BMB
Reports

Analysis of opposing histone modifications H3K4me3 and H3K27me3 reveals candidate diagnostic biomarkers for TNBC and gene set prediction combination

*Hyoung-Min Park, HuiSu Kim, Kang-Hoon Lee & Je-Yoel Cho**

Department of Biochemistry, BK21 Plus and Research Institute for Veterinary Science, School of Veterinary Medicine, Seoul National University, Seoul 08826, Korea

Breast cancer encompasses a major portion of human cancers and must be carefully monitored for appropriate diagnoses and treatments. Among the many types of breast cancers, triple negative breast cancer (TNBC) has the worst prognosis and the least cases reported. To gain a better understanding and a more decisive precursor for TNBC, two major histone modifications, an activating modification H3K4me3 and a repressive modification H3K27me3, were analyzed using data from normal breast cell lines against TNBC cell lines. The combination of these two histone markers on the gene promoter regions showed a great correlation with gene expression. A list of signature genes was defined as active (highly enriched H3K4me3), including NOVA1, NAT8L, and MMP16, and repressive genes (highly enriched H3K27me3), IRX2 and ADRB2, according to the distribution of these histone modifications on the promoter regions. To further enhance the investigation, potential candidates were also compared with other types of breast cancer to identify signs specific to TNBC. RNA-seq data was implemented to confirm and verify gene regulation governed by the histone modifications. Combinations of the biomarkers based on H3K4me3 and H3K27me3 showed the diagnostic value AUC 93.28% with P-value of 1.16e-226. The results of this study suggest that histone modification analysis of opposing histone modifications may be valuable toward developing biomarkers and targets for TNBC. [BMB Reports 2020; 53(5): 266-271]

INTRODUCTION

Breast cancer is one of the most common cancers occurring among females and one of the most dominant causes of cancer related deaths alongside lung cancer (1). Known as highly diverse cancers, breast cancers are characterized by distinct genetic variations, clinical symptoms, treatments, and prognosis outcomes. In previous studies, breast cancer has been clinically classified by major changes in expression levels (2) including high expression of the estrogen receptor (ESR), progesterone receptors (PGR), and HER2. Most clinically diagnosed breast cancer types have at least one of these features, but basal-like triple negative breast cancer (TNBC) presents no expression of the three. TNBC is more aggressive and has poor prognosis, but because of its minor occurrence treatments and therapies, are scarce (3). Some patients suffering from TNBC benefit from chemotherapy, but still need a better method of treatment less toxic and dangerous to the patient. Recent studies of TNBC revealed distinct gene mutation patterns and repressive signal pathways (4). Despite the effort of continuing research, the understanding of the governing gene mechanism and systemic regulation of TNBC pathways is lacking compared to other more dominant breast cancer types.

Breast cancer molecular identities can be further specified based on epigenetic features. Epigenetic regulation has been a major factor of gene expression control (5). Various types of epigenetic control such as DNA methylation and histone modification are crucial for the activation and repression of genes in cancer. Recent studies revealed DNA methylation and histone modification profiles as plausible predictors of well-defined subtypes (6). Among the different types of histone modifications, H3K4me3 is a major modification that moderates genes to an active state (7). Conversely, histone modification H3K27me3 is a major modification that downregulates genes when highly enriched (8). Continuing efforts to discover various precursors to breast cancer by comparing five or more histone modifications enabled a more thorough understanding and precision of prediction (9). However, less is known of the histone modifications specifically contributing to TNBC and the expression differences regulated by the histone regulation.

The purpose of this study was to establish an analytical pipeline for discovering TNBC biomarkers from published histone modification peak data. The combination of the two histone modifications, H3K4me3 and H3K27me3, in TNBC cell lines

*Corresponding author. Tel: +82-2-880-1268; Fax: +82-2-886-1268; E-mail: jeycho@snu.ac.kr

https://doi.org/10.5483/BMBRep.2020.53.5.045

Received 26 February 2020, Revised 17 March 2020, Accepted 20 March 2020

Keywords: ChIP-seq, Histone modification, Human, RNA-seq, Triple negative breast cancer

ISSN: 1976-670X (electronic edition)

presented hallmarks of TNBC gene expression against normal breast cell lines. The results providing genes that are epigenetically regulated in TNBC, were proven successfully by quantitative transcriptional analysis, and suggested biomarker candidates that could specifically diagnose TNBC against normal.

RESULTS

Distribution of H3K4me3 and H3K27me3 histone modifications as TNBC-associated epigenomic signatures

ChIP-seq data from HMEC and MDA-MB-436 that represent normal and TNBC cell lines, respectively, were obtained from the public dataset GSE62907. To determine the TNBC-enriched epigenetic alteration, two histone modification signals H3K4me3 and H3K27me3 on the gene promoter regions were compared across the two cell lines. Overall procedures are depicted in Fig. 1A. In brief, we normalized all ChIP-signal data to the corresponding inputs. The regions of differentially modified histones were identified from the comparison of HMEC and MDA-MB-436 cell lines. Up- and down-regulated histone modifications in TNBC were selected when regions had a larger than two-fold difference in H3K4me3 and H3K27me3, compared to the normal HMEC (Fig. 1A). The promoter region was defined by convention as 2 kb upstream of the transcriptional start site (TSS) of a gene.

Identified as up-regulated genes in TNBC were 1,008 genes with highly enriched H3K4me3 regions and 4,954 genes with depleted H3K27me3 signals in MDA-MB-436 cells. Conversely, 1,608 genes with enriched H3K4me3 and 5,082 genes with low H3K27me3 in HMBC were identified as down-regulated genes in TNBC (Table S1). As a result, a list of genes exclusively up-regulated (148) and down-regulated (41) in TNBC was determined by combining high H3K4me3 and low H3K27me3 profiles and vice versa (Fig. 1B). Each potential candidate was scored by its H3K4me3 peak score. Integrative genomics viewer (IGV) depicted histone modifications on the regions of the NOVA1 and IRX2 genes that were scored in the top (Fig. 1C). H3K4me3 signals were enriched and H3K27me3 disappeared on the NOVA1 promoter region in MDA-MB-436, while H3K4me3 signals are very low and H3K27me3 are enriched in HMEC. Oppositely, H3K4me3 signal was highly enriched and the H3K27me3 disappeared on the DUSP6 genes in HMEC, while H3K4me3 signals disappeared and H3K27me3 were enriched in the MDA-MB-436.

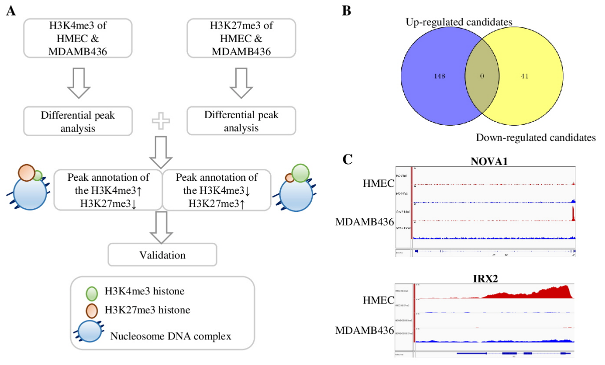

Fig. 1. Project workflow and ChIP-seq analysis. (A) Workflow of the ChIP-seq analysis. (B) Venn diagram of 148 potential up-regulated genes and 41 potential down-regulated genes in triple negative breast cancer. (C) The histone profile H3K4me3 (red) and H3K27me3 (blue) of candidate gene in the HMEC and MDA-MB-436 ChIP-seq datasets.

Integration of transcriptome data revealed TNBC-associated signature genes

Epigenetic profiles such as histone modification represent convincing evidence as biomarker candidates. However, epigenome profiles are not perfectly aligned to their corresponding expression data such as transcriptomic or proteomic data (10). To investigate the effects of epigenomic aberrations on gene expression, matching RNA-seq data of HMEC and MDA-MB-436 were merged with two additional transcriptome datasets obtained from other sub-types of breast cancer cell lines, SK-BR-3 (luminal type) and ZR-75-1 (HER2 expressing) (11). The influence of histone modification on gene expression was examined by calculating the percentage of RNA-seq expression patterns that match with histone peak fold changes. Overall, H3K4me3 has better correlation than H3K27me3 with gene expression levels in up- and down-regulation (Fig. S1A). Of note, the combination of histone markers, high H3K4me3 and low H3K27me3 for up-regulated genes and vise-versa for down-regulated genes, presented a remarkable improvement in the correlation with gene expression. The top 10 highest scored genes are indicated by red and blue dots for up- and down-regulated genes (Fig. S1). For further analysis, the top 10 scored genes (up- and down- regulated) in ChIP-seq were selected and listed with corresponding RNA expressions in Table 1.

Quantitative RT-PCR validation of candidates' gene expression

The expression of selected candidate genes was confirmed by quantitative real time RT-PCR in the corresponding breast cancer related cell lines, MCF-10A (normal), MDA-MB-436 (TNBC) and SK-BR-3 (HER2+). For the genes selected by highly-enriched H3K4me3 by depleted H3K27me3, the gene expressions of 10 up-regulated candidates were confirmed by real time RT-PCR. Except for the DNER gene which showed half of the expression in TNBC than in normal cell line, the nine remaining genes expressed highly in TNBC (Fig. 2A). Interestingly, we found two genes (MMP16, NAT8L), almost exclusively expressed in TNBC. The largest discrepancy in relative gene expression levels between TNBC and normal was in MMP16 (-3,000 fold) followed by NAT8L (-2,500 fold). DCLK2 and SYTL4 showed higher gene expression levels in cancer cell lines SK-BR-3 (HER2+) and MDA-MB-436 (TNBC). CDH2, NOVA1, PLCL2, SOX5, and SALL1 were highly expressed in TNBC, but not in

Table 1. Differentially expressed candidates

Gene name	Score	MDAMB436	SKBR3	ZR751	HMEC
CDH2	1619.5	757.46	428.79	10.67	35.11
DCLK2	857.0	436.45	43.76	1.93	1.28
NOVA1	850.3	120.23	1616.05	27.63	0.53
PLCL2	692.7	121.57	0.00	8.09	0.54
SOX5	826.3	37.11	1.07	6.66	0.05
SALL1	806.2	235.11	0.00	0.08	0.54
SYTL4	780.1	360.80	69.22	112.64	22.07
DNER	744.5	943.92	0.00	15.72	63.47
NAT8L	700.4	284.23	1827.86	320.66	3.46
MMP16	680.3	188.47	99.50	10.53	21.28
DUSP6	2052.3	58.25	9.73	38.26	2706.05
IRX2	1564.5	0.58	0.00	180.80	464.21
ATP2B1	1445.1	212.46	34.82	261.43	1771.94
VSNL1	1296.7	0.02	72.81	0.71	430.11
ADRB2	1248.1	0.01	47.46	0.07	115.33
PLXDC2	1139.4	24.44	0.00	384.42	1078.95
PLD5	1124.9	0.80	196.58	0.79	125.28
TPD52L1	1007.3	59.51	0.00	917.35	332.26
FAM84A	743.7	0.12	4.15	25.43	414.37
SNX19	648.5	553.93	717.26	853.20	1388.00

Candidate selection. Each gene is sorted by a combined data of ChIP-seq data and RNA-seq data. Scores represent ChIP-seq H3K4me3 scores calculated by HOMER. The four FPKM data represent expression profiles retrieved from for cell lines HMEC, SK-BR-3, ZR-75-1 and MDA-MB-436. Ranked by score, potential up-regulating (left) and down-regulating (right) candidates were achieved.

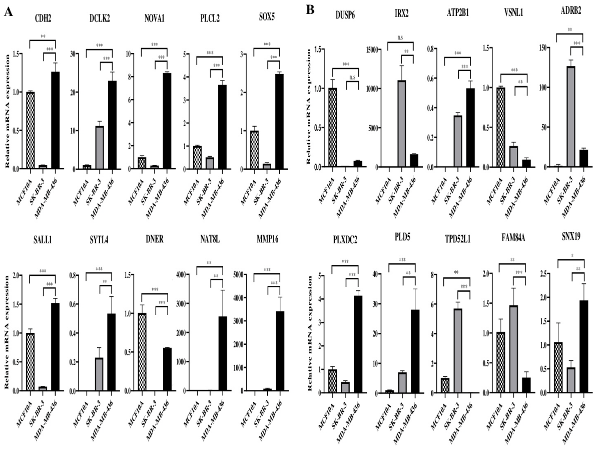

Fig. 2. RNA expression validation. The bar plots of relative RNA expression of 10 genes considered as up-regulating biomarkers (A) and down-regulating biomarkers (B) for the TNBC in MCF10A (black-stripes), SK-BR-3 (grey), and MDA-MB-436 (black) cell lines.

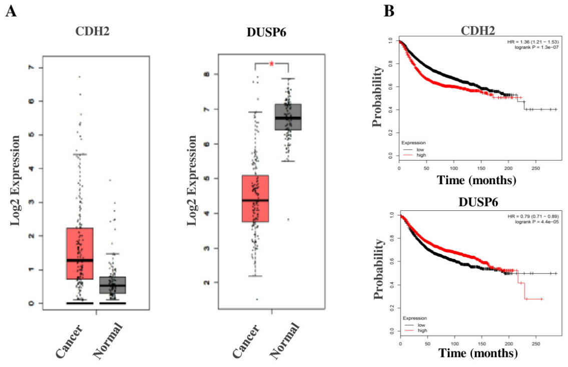

Fig. 3. Diagnostic and prognostic values of candidates. The highest ranked genes in up- and down-regulated were subjected to the TCGA data. (A) The gene expression box plot in the TNBC and basal-like breast cancer (red) and normal (grey). (B) The overall survival of breast cancer patients expressing CDH2 and DUSP6 (high: red, low: black) in the KM plot.

HER2+. Conversely, gene expressions down-regulated in the MDA-MB-436 cells selected from the combination of low H3K4me3 and high H3K27me3 were validated in four of 10 candidates (Fig. 2B). DUSP6 and VSNL1 gene expressions were significantly down-regulated in HER2+ and TNBC breast cancer cell lines. Only TPD52L1, and FAM84A were most significantly down regulated in TNBC compared to MCF10A and SK-BR-3. The expressions of ATP2B1, ADRB2, PLXDC2, PLD5 and SNX19 grouped in down-regulated genes were not correlated with histone states in TNBC and normal.

Clinical correlation of the genes selected by histone modification data: TCGA data correlation of up- and down-regulated genes and cancer patient survival data

To expand the RT-PCR validated gene set data to cancer patient data in the TCGA public domain, we analyzed the TCGA data for the selected gene. Unfortunately, since TCGA data have not been classified by TNBC, the gene expression pattern in overall breast cancer patients was not nicely correlated with the results in this study targeting TNBC. For example, gene expressions of NOVA1, SOX5 and NAT8L were found higher in the TNBC cell line than in the normal cell line while expressed lower in overall breast cancer than the healthy population (Fig. 2 and Fig. S2). This discrepancy may come from the absence of corresponding classifications in TCGA data, since these three genes whose expressions were upregulated in TNBC were lower in the other cancer cell line (SK-BR-3; HER2+) than the normal cell line.

To further correlate the candidate genes with cancer patient data and examine the prognostic value of the candidate genes in breast cancer patient databases, we used the expression box plots (Box-plots: http://gepia2.cancer-pku.cn) of cancer patients and the normal population (12) and their Kaplan-Meier plots (KMplot; https://www.kmplot.com). Aberrant gene expression and its influence on overall survival (OS) was presented in Fig. S2 and S3. Top two representative genes are shown in Fig. 3; CDH2 in up-regulated and DUSP6 in down-regulated. The CDH2 found as an up-regulated gene in MDA-MB-436 was highly expressed in basal-like and TNBC patients. Also, patients with higher CDH2 expression have high mortality when compared to low expressed patients (HR = 1.36, logrank P = 1.3e-07). Conversely, DUSP6 down-regulated in MDA-MB-436 presented significantly low expression levels in basal-like and TNBC patients when compared to healthy controls (Fig. 3A). Survival curves associated with DUPS6 gene expression indicated that lower DUPS6 expression in breast cancer patients has an association with worsening OS (Fig. 3B). Also, we implemented a classification model based on the expression of CDH2 and DUSP6 using 1,222 normal and breast cancer patients from the TCGA database (Fig. 3C). ROC curves from the individual genes had a high AUC curve with 79% in CDH2 and 92% in DUSP6. When these two genes were combined using the binary logistic regression method, AUC curve of sensitivity/1-specificity was up to 93%.

DISCUSSION

Recently, large amounts of omics data have been produced globally and made publicly available. Most are genomic, such as single nucleotide polymorphism (SNP), copy number variation (CNV), and transcriptomic data. But recently epigenomic data such as histone modification, methylation status, and chromatin profiles have seen a continuing increase in various cancer studies (13), since epigenetic mechanisms are recognized as critical risk factors in the development of cancers.

In this study, we developed a strategy to investigate triple negative breast cancer (TNBC) biomarkers based on the epigenome dataset of histone modifications (11). The combination of two different histone modification markers, high H3K4me3 and low H3K27me3 and vice versa, made significant improvement in the correlation with transcriptomic data compared to each marker only (Fig. S1). This strategy is similar to the concept of bivalent chromatin that has a role in developmental regulation in pluripotent cells and is defined wherein the region of DNA is bound to histone proteins with repressing and activating epigenetic regulators. However, there were some discrepancies such as the range of region modification that occurred and the combination of epigenetic markers.

We predicted most likely highly up-regulated or down-regulated genes in terms of transcriptome expression based on two histone marks of H3K4me3 and H3K27me3 and listed up top 10 candidates up-and down-regulated in TNBC. Then we tested the gene expression using RNA-seq data and quantitative Real-Time PCR in three breast cancer related cell lines (MCF-10A, MDA-MB-436 and SK-BR-3). Because of availability, the MCF-10A cell line was used as normal breast cell line in our study instead of HMEC used in RNA-seq and ChIP-seq data. This may present unexpected high expression levels of CDH2, SALL1 and DNER, and low expression of IRX2, PLXDC2 and SNX19 in the normal cell line. This result should be confirmed by extended numbers and types of cell lines to exclude cell line specific features.

The *in vitro* cell line analysis of histone modifications and gene expression was applied to the public clinical data to retrieve the prognostic significance of individual candidates. The top scored genes, CDH2 and DUSP6, up- and down-regulated respectively in TNBC, successfully represented aggressive pathological phenotype of TNBC which may directly link to general breast cancer patient's overall survival in TCGA expression plots and KM-plotter (Fig. 3).

Many of the candidate genes we selected have been studied regarding their breast cancer-related molecular functions. The remarkable increase of aspartate N-acetyltransferase (NAD8L) is reported to develop cancer growth in overall cancer types and is a valuable target for cancer treatment (14). Up-regulation of matrix metallopeptidase 16 (MMP16) from miR-155 is reported to enhance proliferation and migration in triple negative breast cancers. CDH2, commonly known as N-cadherin contributes significantly towards transitioning from the epithelial state to the mesenchymal state (EMT) and enacting abnormal cells to invade and metastasize to nearby as well as distant tissues. Sex determining region Y-box protein 5 (SOX5) expression is reported to increase EZH2 expression inducing breast cancer cell proliferation and invasion (15). Controversially, SALL1 is a tumor suppressor in luminal breast cancer types, as well as in TNBCs (16, 17). Notably, we newly identified four novel candidate genes never been reported in breast cancer (NOVA Alternative Splicing Regulator 1 (NOVA1), Phospholipase C Like 2 (PLCL2), Synaptotagmin Like 4 (SYTL4), and Delta/Notch Like EGF Repeat Containing (DNER)). Since NOVA1 (52.37%) and DNER (34.45%) have been studied in various other cancers, but not in breast cancer.

Breast cancers are continuously separated by different measures for more precise classification. We used the top-ranked genes in up- and down-regulated markers to observe if it could contribute to enhancing breast cancer classification. Each gene showed high differentiation, but the combination of differentially expressed candidate genes, predicted by H3K4me3 and H3K27me3 histone marks analysis, using the logistic regression models further improved the accuracy of breast cancer diagnosis.

CONCLUSION

We suggested a bio-informatical strategy to reveal TNBC biomarkers using histone modifications of H3K4me3 and H3K27me3 and combining transcriptomic datasets. The functional study of the candidate genes found in this study in breast cancer, especially in TNBC, is necessary in more extensive datasets and cancer types for better understanding and discovering novel biomarkers and therapeutic targets.

MATERIALS AND METHODS

DATA acquisition and bioinformatics analysis

H3K4me3 and H3K27me3 ChIP-seq data from human breast cancer cell lines, MDA-MB-436, SK-BR-3, ZR-75-1, and human normal breast cell line, HMEC, was obtained from the GEO database GSE62907 (11). RNA-seq data for all the cancer cell lines was also obtained from the same database. Human normal breast cell line HMEC RNA-seq data was obtained from dataset GSE62820 (18).

Each ChIP-seq raw dataset was aligned with human reference hg19 using the HISAT2. The peak finding was performed using the 'findPeaks' command of HOMER. Differential peaks of TNBC cell line MDA-MB-436 was analyzed by using HMEC ChIP-seq data as a control group. HOMER software command 'getDifferentialPeaks' was used to identify H3K4me3 enriched peaks, H3K27me3 repressed peaks for activated regions and H3K4me3 repressed, H3K27me3 enriched peaks for downregulated regions. The fold change cutoff was $\geq$ 4 for enriched and $\geq$ 2 for the repressed peak regions. Annotation of all regions was performed using the 'annotatePeaks.pl' function of HOMER. Within the annotated list, H3Kme3 regions associated with transcription such as TSS, promoter, and exon regions were selected as potential targets. RNA-seq data was aligned and peak analysis performed using the HOMER transcriptome analysis pipeline. From the ChIP-seq sorted genes, candidates were selected by matching profiles that were high-expressed or low-expressed specifically in the MDA-MB-436 dataset.

Among the histone modifications upregulation of histone H3K4me3 and downregulation of histone H3K27me3 were selected for the peak comparison. After normalization, TNBC H3K4me3 peak data was compared against HMEC H3K4me3 peak data for the differential histone enriched regions. To identify statistically high or low enriched regions, HMEC and MDA-MB-436 ChIP-seq data was used as control groups and experimental groups. As for the potential upregulated regions, only locations in TNBC peaks enriched more than four-fold compared to the HMEC ChIP-seq data and HMEC H3K27me3 locations with a fold enrichment more than two compared to HMEC were sorted (Fig. 1A). The opposite method was implemented to sort potential down-regulated regions. HMEC H3K4me3 peaks that were four-fold higher than TNBC and H3K27me3 peaks of TNBC two-fold higher than HMEC were selected. Because the H3K27me3 profile is dispersed across the entire gene structure, highly enriched peaks are difficult to locate. As a result, histone modification H3K27me3 are sorted by a two-fold degree. After differential analysis, sorted regions were annotated with gene names and enriched gene positions. Among the DNA structure, H3K4me3 regions with expression influence were selected as potential histone enriched regions; promoter, transcriptional start sites, and exon.

Correlation of genes matching the up or down regulating prediction was calculated by comparing each individual ChIP-seq peak log2 foldchange and its matching RNA-seq log2 expression foldchange. The combined method of sorting candidate genes using H3K4me3 and H3K27me3 histone modification was compared with the methods that sorted the genes using only H3K4me3 or H3K27me3. The accuracy was estimated by a percentage of genes that matched its predicted RNA expression pattern.

Cell culture

MCF-10A normal cell line was cultured in Mammary Epithelial Cell Growth Basal Medium (MEBM) BulletKit (Lonza cat # CC-3150) with an additional 10% fetal bovine serum (FBS) and 1% Antibiotic-Antimycotic product. The cell line SK-BR-3 was cultured using the RPMI media with an additional 10% FBS and 1% Antibiotic-Antimycotic product. MDA-MB-436 was cultured in the DMEM media with an additional 10% FBS and 1% Antibiotic-Antimycotic product.

Quantitative RT-qPCR

The RNA isolation was processed using the Rneasy Plus Mini Kit (Qiagen). The genomic DNA contamination was eliminated by using the gDNA elimination columns. Also, 2 μg of the total RNA was used for the cDNA synthesis using the OMMISCRIPT RT KIT (Qiagen, CA). The primers for each target gene were designed spanning two different exons (Table S3). The real-time PCR was performed using the CFX96 Touch Real-Time PCR Detection System (Bio-Rad). The relative gene expression was measured by the delta-delta CT method. The data were normalized to the 18srRNA.

Expression box plot analysis

The gene expression data in the TCGA cancer patients samples were analyzed with GEPIA (http://gepia2.cancer-pku.cn). The gene expression in the normal data was compared only with the basal-like and triple negative breast cancer subtypes. The log2 foldchange cutoff was set to 1. The p-value cutoff was set to less than 0.01.

Kaplan-Meier plot analysis

The web-based Kaplan–Meier plotter was used to evaluate the effect of candidate genes on survival rates in more than 3,000 breast cancer samples. The hazard ratio (HR) was given with 95% confidence intervals, and log rank P value was calculated and displayed on the web page. The log rank P-values were calculated by auto-selecting the best cutoff option. The affymetrix ID of the top 10 potential up- and down-regulated candidates were listed in Table S2.

ROC analysis

Classification using receiver operating characteristic (ROC) curves was performed using 1,222 normal and breast cancer patients in the TCGA database. The area under the curve (AUC) scores and p-values were calculated using the easyROC web-based tool. The gene set combined logistic regression model was achieved using SPSS statistical analysis software.

AUTHOR CONTRIBUTIONS

J.Y. Cho conceived the presented idea. H.M Park developed theory and performed major computations. H.M Park and K.H. Lee wrote the manuscript and H.S. Kim performed cell culture and verified RNA-seq expression data. K.H. Lee provided scientific discussion. All authors discussed the results and contributed to the final manuscript.

ACKNOWLEDGEMENTS

This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Ministry of Science and ICT (#2016M3A 9B6026771).

CONFLICTS OF INTEREST

The authors have no conflicting interests.

REFERENCES

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E and Forman D (2011) Global cancer statistics. CA Cancer J Clin 61, 69-90
2. O'Brien KM, Cole SR, Tse CK et al (2010) Intrinsic breast tumor subtypes, race, and long-term survival in the Carolina Breast Cancer Study. Clin Cancer Res 16, 6100-6110
3. Perou CM (2011) Molecular stratification of triple-negative breast cancers. Oncologist 16 Suppl 1, 61-70
4. Carey LA, Perou CM, Livasy CA et al (2006) Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA 295, 2492-2502
5. Jones PA and Baylin SB (2007) The epigenomics of cancer. Cell 128, 683-692
6. Chen X, Hu H, He L et al (2016) A novel subtype classification and risk of breast cancer by histone modification profiling. Breast Cancer Research and Treatment 157, 267-279
7. Koch CM, Andrews RM, Flicek P et al (2007) The landscape of histone modifications across 1% of the human genome in five human cell lines. Genome Res 17, 691-707
8. Barski A, Cuddapah S, Cui K et al (2007) High-resolution profiling of histone methylations in the human genome. Cell 129, 823-837
9. Xi Y, Shi J, Li W et al (2018) Histone modification profiling in breast cancer cell lines highlights commonalities and differences among subtypes. BMC Genomics 19, 150
10. Gomez-Cabrero D, Abugessaisa I, Maier D et al (2014) Data integration in the era of omics: current and future challenges. BMC Syst Biol 8 Suppl 2, I1
11. Chaligne R, Popova T, Mendoza-Parra MA et al (2015) The inactive X chromosome is epigenetically unstable and transcriptionally labile in breast cancer. Genome Res 25, 488-503
12. Tang Z, Li C, Kang B, Gao G, Li C and Zhang Z (2017) GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res 45, W98-W102
13. Nebbioso A, Tambaro FP, Dell'Aversana C and Altucci L (2018) Cancer epigenetics: Moving forward. PLoS Genet 14, e1007362
14. Zand B, Previs RA, Zacharias NM et al (2016) Role of Increased n-acetylaspartate Levels in Cancer. J Natl Cancer Inst 108, djv426
15. Sun C, Ban Y, Wang K, Sun Y and Zhao Z (2019) SOX5 promotes breast cancer proliferation and invasion by transactivation of EZH2. Oncol Lett 17, 2754-2762
16. Wolf J, Muller-Decker K, Flechtenmacher C et al (2014) An in vivo RNAi screen identifies SALL1 as a tumor suppressor in human breast cancer with a role in CDH1 regulation. Oncogene 33, 4273-4278
17. Ma C, Wang F, Han B et al (2018) SALL1 functions as a tumor suppressor in breast cancer by regulating cancer cell senescence and metastasis through the NuRD complex. Mol Cancer 17, 78
18. Rahman M, Jackson LK, Johnson WE, Li DY, Bild AH and Piccolo SR (2015) Alternative preprocessing of RNA-Sequencing data in The Cancer Genome Atlas leads to improved analysis results. Bioinformatics 31, 3666-3672