive correlation between the CSC signature gene and MUC1 among different races of the population as well.

Differential methylation of MUC1 core promoter region in CSCs

In an attempt to delineate the possible reason for overexpression of MUC1 in CSCs we looked at the promoter methylation pattern of the MUC1 core promoter region. Our analysis revealed putative CpG islands in the core promoter region of

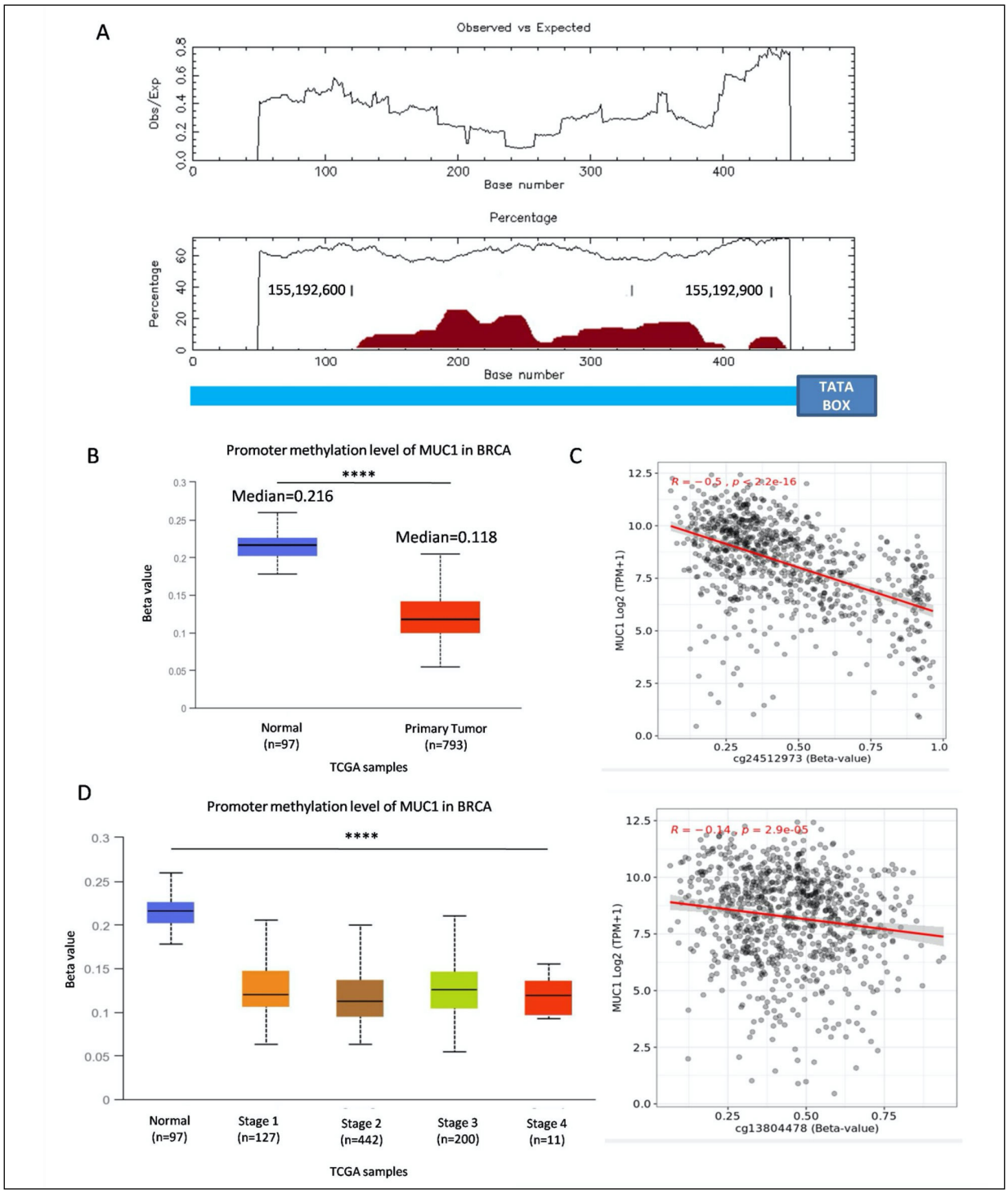

Figure 5. *A*, MUC1 promoter analysis using EMBOSS CpG plot revealed CpG island present in the core promoter region and also the H3K4me1 methylation pattern of that region (-499 to 0) using Eukaryotic Promoter Database (EPD) (***https://epd.epfl.ch/cgibin/get_doc?db=hgEpdNew&format=genome&entry=MUC1_1***) *B*, Box-plot showing the promoter methylation pattern of MUC1 in BRCA. C) Correlation between CpG sites and MUC1 expression. D) Boxplot showing promoter methylation pattern of MUC1 in BRCA at different stages of the disease.

MUC1 (Figure 5A). Subsequently, the MUC1 promoter region was also found to be hypomethylated in tumor tissues including breast tumor tissues (n=793) as compared to its paired normal counterpart (n=97) (Figure 5B and ***Figure S3A***). Analysis using SMART (http://www.bioinfo-zs.com/smartapp/) divulged a negative correlation between methylation in CpG sites (cg24512973 and cg13804478) and gene expression suggesting an increase in expression of MUC1 in a hypometh-

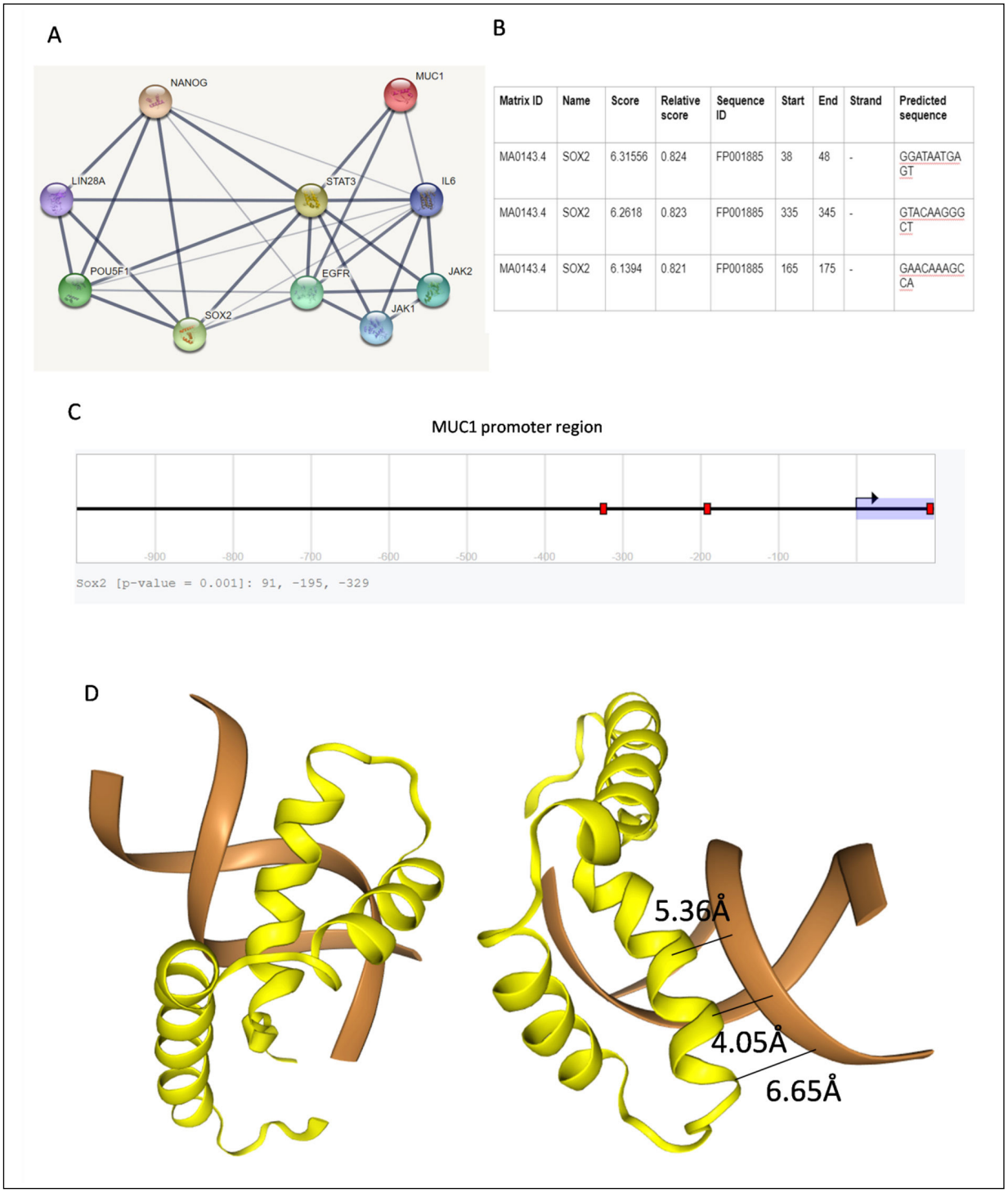

Figure 6. *A*, STRING analysis showing the interacting partners of MUC1. *B*, Binding site of SOX2 on MUC1 core promoter region. *C*, Diagrammatic representation of SOX2 binding site on MUC1 promoter. *D*, Representative structure of SOX2 binding on MUC1 core promoter region.

ylated state (Figure 5C)[15]. Extrapolating this data onto the stage-wise methylation pattern of MUC1 revealed a decrease in methylation of MUC1 with an increase in the stage of the tumor. An increase in tumor stage increases CSC signature genes and is also known to increase CSC number (Figure 5D). Therefore, it will be reasonable to speculate that the MUC1 core promoter region will be hypomethylated in CSCs as compared to differentiated cancer cells.

Putative binding of SOX2 on the MUC1 promoter region

In an attempt to delineate the mechanism of overexpression of MUC1 in CSCs, we constructed a STRING diagram to bring into perspective high confidence interacting partners of MUC1[27]. Our analysis affirmed a strong interaction between MUC1 and STAT3[28] which is an important transcription factor known to be overexpressed in CSCs. Interestingly, we also found a high confidence interaction between MUC1 and SOX2 via the STAT3 pathway (Figure 6A). This led us to scout for the possible binding sites of SOX2 on the MUC1 promoter region. Our analysis revealed 3 putative binding sites for SOX2 in the MUC-1 core promoter region (Figure 6B and 6C). Our docking analysis also revealed a strong binding of SOX2 on the MUC1 promoter region with a docking energy score of -268.0 (Figure 6D). Since more negative dock score corresponds to more preferential binding, these results indicate the preference of SOX2-binding at the core promoter. However, in the best models, the H1 helix was found to be inserted into the major groove of the promoter DNA (Figure 6D), similar to the Nanog-Oct4 promoter complex[29]. Close interaction was observed between Gln17 (5.36Å), Asn19 (4.05Å), and Asn24 (6.65Å) within the proximity of <7Å. Further studies are warranted to confirm the molecular mechanism and epigenetic regulations that governs MUC1 expression in CSCs. Furthermore, in-depth investigations are needed to confirm whether Sox2 plays an integral part in regulating the expression of MUC1 in CSCs.

DISCUSSION

MUC1 oncoprotein overexpression correlates with the aggressiveness of tumors and poor overall survival of cancer patients[20]. A recent study conducted by the National Cancer Institute (NCI) Translational Research Working Group ranked MUC1 as the second-best potential target out of the 75 targets studied involving tumor-associated antigens for the development of cancer vaccine[30]. Currently, there are 322 MUC1-based clinical trials registered with ClinicalTrials.gov (https://clinicaltrials.gov/) which suggest the increasing number of therapies being developed targeting MUC1[31]. MUC1 is known to promote tumor progression and invasiveness through the activation of β-catenin, NF-κB, pyruvate kinase muscle isozyme M2, EGFR and other pathways[32,33].

In our meta-analysis, we have not only corroborated previous studies which highlighted the positive correlation between MUC1, tumor metastasis and "stemness" but we have also extrapolated this data to verify the significant correlation between MUC1 and CSC signature genes which is in concurrence with previous studies[34]. Therefore, targeting MUC1 will allow us to not only target cancer cells but at the same time will also allow us to precisely target the much more resilient CSC population. We have also identified CpG islands in the core promoter region of MUC1 which was found to be hypomethylated in the tumor tissue as compared to the normal counterpart which is in concurrence with previous reports[21]. The subsequent analysis also revealed a negative correlation between MUC1 expression and promoter methylation. Our results are in agreement with previous reports which observed a high level of CpG methylation in MUC1-negative/low-expression cell lines only in the vicinity of the transcriptional start site[35]. All together these data suggest that a hypomethylated MUC1 promoter promotes its overexpression in the cancer cells. Furthermore, in an attempt to delineate the promoter methylation pattern of MUC1 in CSCs we indirectly correlated the methylation pattern of the MUC1 promoter with the stage of disease progression. An advanced stage often correlates with more aggressive disease and is related to a higher number of CSCs along with CSC-associated genes[36]. Recently Hata et al[34] demonstrated that MUC1-C activates STAT3 which then drive induction of the TWIST1 gene. In turn, the MUC1-C/TWIST1 circuit drives (i) self-renewal capacity (ii) expression of the stem cell markers SOX2, BMI1, ALDH1 and CD44, and (iii) tumorigenicity[34]. In concert with these results, we also predicted interaction of MUC1 and SOX2 via STAT3. Interestingly, we identified putative binding sites for SOX2 on the core promoter region of MUC1. Studies reported thus far indicate that high SOX2 levels correlate with poor prognosis for patients with many different cancers, including breast[37]. MUC1 is known to suppress the p53 pathway, induce the Yamanaka pluripotency factors (OCT4, SOX2, KLF4, and MYC) and drive stemness[38]. This interaction between SOX2 and MUC1 core promoter may be responsible for MUC1 overexpression in CSCs. SOX2 is reported as an important factor for upregulation of gastric foveolar mucin, MUC5AC in colorectal mucinous and signet ring cell carcinomas[39]. SOX2 is also known to induce passive demethylation at SOX2 binding sites and SOX2 binding motifs can attract demethylases[40,41] which can be the possible reason for the hypomethylated promoter of MUC1 in cancer cells. Thus, it will be interesting to determine the role of SOX2 in overexpression of MUC1 as well. There are many reports suggesting

the role of MUC1 in driving SOX2 expression in CSCs, our work further predicts the cross-interaction of SOX2 with MUC1 and how these two molecules may be interlinked in a unique two to tango relationship. The present study was hypothetically driven and performed using experimental generated data available in public databases. Therefore, further studies to elucidate the role of SOX2 in promoting demethylation of the MUC1 promoter region and its overexpression may give us detailed insights into the mechanisms controlling MUC1 expression in CSCs and thus providing us with a unique opportunity to target CSCs.

CONCLUSIONS

Our meta-analysis has revealed the previously unknown role of SOX2 in regulating the expression of MUC1. Further *in-vitro* and *in-vivo* studies will be able to bring forth the exact molecular mechanism governing MUC1 regulation by SOX2. It will also be interesting to study the MUC1/SOX2/STAT3 axis in driving hallmarks of CSCs. This study hopes to initiate the conversation about the factors controlling MUC1 in CSCs which may become a lucrative target for future anti-CSC therapy. This work should be further extended to include patient data and corroborate our findings.

Financial support:
The authors received no financial support for the research.

Ethics approval and consent to participate:
Not applicable.

Consent for publication:
The consent to publish had been taken from each participant in this work.

Availability of data and material:
Data analyzed in this study were re-analysis from existing data, which are openly available at locations cited in the reference section.

Competing interests:
The authors declare that they have no competing interests.

Funding:
This study was funded by Council of Scientific and Industrial Research (CSIR), Government of India.

Authors Contributions:
AD analyzed and interpreted the patient data as well as bioinformatics and conceived the idea of the manuscript. SP was responsible for data analysis, bioinformatics, and was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

Acknowledgements:
The authors would like to thank CSIR, Government of India and DBT, Government of India

REFERENCES

1. Walcher L, Kistenmacher AK, Suo H, Kitte R, Dluczek S, Strauß A, Blaudszun AR, Yevsa T, Fricke S, & Kossatz-Boehlert U. Cancer Stem Cells-Origins and Biomarkers: Perspectives for Targeted Personalized Therapies. Front. immunol 2020; 11: 1280.
2. Francescangeli F, Contavalli P, De Angelis ML, Careccia S, Signore M, Haas TL, Salaris F, Baiocchi M, Boe A, Giuliani A, Tcheremenskaia O, Pagliuca A, Guardiola O, Minchiotti G, Colace L, Ciardi A, D'Andrea V, La Torre F, Medema J, De Maria R, Zeuner A. A pre-existing population of ZEB2+ quiescent cells with stemness and mesenchymal features dictate chemoresistance in colorectal cancer. J Exp Clin Cancer Res 2020; 39: 2.
3. Nath S, Mukherjee P. MUC1: a multifaceted oncoprotein with a key role in cancer progression. Trends Mol Med 2014; 20: 332–342.
4. Hattrup CL, Gendler SJ. Structure and function of the cell surface (tethered) mucins. Annu Rev Physiol 2008; 70: 431–457.
5. Ganguly K, Rauth S, Marimuthu S, Kumar S, Batra SK. Unraveling mucin domains in cancer and metastasis: when protectors become predators. Cancer Metastasis Rev 2020; 39: 647–659.
6. Gnemmi V, Bouillez A, Gaudelot K, Hémon B, Ringot B, Pottier N, Glowacki F, Villers A, Vindrieux D, Cauffiez C, Van Seuningen I, Bernard D, Leroy X, Aubert S, Perrais M. MUC1 drives epithelial-mesenchymal transition in renal carcinoma through Wnt/βcatenin pathway and interaction with SNAIL promoter. Cancer Lett 2014; 346: 225–236.
7. Rajabi H, Kufe D. MUC1-C Oncoprotein Integrates a Program of EMT, Epigenetic Reprogramming and Immune Evasion in Human Carcinomas. Biochim Biophys Acta Rev Cancer 2017; 1868: 117–122.
8. Hikita ST, Kosik KS, Clegg DO, Bamdad C. MUC1* mediates the growth of human pluripotent stem cells. PloS one 2008; 3: e3312.
9. Li W, Zhang N, Jin C, Long MD, Rajabi H, Yasumizu Y, Fushimi A, Yamashita N, Hagiwara M, Zheng R, Wang J, Kui L, Singh H, Kharbanda S, Hu Q, Liu S, Kufe D. MUC1-C drives stemness in progression of colitis to colorectal cancer. JCI insight 2020; 5: e137112.
10. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res 2017; 45: W98–W102.
11. Chandrashekar DS, Bashel B, Balasubramanya S, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, Varambally S. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. Neoplasia (New York, N.Y.) 2017; 19: 649–658.
12. Győrffy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q, Szallasi Z. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. Breast Cancer Res Treat 2010; 123: 725–731.
13. Yuan H, Yan M, Zhang G, Liu W, Deng C, Liao G, Xu L, Luo T, Yan H, Long Z, Shi A, Zhao T, Xiao Y, Li X. CancerSEA: a cancer single-cell state atlas. Nucleic Acids Res 2019; 47: D900–D908

14. Dreos R, Ambrosini G, Périer RC, Bucher P. The Eukaryotic Promoter Database: expansion of EPDnew and new promoter analysis tools. Nucleic Acids Res 2015; 43: D92–D96.
15. Li Y, Ge D, Lu C. The SMART App: an interactive web application for comprehensive DNA methylation analysis and visualization. Epigenetics chromatin, 2019; 12: 71.
16. Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, Modi BP, Correard S, Gheorghe M, Baranašić D, Santana-Garcia W, Tan G, Chèneby J, Ballester B, Parcy F, Sandelin A, Lenhard B, Wasserman WW, Mathelier A. JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res 2020; 48: D87–D92.
17. Yan Y, Zhang D, Zhou P, Li B, Huang S-Y. HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. Nucleic Acids Res 2017; 45: W365-W373.
18. Daniel JW, David VDB, Fei P, Berman BP, Laird PW. Comprehensive DNA methylation analysis on the Illumina® infinium® assay platform. Illumina. 2008.
19. McGuckin MA, Walsh MD, Hohn BG, Ward BG, Wright RG. Prognostic significance of MUC1 epithelial mucin expression in breast cancer. Hum Pathol 1995; 26: 432–439.
20. Horm TM, Schroeder JA. MUC1 and metastatic cancer: expression, function and therapeutic targeting. Cell Adh Migr 2013; 7: 187–198.
21. Jing X, Liang H, Hao C, Yang X, Cui X. (2019). Overexpression of MUC1 predicts poor prognosis in patients with breast cancer. Oncol Rep 2019; 41: 801–810.
22. Smagghe BJ, Stewart AK, Carter MG, Shelton LM, Bernier KJ, Hartman EJ, Calhoun AK, Hatziioannou VM, Lillacci G, Kirk BA, DiNardo BA, Kosik KS, Bamdad C. MUC1* ligand, NM23-H1, is a novel growth factor that maintains human stem cells in a more naïve state. PloS one 2013; 8: e58601.
23. Wei X, Xu H, Kufe D. Human mucin 1 oncoprotein represses transcription of the p53 tumor suppressor gene. Cancer Res 2007; 67: 1853–1858.
24. Ham SY, Kwon T, Bak Y, Yu JH, Hong J, Lee SK, Yu DY, Yoon DY. Mucin 1-mediated chemo-resistance in lung cancer cells. Oncogenesis 2016; 5: e185.
25. Alam M, Ahmad R, Rajabi H, Kharbanda A, Kufe D. MUC1-C oncoprotein activates ERK→C/EBPβ signaling and induction of aldehyde dehydrogenase 1A1 in breast cancer cells. J Biol Chem 2013; 288: 30892–30903.
26. Tagde A, Rajabi H, Bouillez A, Alam M, Gali R, Bailey S, Tai YT, Hideshima T, Anderson K, Avigan D, Kufe D. MUC1-C drives MYC in multiple myeloma. Blood 2016; 127: 2587–2597.
27. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. V. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res, 2019; 47: D607–D613.
28. Ahmad R, Rajabi H, Kosugi M, Joshi MD, Alam M, Vasir B, Kawano T, Kharbanda S, Kufe D. MUC1-C oncoprotein promotes STAT3 activation in an autoinductive regulatory loop. Sci Signal 2011; 4: ra9.
29. Hayashi Y, Caboni L, Das D, Yumoto F, Clayton T, Deller MC, Nguyen P, Farr CL, Chiu HJ, Miller MD, Elsliger MA, Deacon AM, Godzik A, Lesley SA, Tomoda K, Conklin BR, Wilson IA, Yamanaka S, Fletterick RJ. Structure-based discovery of NANOG variant with enhanced properties to promote self-renewal and reprogramming of pluripotent stem cells. Proc Natl Acad Sci U.S.A 2015; 112: 4666–4671.
30. Cheever MA, Allison JP, Ferris AS, Finn OJ, Hastings BM, Hecht TT, Mellman I, Prindiville SA, Viner JL, Weiner LM, Matrisian LM. The prioritization of cancer antigens: a national cancer institute pilot project for the acceleration of translational research. Clin Cancer Res 2009; 15: 5323–5337.
31. Gatti-Mays ME, Redman JM, Donahue RN, Palena C, Madan RA, Karzai F, Bilusic M, Sater HA, Marté JL, Cordes LM, McMahon S, Steinberg SM, Orpia A, Burmeister A, Schlom J, Gulley JL, Strauss J. A Phase I Trial Using a Multitargeted Recombinant Adenovirus 5 (CEA/MUC1/Brachyury)-Based Immunotherapy Vaccine Regimen in Patients with Advanced Cancer. The oncologist 2020; 25: 479–e899.
32. Bernier AJ, Zhang J, Lillehoj E, Shaw AR, Gunasekara N, Hugh JC. Non-cysteine linked MUC1 cytoplasmic dimers are required for Src recruitment and ICAM-1 binding induced cell invasion. Mol Cancer 2011; 10: 93.
33. Roy LD, Sahraei M, Subramani DB, Besmer D, Nath S, Tinder TL, Bajaj E, Shanmugam K, Lee YY, Hwang SI, Gendler SJ, Mukherjee P. MUC1 enhances invasiveness of pancreatic cancer cells by inducing epithelial to mesenchymal transition. Oncogene 2011; 30: 1449–1459.
34. Hata T, Rajabi H, Yamamoto M, Jin C, Ahmad R, Zhang Y, Kui L, Li W, Yasumizu Y, Hong D, Miyo M, Hiraki M, Maeda T, Suzuki Y, Takahashi H, Samur M, Kufe D. Targeting MUC1-C Inhibits TWIST1 Signaling in Triple-Negative Breast Cancer. Mol Cancer Ther 2019; 18: 1744–1754.
35. Yamada N, Nishida Y, Tsutsumida H, Hamada T, Goto M, Higashi M, Nomoto M, Yonezawa S MUC1 expression is regulated by DNA methylation and histone H3 lysine 9 modification in cancer cells. Cancer Res 2008; 68: 2708–2716.
36. Ayob AZ, Ramasamy TS Cancer stem cells as key drivers of tumour progression. J Biomed Sci 2018; 25: 20.
37. Wuebben EL, Rizzino A. The dark side of SOX2: cancer - a comprehensive overview. Oncotarget 2017; 8: 44917–44943.
38. Yasumizu Y, Rajabi H, Jin C, Hata T, Pitroda S, Long MD, Hagiwara M, Li W, Hu Q, Liu S, Yamashita N, Fushimi A, Kui L, Samur M, Yamamoto M, Zhang Y, Zhang N, Hong D, Maeda T, Kosaka T, Kufe D. MUC1-C regulates lineage plasticity driving progression to neuroendocrine prostate cancer. Nat Commun 2020; 11: 338.
39. Park ET, Gum JR, Kakar S, Kwon SW, Deng G, Kim YS. Aberrant expression of SOX2 upregulates MUC5AC gastric foveolar mucin in mucinous cancers of the colorectum and related lesions. Int J Cancer 2008; 122: 1253–1260.
40. Hori N, Yamane M, Kouno K, Sato K. Induction of DNA demethylation depending on two sets of Sox2 and adjacent Oct3/4 binding sites (Sox-Oct motifs) within the mouse H19/insulin-like growth factor 2 (Igf2) imprinted control region. J Biol Chem 2012; 287: 44006–44016.
41. Vanzan L, Soldati H, Ythier V, Anand S, Francis N, Murr R. High throughput screening identifies SOX2 as a Super Pioneer Factor that inhibits DNA methylation maintenance at its binding sites. bioRxiv 2020.