

RESEARCH PAPER



Lowly methylated region analysis identifies EBF1 as a potential epigenetic modifier in breast cancer

Nora Fernandez-Jimenez[†], Athena Sklias, Szilvia Ecsedi, Vincent Cahais, Davide Degli-Esposti, Antonin Jay, Pierre Benoit Ancey, Hae Dong Woo, Hector Hernandez-Vargas, and Zdenko Herceg[†]

Epigenetics Group, International Agency for Research on Cancer (IARC), Lyon Cedex 08, France

ABSTRACT

Breast cancer (BC) encompasses heterogeneous pathologies with different subtypes exhibiting distinct molecular changes, including those related to DNA methylation. However, the role of these changes in mediating BC heterogeneity is poorly understood. Lowly methylated regions (LMRs), non-CpG island loci that usually contain transcription factor (TF) binding sites, have been suggested to act as regulatory elements that define cellular identity. In this study, we aimed to identify the key subtype-specific TFs that may lead to LMR generation and shape the BC methylome and transcription program. We initially used whole-genome bisulfite sequencing (WGBS) data available at The Cancer Genome Atlas (TCGA) portal to identify subtype-specific LMRs. Differentially methylated regions (DMRs) within the BC PAM50 subtype-specific LMRs were selected by comparing tumors and normal tissues in a larger TCGA cohort assessed by HumanMethylation450 BeadChip (450K) arrays and TF enrichment analyses were performed. To assess the impact of LMRs on gene expression, TCGA RNA sequencing data were downloaded and Pearson correlations between methylation levels of loci presenting subtype-specific TF motifs and expression of the nearest genes were calculated. WGBS methylome data revealed a large number of LMRs for each of the BC subtypes. Analysis of these LMRs in the 450K datasets available for a larger sample set identified 7,765, 5,657, and 19 differentially methylated positions (DMPs) between normal adjacent tissues and tumor tissues from basal, luminal, and HER2-enriched subtypes, respectively. Unsupervised clustering showed that the discriminatory power of the top DMPs was remarkably strong for basal BC. Interestingly, in this particular subtype, we found 4,409 differentially hypomethylated positions grouped into 1,185 DMRs with a strong enrichment for the early B-cell factor 1 (EBF1) motifs. The methylation levels of the DMRs containing EBF1 motifs showed a strong negative correlation with the expression of 719 nearby genes, including *BTS2* and *CD74*, two oncogenes known to be specific for basal BC subtype and for poor outcome. This study identifies LMRs specific to the three main BC subtypes and reveals EBF1 as a potentially important regulator of BC subtype-specific methylation and gene expression program.

Abbreviations: BC, breast cancer; CGI, CpG island; DMP, differentially methylated position; DMR, differentially methylated region; EBF1, early B cell factor 1; ER, estrogen receptor; LMR, lowly methylated region; TCGA, The Cancer Genome Atlas; TF, transcription factor; TFBS, transcription factor binding site; TSS, transcription start site; WGBS, whole-genome bisulfite sequencing

ARTICLE HISTORY

Received 9 June 2017
Revised 21 August 2017
Accepted 28 August 2017

KEYWORDS


breast cancer subtypes; DNA methylome; EBF1; lowly methylated regions; transcription factor binding sites

Background

Despite an important progress in prevention and surveillance, breast cancer (BC) remains the most common invasive cancer and cause of cancer-related mortality of women worldwide.^{1,2} Therefore, improving our understanding of the mechanisms underlying BC onset and progression, as well as tumor phenotypes, is fundamental to improving our abilities to treat and prevent the disease. Clinically, breast cancer encompasses heterogeneous pathologies, and the traditional categorization of human breast cancer on the basis of histological criteria is confounded by different factors, including intra-tumor heterogeneity, lack of understanding of the target cells, and absence of common genetic events.³

Previous molecular studies of breast cancer have primarily focused on mRNA expression profiling and genetic changes, including mutation and DNA copy number analysis.^{4,5} Beside genetic changes, epigenetic mechanisms are essential for development and cell differentiation and seem to be critical for the integration of endogenous and environmental signals during the life of a cell or an organism.^{6,7} By analogy, deregulation of epigenetic mechanisms has been associated with a variety of human malignancies, including BC.^{8–10} Technological advances in epigenomics have created the opportunity to comprehensively characterize the epigenetic landscape of this disease,¹¹ and recent DNA methylome profiling of a large series of breast tumors suggested the

CONTACT Zdenko Herceg  hercegz@iarc.fr; Nora Fernandez-Jimenez  nora.fernandez@ehu.es  Epigenetics Group, International Agency for Research on Cancer (IARC), 150 Cours Albert Thomas, F-69008, Lyon, France.

 Supplemental data for this article can be accessed on the [publisher's website](#).

[†]Equal senior authorship.

© 2017 Taylor & Francis Group, LLC

existence of previously unrecognized BC types beyond the currently known expression subtypes.^{12,13}

This and other studies of BC samples demonstrated that DNA methylation changes are ubiquitous in primary BC cells and that the phenotypic diversity may reflect intra-tumor (epi) genetic heterogeneity. The list of DNA methylation changes in this disease has further increased with the completion of major international sequencing initiatives; however, the underlying mechanisms and functional importance of DNA methylation changes have not been systematically evaluated.^{8,14}

DNA methylation is dynamically regulated at tissue-specific promoters and *cis*-regulatory elements during cell differentiation. A detailed analysis of these dynamic, differentially methylated regions (DMRs) indicated that DMRs are often short regions of DNA enriched in transcription factor binding sites (TFBS).¹⁵ This observation points to the role of TF binding events in reshaping the DNA methylation profile in normal and cancer cells.¹⁶ This notion is further supported by the identification of the lowly methylated regions (LMRs), the localized CpG-poor distal regulatory regions exhibiting an average methylation of 30%, DNase I hypersensitivity, and presence of enhancer chromatin marks, that tend to be occupied by cell type-specific TFs.¹⁷

Several lines of evidence point to the role of TF occupancy in heterogeneous methylation at LMRs. Feldmann et al.¹⁸ found that binding strength of transcriptional repressor CTCF is inversely correlated with DNA methylation within the CTCF motifs, and that 5-hydroxymethylcytosine, an intermediate modification during active demethylation, is enriched at LMRs. These findings are consistent with a model in which TFs and other DNA-binding factors can mediate DNA methylation levels as a mechanism for maintaining and reprogramming gene regulatory regions.¹⁸ In addition, many TFBSs identified in a particular LMR cluster were found highly enriched in the most highly expressed genes of specific cancer subtypes, demonstrating a key role for these TFs in determining subgroup identity.¹⁹

In the present study, we aimed to investigate the presence of LMRs in the main BC subtypes and identify the most relevant TFs that could be involved in BC-related LMRs. We also aimed to ascertain whether these specific regions have a functional impact on the gene expression program and the molecular subtypes of the disease.

Methods

TCGA data

The outline of our *in silico* analysis is summarized in Fig. 1. We first downloaded whole-genome bisulfite sequencing (WGBS) data of BC samples available at TCGA data portal (<https://tcga-data.nci.nih.gov/docs/publications/tcga/>) for LMR identification. This included 5 WGBS methylomes, two of which were luminal, two HER2-enriched and one basal breast cancer subtype (Supplementary File 1). The characteristics of the cohort for which Illumina HumanMethylation450 BeadChip (450K) data were available are summarized in Supplementary File 2. These include TCGA barcodes, PAM50 classification, and the codes of the samples used for meta-analysis. Out of the total 614 samples, only those for which we had PAM50 classification

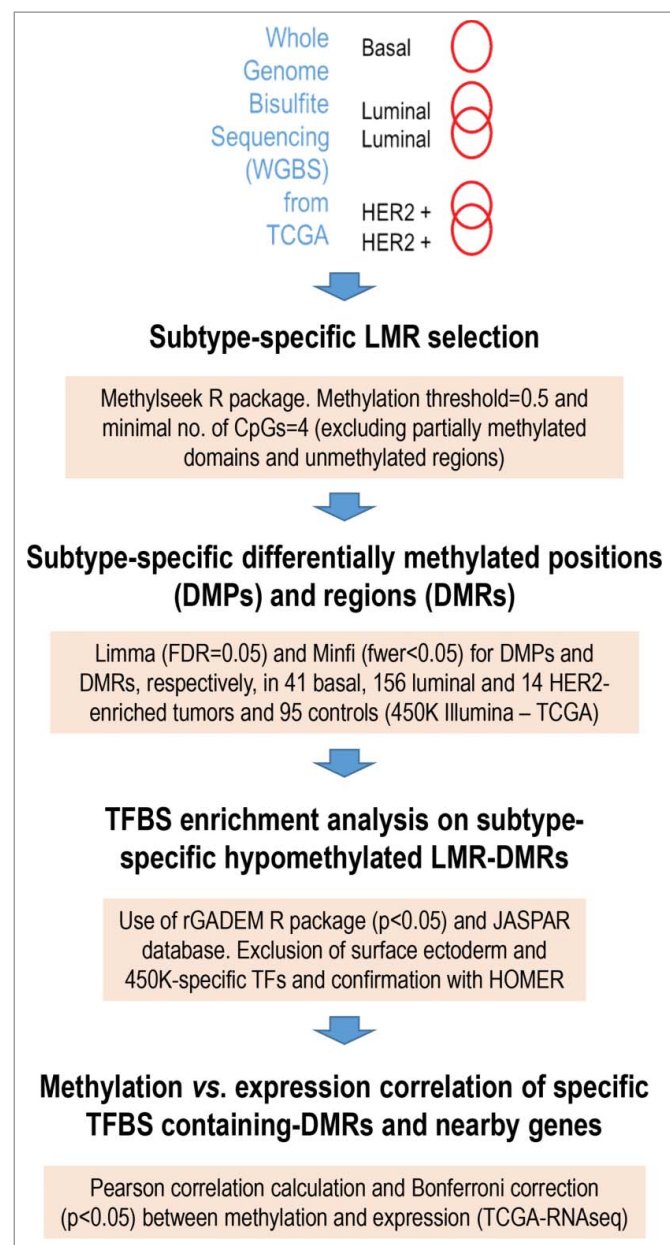


Figure 1. Pipeline of the *in silico* approach. Information on the packages, thresholds, and general criteria used for the analysis are provided.

data were used in the 450K subtype-specific analysis. This includes 41 basal, 156 luminal, and 14 HER2-enriched tumors as well as 95 surrounding normal tissues.

LMR identification

LMR identification was carried out by using the MethylSeekR package.²⁰ This package allows for the identification of active regulatory regions from high-resolution WGBS methylomes and relies on the idea of transcription factor binding leading to defined reduction in DNA methylation. In particular, all WGBS datasets available at TCGA have a coverage per sample greater than 15x, as was the case of the samples used in this study. Methylation threshold was established at 0.5 (50% methylation), as suggested by developers, with a minimal number of 4 consecutive CpGs. Partially methylated and unmethylated regions were excluded from further analyses. In the BC

subtypes in which we had two samples, only overlapping LMRs were taken into consideration to increase the chances of finding differential methylation in a larger cohort.

Differential methylation analysis

The analysis of 450K for the assessment of differential methylation between tumors and normal adjacent mucosa in the union LMRs was performed in the whole sample collection for which data was available (614 samples – 95 adjacent mucosa/519 tumor). Subtype-specific methylation analysis was carried out comparing each subtype (41 basal-like, 14 HER2-enriched, 156 luminal) to the same 95 normal sample collection in a non-matched way.

All analyses were performed with Limma [false discovery rate (FDR) = 0.05] and Minfi (fwer < 0.05) R packages²¹ for DMP and DMR assessment, respectively. Statistically significant DMRs with at least two consecutive CpGs included in a bookend of 1,000 nucleotides were retained for further analysis.

Transcription factor enrichment analysis of the subtype-specific differentially methylated LMRs

TFBS enrichment analysis was performed using rGADEM R package (P value < 0.05)²² and JASPAR database (<http://jaspar.genereg.net/>) with 200 bp flanking regions downstream and upstream of each DMR, and confirmed with HOMER tool (<http://homer.salk.edu/homer/motif/>) using standard thresholds.

To avoid picking a candidate enriched by *default* in the array, we built up random DMR sets of the same size as those found in each comparison (tumor vs. normal for each subtype) and run the TFBS enrichment analysis with the parameters described above. Resulting TFs were excluded from further analyses. Additionally, we compared our candidate TF list to the previously published surface ectoderm-specific TF set,²⁰ to discard those TFs that could be enriched in our DMRs due to cell type-of-origin and not to disease.

Assessment of the correlation between differential methylation and RNA sequencing expression data

To examine whether differentially methylated CpGs were correlated with the expression of adjacent genes, we used the RNA sequencing data of 594 samples out of the previously described 614 samples for which expression data was available (Supplementary File 3). Methylation and expression level correlation was assessed by Pearson correlation calculation followed by Bonferroni correction (P value < 0.05). All the TCGA data used in the present study was normalized, level 3 data. All pathway analyses were performed at <https://david.ncicrf.gov/>.

Results

Identification of differentially methylated BC-LMRs

We first downloaded WGBS data for different subtypes of BC available at TCGA data portal, among which 2 were luminal, 2 were HER2-enriched and one basal BC subtype (Fig. 1 and Supplementary File 1 and 2). Our analysis of WGBS methylation data by the MethylSeekR package identified the presence

of a large number of LMRs that are characterized by 30% median methylation levels as well as low CpG density, unlike fully unmethylated regions, usually present at CpG islands (CGI) and high density regions (Fig. 2A).

In particular, identifying the intersection among the five WGBS datasets provided a list of 25,923 union LMRs that overlapped 37,211 probes in the 450K methylation array. Interestingly, 23,783 out of these 37,211 probes were differentially methylated when their methylation levels were assessed in a larger 450K cohort, as described in Methods. In summary, the unsupervised clustering performed in the differentially methylated probes demonstrated that basal tumors clustered together, while luminal A and B appeared to be mixed. The rest of the subtypes were poorly represented (Supplementary Figure 1A). Additionally, BC-related differentially methylated LMRs seemed to be enriched in TFBSs as well as in intergenic and promoter-distal regions when compared to a random differentially methylated probe set of the array or to the whole 450K probe collection (Supplementary Figure 1B and C).

Identification of differentially methylated basal BC-specific LMRs

Given that the unsupervised clustering of the differentially methylated union LMRs revealed a marked subtype specificity, we decided to focus on LMRs specific for each of the BC subtypes, which, however, presented a significant overlap among them (Fig. 2B). Overlap between samples of the same subtype was greater for the two luminal tumors (38,335 common LMRs out of the 59,925 and the 67,044 called in individual samples) when compared to HER2-enriched (25,101 common LMRs out of the 41,830 and the 61,409 called in individual samples).

In this context, we performed a case vs. control differential methylation analysis in those regions overlapping between the subtype-specific LMR sets in WGBS data and those positions present in the 450K array. Therefore, we compared cases and controls only in those probes available at the array that were previously identified to be located in at least a subtype-specific LMR. To this end, we downloaded all the 450K methylation datasets for which PAM50 BC classification data was available. The analysis included 41 basal, 156 luminal, and 14 HER2-enriched tumors as well as 95 normal tissues.

Our analysis revealed that 13,848, 10,436, and 7,770 regions were common between WGBS-identified LMRs and 450K probes for basal, luminal, and HER2-enriched subtypes, respectively. Out of these, 7,765, 5,657, and 19 were differentially methylated when comparing individual BC subtypes to normal tissue samples. We further performed an unsupervised clustering with the top DMPs included for each of the groups and found that the identified changes were capable to discriminate between cases and controls. Again, the discriminating power of the top DMPs was particularly strong for basal BC datasets (Fig. 2C) and, therefore, we decided to focus on this subtype for further analyses.

Differentially methylated LMRs are enriched in EBF1 motifs

We found that 4,409 out of the total 7,665 DMPs were hypomethylated in basal BC subtype (Supplementary File 4) and that these hypomethylated sites were clustered into 1,185 DMRs

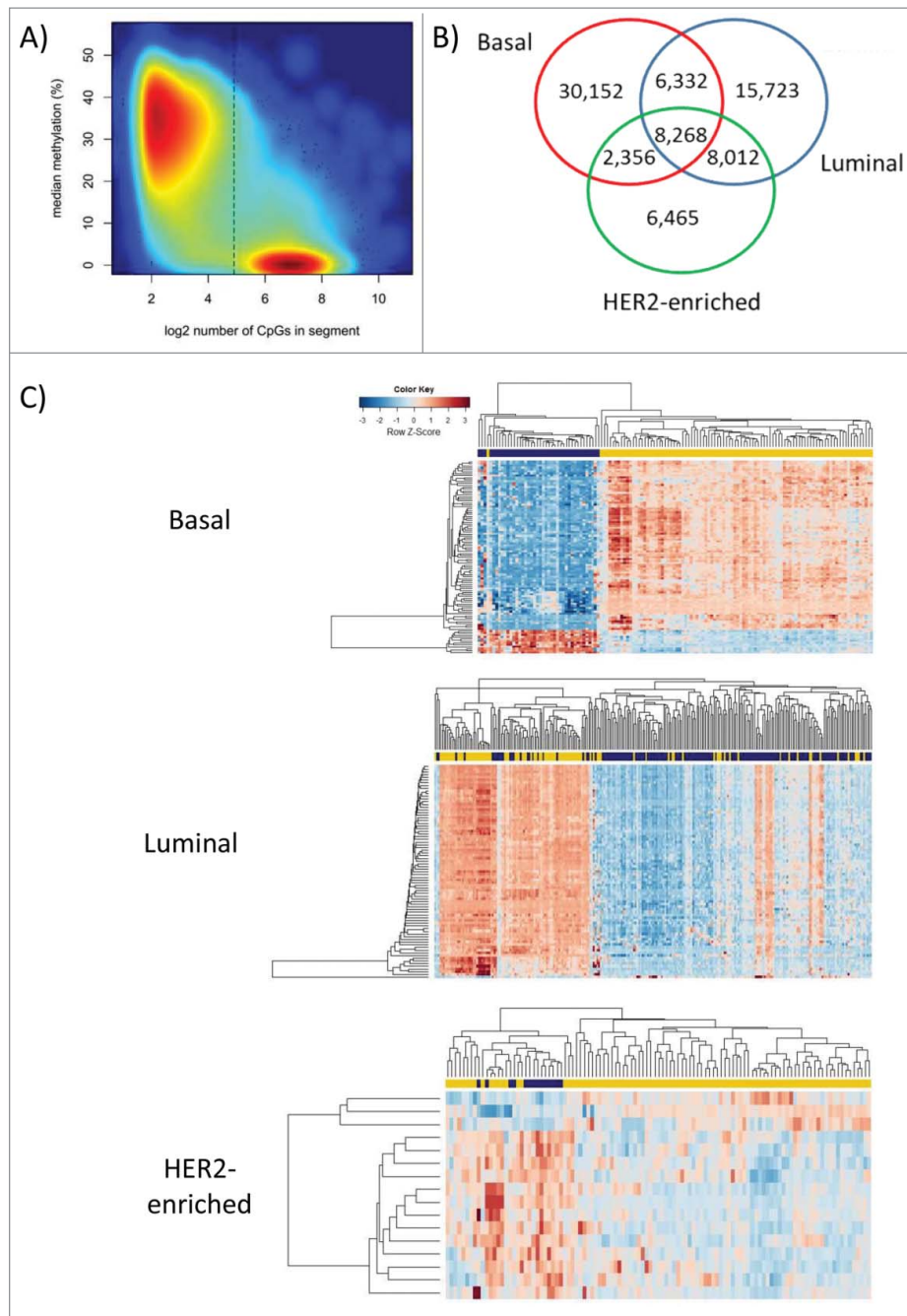


Figure 2. (A) CpG density plot per median methylation percentage showing the LMRs (upper left high-density cluster) and the unmethylated loci (bottom right high-density cluster). The latter are excluded from further analyses. The red to blue color gradient represents higher to lower CpG density. (B) Venn-Diagram showing BC subtype-specific LMRs. (C) Heatmap showing BC subtype-specific top differentially methylated loci in previously identified LMRs, with tumors (blue) and adjacent mucosa (yellow) in columns. The red to blue color gradient represents higher to lower methylation. Specifically, the 100 DMPs showing the lowest FDR corrected P values are shown for basal and luminal subtypes (all the DMPs including these are $FDR < 0.05$). In the case of HER2-enriched subtype, we show all the $FDR < 0.05$ DMPs identified (19 in total).

(Supplementary File 5). To further investigate the relationship between LMRs and TF binding in specific BC subtypes, we next performed a TFBS enrichment analysis on the hypomethylated DMRs in basal BC. Our analysis identified a high enrichment in EBF1 motifs (P value = $8.88e-0.5$) in the basal hypomethylated DMRs. The binding sites specific for other TFs, including those related to surface ectoderm-derived cells (EWSR1-FLI1, TFAP2A, SP1, ESR1, and Myf), were also identified (Table 1). These results were confirmed with two different methods, rGADEM and HOMER, as described in the Methods section. We then identified the hypomethylated loci containing EBF1

motifs in our dataset. In particular, we identified those 450K probes contained in the hypomethylated regions where EBF1 was predicted to be able to bind in the previous TFBS enrichment analysis. In total, we obtained 3,706 individual CpG positions for which we identified the nearest gene (Supplementary File 6).

Differentially methylated LMRs containing EBF1 motifs have an effect in nearby gene expression

To assess whether methylation levels at LMRs containing EBF1 motifs are associated with gene expression, we downloaded the

Table 1. Results of the transcription factor enrichment analysis performed in the hypomethylated LMR-DMRs. *SE = surface ectoderm.

Trancription Factor	Motif	p value	SE-specific*	Luminal
EBF1		8.88e-05	no	no
ESR1		8.04e-04	yes	no
EWSR1-FLI1		5.19e-08	yes	no
FEV		2.24e-02	no	no
Hand1::Tcf2a		2.24e-02	no	no
IRF1		1.62e-02	no	yes
IRF2		3.40e-02	no	no
Myf		9.84e-06	yes	no
NHLH1		7.75e-04	no	yes
NR2F1		2.73e-03	no	no
PLAG1		1.03e-02	no	no
PPARG::RXRA		5.62e-04	no	no
SP1		3.30e-05	yes	no
SPI1		8.80e-03	no	yes
Stat3		1.30e-02	no	no
Tal1::Gata1		3.95e-03	no	yes
TFAP2A		1.97e-03	yes	no
TLX1::NFIC		1.23e-02	no	no
Zfp423		9.31e-03	yes	no

RNA sequencing data available at TCGA portal for up to 594 samples with 450K array information and calculated pairwise Pearson correlations in the entire datasets (Supplementary File 7). Our analysis revealed that the methylation levels of 719 positions were inversely correlated with the expression level of their nearest gene, while 315 were positively correlated. We found an overrepresentation of transcription start sites (TSS) and promoter regions in negatively correlated positions, while gene body regions were enriched in those showing positive correlations with gene expression, with a decrease in promoters (Supplementary Figure 1).

To gain an insight into the identity of the genes and the pathways that may be deregulated by methylation changes in basal BC, we performed a pathway enrichment analysis of those genes whose expression levels were correlated to differentially methylated LMRs, as described in Methods. Our analysis revealed that small GTPase-mediated signal transduction, programmed cell death, and cell proliferation were among the top enriched pathways (Table 2). Among the top genes whose DNA methylation was negatively correlated with expression we found the *BST2* gene (Fig. 3) and the *CD74* gene (Supplementary

Figure 2), whereas among those positively correlated, we identified *ZBTB20* (Fig. 4) and *KLF6* (Supplementary Figure 3). Consistent with our previous observations, the regions negatively correlated with gene expression tended to be longer and contained a higher number of differentially methylated CpG positions. We also found that these regions are frequently located in promoters, in close proximity to TSSs (Supplementary Figure 1). In contrast, the DMRs positively correlated with expression contained fewer CpGs and were enriched in intergenic and gene body regions. These results are coherent with previous knowledge on methylation-related gene expression regulation.

Discussion

TFs are key players of gene expression program, although the determinants of their binding to DNA sequence motifs in a given cell type remain poorly understood. DNA methylation in the gene promoter region is generally associated with gene silencing; however, it remains debated whether this is due to inhibition of transcription factor binding.¹⁵ Intriguing recent evidence suggested that some TFs may bind specific methylated

Table 2. Pathway analysis of the genes correlated to nearby CpG methylation levels within EBF1 motif-containing basal LMRs. Count and % refer to the number and percentage of genes identified in our list for each corresponding term.

Category	Term	Count	%	P value	Benjamini
GOTERM_BP_FAT	GO:0051056~regulation of small GTPase mediated signal transduction	27	0.39370079	2.82E-06	0.00758966
GOTERM_BP_FAT	GO:0010941~regulation of cell death	56	0.8165646	1.43E-05	0.00961764
GOTERM_BP_FAT	GO:0043067~regulation of programmed cell death	56	0.8165646	1.28E-05	0.01145541
GOTERM_BP_FAT	GO:0042981~regulation of apoptosis	56	0.8165646	9.68E-06	0.0129961
GOTERM_BP_FAT	GO:0008283~cell proliferation	35	0.51035287	4.24E-05	0.02268137
GOTERM_BP_FAT	GO:0007242~intracellular signaling cascade	75	1.0936133	5.10E-05	0.02271459
GOTERM_BP_FAT	GO:0006915~apoptosis	43	0.62700496	7.40E-05	0.02818764
GOTERM_BP_FAT	GO:0012501~programmed cell death	43	0.62700496	1.04E-04	0.03440995

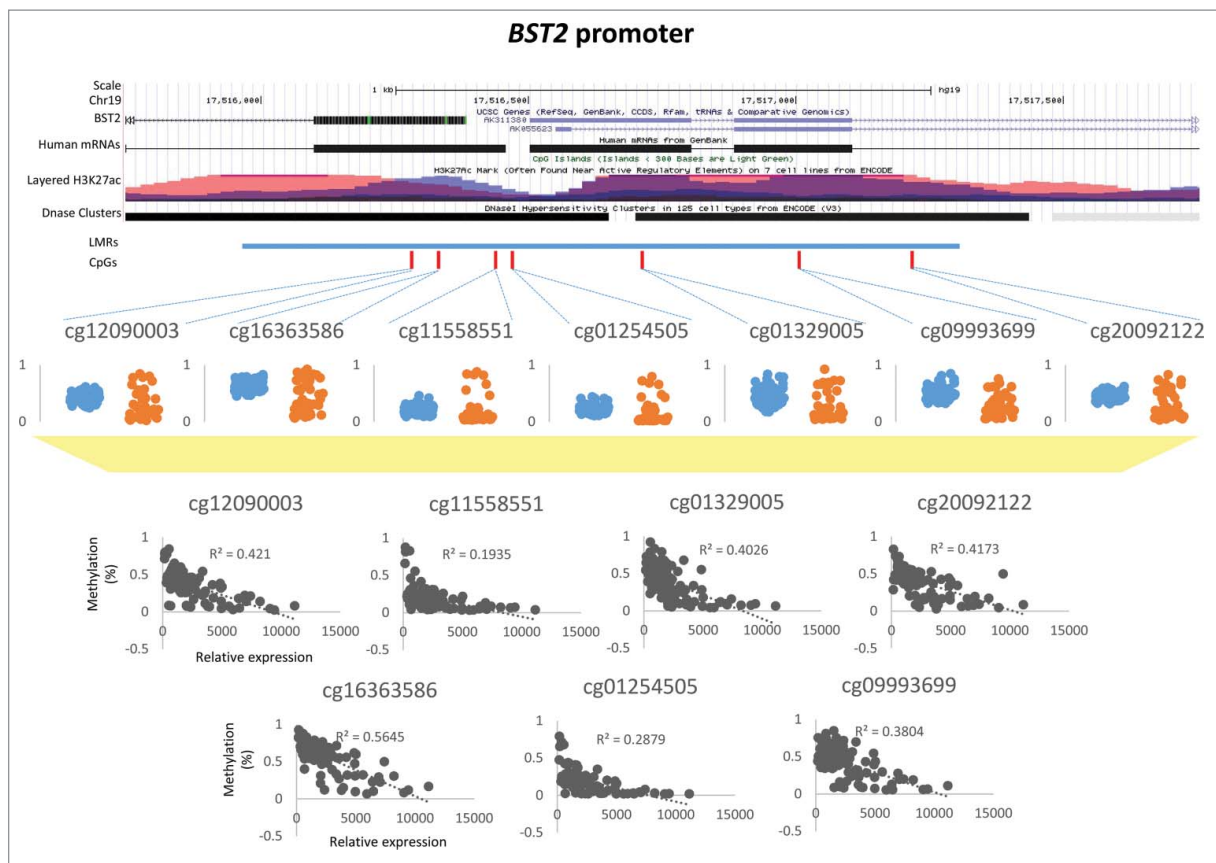


Figure 3. Example of a gene whose expression is negatively correlated to the methylation level of an adjacent differentially methylated LMR. *BST2* promoter showing transcripts, layered H3K27Ac activating marks, and DNase I hypersensitivity clusters (grey gradient according to sensitivity level) in the region. LMRs previously identified in BC subtype-specific WGBS data are shown in blue, while CpGs represented in the 450K array overlapping those regions are shown in red. Methylation differences per CpG site are shown in individual graphs, with adjacent tissues in blue dots and tumors in orange. All the differences observed are significant (FDR<0.05). Correlation between those methylation levels and the expression of the *BST2* gene is shown in the bottom graphs, with grey dots. Pearson correlation coefficients (all *P* value <0.05) are shown in each graph.

regions in the genome and trigger their demethylation.^{17,23} Furthermore, it has been suggested that binding and activity of some TFs that are sensitive to DNA methylation relies on additional determinants to induce local hypomethylation.²⁴ Among these additional factors there is a subset that possess the remarkable ability to activate their target genes in closed chromatin, thus behaving as pioneer factors to initiate local demethylation in such chromatin configuration.^{25,26} Since pioneer transcription factors play a key role in the establishment and maintenance of gene expression programs, their deregulation could trigger aberrant cell reprogramming and emergence of functional sub-populations, including those with metastatic capabilities or stem cell-like features.

In the present work, we first identified the union LMRs that overlapped the five WGBS sample sets taken individually, so that the analysis was related to BC in general and completely independent from disease subtype. As one could expect, LMRs were located far away from promoters, in open sea regions and enriched in TFBSs. Additionally, the top differentially methylated LMRs conducted a clustering where basal tumors clearly separated from the rest of the tumors, while luminal A and B samples appeared to be mixed. Therefore we decided to identify the LMRs specific to the three main BC subtypes, consistent with the idea that these regions may be important in mediating TF-induced regulation of the methylome and emergence of functional subpopulations of some types of tumors.

Again, methylation states at subtype-specific LMRs with the dominance of hypomethylation showed the capacity to discriminate tumors from normal, adjacent mucosa within different BC subtypes, a feature that was particularly evident for basal BC. This notion is supported by the finding that basal and luminal subtypes showed a marked deregulation of specific genomic regions (characterized by predominant DNA hypomethylation), whereas HER2-enriched subtype presents less numerous, mostly hypermethylated, changes in LMRs. However, the limited number of HER2-enriched tumors in our datasets is an important factor to take into consideration and could be responsible for the low number of significant hits found for this subtype.

Our finding of a notable enrichment of EBF1 motifs in our basal-specific hypomethylated LMRs suggests that the TF EBF1 could play a role in the subtype-specific methylation profile. EBF1 is a member of TF network that together with E2A and Foxo1 orchestrates B cell fate,²⁷ and it has been suggested that the enrichment of EBF1 motifs around hypomethylated sites may imply the role of its occupancy in regulating methylation levels.²⁸ In addition, EBF1 has recently been implicated in shaping the chromatin landscape in the context of B cell programming and proposed to be a pioneer TF.²⁹ Moreover, EBF1 has been identified as a potential player in sequence-specific regulation of methylation in other models and different cancer types. Meta-analysis of acute myeloid leukemia, low-grade glioma

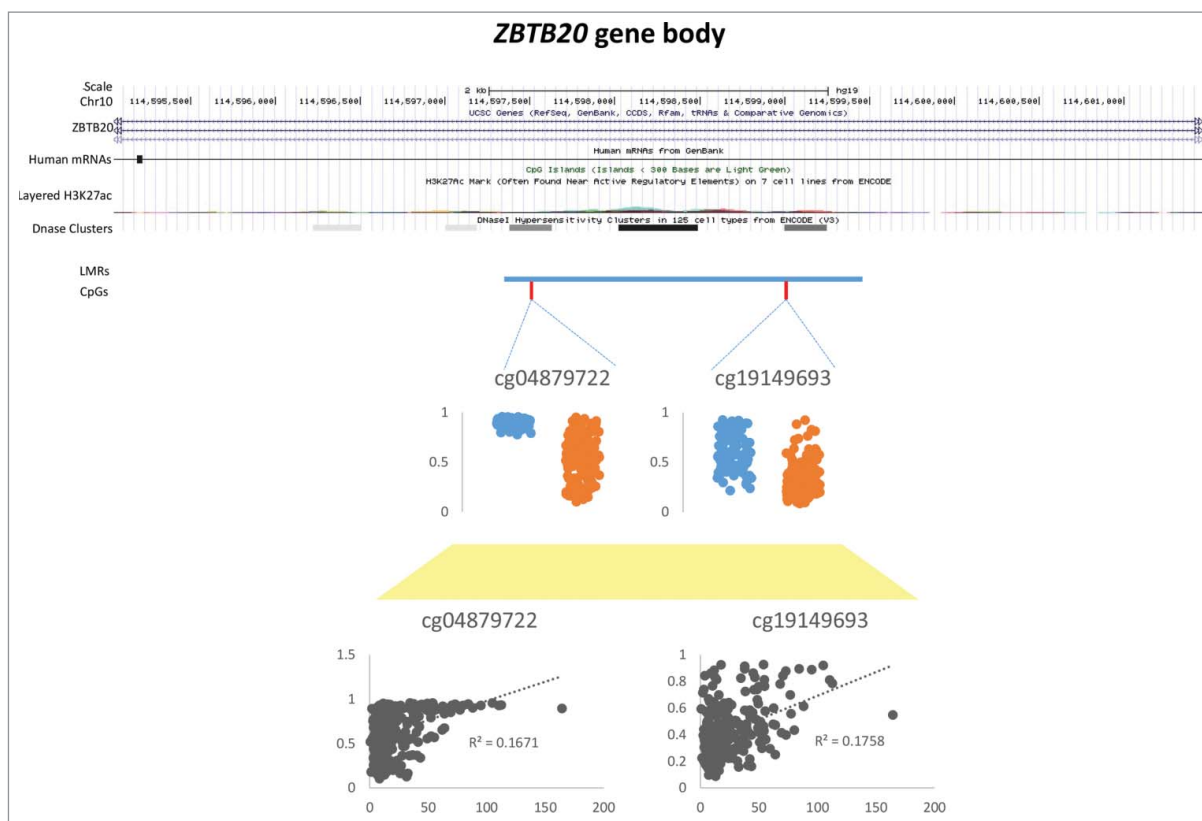


Figure 4. Example of a gene whose expression is positively correlated to the methylation level of an adjacent differentially methylated LMR. *ZBTB20* gene body showing transcripts, layered H3K27Ac activating marks, and DNase I hypersensitivity clusters (grey gradient according to sensitivity level) in the region. LMRs previously identified in BC subtype-specific WGBS data are shown in blue, while CpGs represented in the 450K array overlapping those regions are shown in red. Methylation differences per CpG site are shown in individual graphs, with adjacent tissues in blue dots and tumors in orange. All the differences observed are significant ($FDR < 0.05$). Correlation between those methylation levels and the expression of the *ZBTB20* gene is shown in the bottom graphs, with grey dots. Pearson correlation coefficients (all P value < 0.05) are shown in each graph.

and cholangiocarcinoma data showed that EBF1 is an interaction partner for the methylcytosine dioxygenase enzyme TET2 and often binds the motifs surrounding the hypermethylated regions specific to these cancer types.³⁰ These results indicate that EBF1 may play an important role in methylation regulation of cell types other than B-cells and that its deregulation may be involved in the development of different cancers.

A previous genetic study identified an association between triple-negative BC and a SNP in the proximity of the EBF1 locus, suggesting a possible role of the *EBF1* gene in BC, although no functional implication has been established.³¹ Our identification of a highly significant enrichment of EBF1 motifs in basal subtype-specific hypomethylated regions suggests a potential functional role for EBF1 in subtype-specific methylation regulation in BC. The finding that expression levels of a large number of the genes located in close proximity to the DMRs containing EBF1 motifs are correlated with the methylation of the variable CpG positions is consistent with the functional relevance of methylation states in transcriptional regulation. Although functional characterization of LMRs specific to BC subtypes requires further studies, the significant enrichment of genes involved in cell death and proliferation among the candidates found to be under the control of differentially methylated LMRs supports the notion of methylation alterations being important in regulation of critical cellular processes relevant to cancer development and progression. Another limitation is that differences between tumor and

normal mucosa cell purity could somehow affect the TFBS enrichment. Therefore, additional studies would require *in vitro* analyses confirming the impact of EBF1 in BC subtypes.

Of note, identification of *BST2* as the top hit whose DNA methylation was correlated with expression exemplifies a potential pathway through which deregulated expression or activity of EBF1 may contribute to BC development. These results corroborate previous findings that aberrant expression of *BST2* may both enhance cell proliferation and decrease apoptosis rates in high grade BC³² as well as activate migration in tamoxifen-resistant BC cells.³³ Furthermore, identification of the *CD74* gene among the top candidates potentially regulated by DNA methylation in BC also provides a plausible causal pathway in basal BC. *CD74* was proposed to be a potential target in triple-negative BC,³⁴ as it enhances invasion and lymph node metastasis.³⁵ Therefore, our finding that these two genes are lowly methylated and overexpressed in our dataset reinforce the notion that DNA methylation changes may be an important mechanism that contributes to development and phenotype of BC, particularly basal subtype.

Our identification of several genes whose expression levels are positively correlated with methylation is more puzzling. *KLF6* is such an example and it is interesting to note that it seems to be under the control of a short LMR located downstream of its promoter that is enriched in activating histone marks. *KLF6* has been shown to act as a tumor suppressor gene in estrogen receptor (ER)-positive breast tumors³⁶ and it is

therefore plausible that its expression is positively correlated with methylation levels at LMRs found in basal, ER-negative tumors. The putative regulatory region of *ZBTB20*, in contrast, is located in the body of the gene. This is in line with the enrichment analysis performed in the DMR set positively correlated with nearby gene expression that showed a loss of promoter and TSS-proximal regions. These results suggest that low methylation states at LMRs located in gene bodies and downstream of gene promoters may promote transcriptional repression, although future studies are needed to investigate the precise underlying mechanism of gene regulation.

Conclusions

In summary, this study identifies the LMRs specific to the three main BC subtypes and underscores the importance of TF-mediated regulation of the methylome in tumor cells. It also identifies EBF1 as an important TF potentially involved in the epigenetic modulation of a specific subtype of BC. This notion is further supported by the finding that many targets of EBF1, including oncogenes related to aggressive types of BC, such as *BST2* and *CD74*, appear to be under the control of LMRs. Although further studies are required to test the functional impact of LMRs and EBF1 in BC, it is reasonable to propose that EBF1 may be involved in the regulation of methylation states at its targets and in driving biology of basal BC subtype.

Declarations

Ethics approval and consent to participate

N/A

Consent for publication

All authors read and approved the manuscript.

Availability of data and material

All data is available at TCGA data portal (<https://gdc.cancer.gov/>) and/or at the supplementary files provided.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Conceived the study: NFJ, HHV, ZH.

Coordinated the study: ZH.

Performed the analyses: NFJ, HHV, AS, SE, VC, DDE, AJ, PBA, HDW.

Wrote the manuscript: NFJ, ZH.

Read and approved the manuscript: NFJ, AS, SE, VC, DDE, AJ, PBA, HDW, HHV, ZH.

Funding

This work was partially supported by grants from the Institut National du Cancer (INCa, France) and the European Commission (EC) Seventh Framework Programme (FP7) Translational Cancer Research (TRANSCAN)

Framework and the Fondation ARC pour la Recherche sur le Cancer (France) to ZH. NFJ was partially supported by an IARC Fellowship (Marie Curie actions – People – COFUND) and a Postdoctoral Fellowship of the Basque Government. AS is supported by the Fonds National de la Recherche, Luxembourg (AFR code: 10100060).

References

1. Ferlay J, Shin H-R, Bray F, Forman D, Mathers C, Parkin MD. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer*. 2010;127:2893-917. doi:10.1002/ijc.25516. PMID:21351269
2. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015;136:E359-86. doi:10.1002/ijc.29210. PMID:25220842
3. Dawson SJ, Rueda OM, Aparicio S, Caldas C. A new genome-driven integrated classification of breast cancer and its implications. *Embo J*. 2013;32:617-28. doi:10.1038/emboj.2013.19. PMID:23395906
4. Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, Harris CC, McLellan MD, Fulton RS, Fulton LL, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature*. 2010;464:999-1005. doi:10.1038/nature08989. PMID:20393555
5. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, Turashvili G, Ding J, Tse K, Haffari G, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature*. 2012;486:395-9. PMID:22495314
6. Feinberg AP, Koldobskiy MA, Gondor A. Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nat Rev Genet*. 2016;17:284-99. doi:10.1038/nrg.2016.13. PMID:26972587
7. Herceg Z. Epigenetics and cancer: towards an evaluation of the impact of environmental and dietary factors. *Mutagenesis*. 2007;22:91-103. doi:10.1093/mutage/gel068. PMID:17284773
8. Jovanovic J, Ronneberg JA, Tost J, Kristensen V. The epigenetics of breast cancer. *Mol Oncol*. 2010;4:242-54. doi:10.1016/j.molonc.2010.04.002. PMID:20627830
9. Kristensen VN, Lingjaerde OC, Russnes HG, Vollan HK, Frigessi A, Borresen-Dale AL. Principles and methods of integrative genomic analyses in cancer. *Nat Rev Cancer*. 2014;14:299-313. doi:10.1038/nrc3721. PMID:24759209
10. Jones PA, Issa JP, Baylin S. Targeting the cancer epigenome for therapy. *Nat Rev Genet*. 2016;17:630-41. doi:10.1038/nrg.2016.93. PMID:27629931
11. Umer M, Herceg Z. Deciphering the epigenetic code: an overview of DNA methylation analysis methods. *Antioxid Redox Signal*. 2013;18(15):1972-86. doi:10.1089/ars.2012.4923. PMID:23121567
12. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the Infinium Methylation 450K technology. *Epigenomics*. 2011;3:771-84. doi:10.2217/epi.11.105. PMID:22126295
13. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490:61-70. doi:10.1038/nature11412. PMID:23000897
14. Terry MB, Delgado-Cruzata L, Vin-Raviv N, Wu HC, Santella RM. DNA methylation in white blood cells: association with risk factors in epidemiologic studies. *Epigenetics*. 2011;6:828-37. doi:10.4161/epi.6.7.16500. PMID:21636973
15. Ziller MJ, Gu H, Muller F, Donaghey J, Tsai LT, Kohlbacher O, De Jager PL, Rosen ED, Bennett DA, Bernstein BE, et al. Charting a dynamic DNA methylation landscape of the human genome. *Nature*. 2013;500:477-81. doi:10.1038/nature12433. PMID:23925113
16. Marchal C, Miotto B. Emerging concept in DNA methylation: role of transcription factors in shaping DNA methylation patterns. *J Cell Physiol*. 2015;230:743-51. doi:10.1002/jcp.24836. PMID:25283539
17. Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*. 2011;480:490-5. PMID:22170606

18. Feldmann A, Ivanek R, Murr R, Gaidatzis D, Burger L, Schubeler D. Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. *PLoS Genet.* 2013;9:e1003994. doi:10.1371/journal.pgen.1003994. PMID:24367273
19. Hovestadt V, Jones DT, Picelli S, Wang W, Kool M, Northcott PA, Sultan M, Stachurski K, Ryzhova M, Warnatz HJ, et al. Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature.* 2014;510:537-41. doi:10.1038/nature13268. PMID:24847876
20. Burger L, Gaidatzis D, Schubeler D, Stadler MB. Identification of active regulatory regions from DNA methylation data. *Nucleic Acids Res.* 2013;41:e155. doi:10.1093/nar/gkt599. PMID:23828043
21. Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk CL. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics.* 2013;14:293. doi:10.1186/1471-2164-14-293. PMID:23631413
22. Droit A, Gottardo R, Robertson G, Li L. rGADEM: de novo motif discovery. R package version 2.24.0. 2014
23. Wu H, Zhang Y. Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell.* 2014;156:45-68. doi:10.1016/j.cell.2013.12.019. PMID:24439369
24. Domcke S, Bardet AF, Adrian Ginno P, Hartl D, Burger L, Schubeler D. Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature.* 2015;528:575-9. doi:10.1038/nature16462. PMID:26675734
25. Iwafuchi-Doi M, Zaret KS. Pioneer transcription factors in cell reprogramming. *Genes Dev.* 2014;28:2679-92. doi:10.1101/gad.253443.114. PMID:25512556
26. Schubeler D. Function and information content of DNA methylation. *Nature.* 2015;517:321-6. doi:10.1038/nature14192. PMID:25592537
27. Lin YC, Jhunjhunwala S, Benner C, Heinz S, Welinder E, Mansson R, Sigvardsson M, Hagman J, Espinoza CA, Dutkowski J, et al. A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nat Immunol.* 2010;11:635-43. doi:10.1038/ni.1891. PMID:20543837
28. Lee ST, Xiao Y, Muench MO, Xiao J, Fomin ME, Wiencke JK, Zheng S, Dou X, de Smith A, Chokkalingam A, et al. A global DNA methylation and gene expression analysis of early human B-cell development reveals a demethylation signature and transcription factor network. *Nucleic Acids Res.* 2012;40:11339-51. doi:10.1093/nar/gks957. PMID:23074194
29. Boller S, Ramamoorthy S, Akbas D, Nechanitzky R, Burger L, Murr R, Schubeler D, Grosschedl R. Pioneering Activity of the C-Terminal Domain of EBF1 Shapes the Chromatin Landscape for B Cell Programming. *Immunity.* 2016;44:527-41. doi:10.1016/j.immuni.2016.02.021. PMID:26982363
30. Guilhamon P, Eskandarpour M, Halai D, Wilson GA, Feber A, Teschendorff AE, Gomez V, Hergovich A, Tirabosco R, Fernanda Amary M, et al. Meta-analysis of IDH-mutant cancers identifies EBF1 as an interaction partner for TET2. *Nat Commun.* 2013;4:2166. doi:10.1038/ncomms3166. PMID:23863747
31. Purrington KS, Slager S, Eccles D, Yannoukakos D, Fasching PA, Miron P, Carpenter J, Chang-Claude J, Martin NG, Montgomery GW, et al. Genome-wide association study identifies 25 known breast cancer susceptibility loci as risk factors for triple-negative breast cancer. *Carcinogenesis.* 2014;35:1012-9. doi:10.1093/carcin/bgt404. PMID:24325915
32. Sayeed A, Luciani-Torres G, Meng Z, Bennington JL, Moore DH, Dairkee SH. Aberrant regulation of the BST2 (Tetherin) promoter enhances cell proliferation and apoptosis evasion in high grade breast cancer cells. *PLoS One.* 2013;8:e67191. doi:10.1371/journal.pone.0067191. PMID:23840623
33. Yi EH, Yoo H, Noh KH, Han S, Lee H, Lee JK, Won C, Kim BH, Kim MH, Cho CH, et al. BST-2 is a potential activator of invasion and migration in tamoxifen-resistant breast cancer cells. *Biochem Biophys Res Commun.* 2013;435:685-90. doi:10.1016/j.bbrc.2013.05.043. PMID:23702480
34. Tian B, Zhang Y, Li N, Liu X, Dong J. CD74: a potential novel target for triple-negative breast cancer. *Tumour Biol.* 2012;33:2273-7. doi:10.1007/s13277-012-0489-x. PMID:22935920
35. Greenwood C, Metodieva G, Al-Janabi K, Lausen B, Alldridge L, Leng L, Bucala R, Fernandez N, Metodiev MV. Stat1 and CD74 overexpression is co-dependent and linked to increased invasion and lymph node metastasis in triple-negative breast cancer. *J Proteomics.* 2012;75:3031-40. doi:10.1016/j.jprot.2011.11.033. PMID:22178447
36. Liu J, Du T, Yuan Y, He Y, Tan Z, Liu Z. KLF6 inhibits estrogen receptor-mediated cell growth in breast cancer via a c-Src-mediated pathway. *Mol Cell Biochem.* 2010;335:29-35. doi:10.1007/s11010-009-0237-8. PMID:19707857