



## OPEN

SUBJECT AREAS:  
BREAST CANCER  
CANCER GENETICS  
DNA METHYLATION  
DATA INTEGRATION

Received  
16 September 2014

Accepted  
4 February 2015

Published  
5 March 2015

Correspondence and requests for materials should be addressed to Y.X.L. (yxli@sibs.ac.cn); J.X. (xujuanbioc@ems.hrbmu.edu.cn) or X.L. (lixia@hrbmu.edu.cn)

\* These authors contributed equally to this work.

# Genome-wide DNA methylome analysis reveals epigenetically dysregulated non-coding RNAs in human breast cancer

Yongsheng Li<sup>1\*</sup>, Yunpeng Zhang<sup>1\*</sup>, Shengli Li<sup>1\*</sup>, Jianping Lu<sup>1</sup>, Juan Chen<sup>1</sup>, Yuan Wang<sup>1</sup>, Yixue Li<sup>1,2</sup>, Juan Xu<sup>1</sup> & Xia Li<sup>1</sup>

<sup>1</sup>College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China, <sup>2</sup>Shanghai Center for Bioinformation Technology, Shanghai, People's Republic of China.

Despite growing appreciation of the importance of epigenetics in breast cancer, our understanding of epigenetic alterations of non-coding RNAs (ncRNAs) in breast cancer remains limited. Here, we explored the epigenetic patterns of ncRNAs in breast cancers using published sequencing-based methylome data, primarily focusing on the two most commonly studied ncRNA biotypes, long ncRNAs and miRNAs. We observed widely aberrant methylation in the promoters of ncRNAs, and this abnormal methylation was more frequent than that in protein-coding genes. Specifically, intergenic ncRNAs were observed to comprise a majority (51.45% of the lncRNAs and 51.57% of the miRNAs) of the aberrantly methylated ncRNA promoters. Moreover, we summarized five patterns of aberrant ncRNA promoter methylation in the context of genomic CpG islands (CGIs), in which aberrant methylation occurred not only on CGIs, but also in regions flanking CGI and in CGI-lacking promoters. Integration with transcriptional datasets enabled us to determine that the ncRNA promoter methylation events were associated with transcriptional changes. Furthermore, a panel of ncRNAs were identified as biomarkers that discriminated between disease phenotypes. Finally, the potential functions of aberrantly methylated ncRNAs were predicted, suggesting that ncRNAs and coding genes cooperatively mediate pathway dysregulation during the development and progression of breast cancer.

The development of human breast cancer is mediated by both genetic and epigenetic alterations of the cell<sup>1,2</sup>. Since the discovery of altered DNA methylation in human cancer, DNA methylation studies of breast cancer have used methodologies of varying scale, focusing on a few coding genes or regions assumed to be functionally important, such as promoters and CpG islands (CGIs)<sup>3,4</sup>. Although it is well understood that most of the mammalian genome is transcribed, producing non-coding RNAs (ncRNAs), the genome-wide methylation patterns of ncRNAs in breast cancer remain largely unknown.

NcRNA transcripts have been categorized into several groups based on their size, which is the most popular classification method. These classes include the well-annotated microRNAs (miRNAs) and long ncRNAs (lncRNAs). lncRNAs account for approximately 81.8% of all ncRNAs<sup>5</sup>. Although the molecular basis of the functions of many lncRNAs is just emerging, much evidence indicates that lncRNAs play intricate roles in the regulation of a wide variety of biological processes, such as imprinting and gene expression at the transcriptional level<sup>6–8</sup>. Considering the potential functions of lncRNAs, their transcription must be tightly regulated. Aberrant expression of lncRNAs has appeared in prevalent cancer types, including breast cancer. One notable example is HOTAIR, which is over-expressed in breast cancers; loss of HOTAIR reduces the invasiveness of breast cancer<sup>9</sup>. Another example is MIR31HG, which is expressed abundantly in non-invasive breast cancer cell lines of the luminal subtype<sup>10</sup>. Although lncRNAs have been demonstrated to participate in the modulation of gene expression<sup>11</sup>, the epigenetic regulation of lncRNAs remains poorly understood. Recent studies have described aberrant methylation of specific lncRNAs in breast cancers. However, studies of aberrant epigenetic regulation patterns in lncRNA genes at a global scale are scarce.

In addition, miRNAs are a recently discovered and well-characterized class of ncRNAs<sup>12</sup>. miRNAs are important regulators of gene expression and are frequently dysregulated in cancer<sup>13,14</sup>; aberrant DNA methylation is an



epigenetic mechanism that is involved in the process of miRNA dysregulation<sup>15–17</sup>. Aberrant DNA methylation events associated with the silencing of individual miRNAs have been demonstrated in many cancer types, including breast cancer<sup>18,19</sup>. Some of these miRNAs function as tumor suppressors (such as miR-203, miR-195 and miR-497) and the down-regulation of these miRNAs due to aberrant hypermethylation is associated with increased malignancy or metastatic potential in breast cancer<sup>20,21</sup>. Using 5-methylcytosine immunoprecipitation coupled to miRNA tiling microarray hybridization, Vrba et al. have demonstrated that miRNA gene promoters are frequent targets of aberrant DNA methylation in human breast cancer<sup>22</sup>, indicating an important role of DNA methylation in miRNA dysregulation in cancer. However, only 167 miRNAs were analyzed in their study, accounting for only 10% of all miRNAs in the genome. To our knowledge, the comprehensive analysis of the methylation of miRNA genes in breast cancer has yet to be performed.

Next-generation sequencing technologies have emerged as powerful tools that enable whole-genome profiling of epigenetic modifications, including DNA methylation. For instance, the MBDCap-seq protocol, is a technique used to identify methylated DNAs using a methyl-CpG binding domain (MBD) protein column followed by next-generation sequencing. The low cost and unbiased generation of the methylation profiles of both coding and non-coding regions render this technique as suitable for genome-wide methylation profile analysis. The Cancer Methylome System (CMS)<sup>23</sup> has recently used high-throughput sequencing technology to generate DNA methylation profiles in a cohort of 87 breast samples (77 cancer samples and 10 normal control samples). This study was a comparative analysis of the methylomes generated by the previous unbiased systematic effort to determine the aberrant methylation patterns of ncRNAs, and to provide the precise genomic locations that undergo methylation changes. The data used in this study represent a highly valuable public resource understanding the epigenetic regulation of the breast cancer genome and for identifying ncRNAs as therapeutic targets.

## Results

**Global differences in DNA methylation between breast cancer and normal control samples.** We performed comprehensive comparative analyses of the DNA methylation profiles of normal control and breast cancer samples, which were downloaded from the CMS. First, by computing the average intensity correlation of methylation across all breast cancer and normal control samples, we found that the control and breast cancer samples were classified into two different groups (Figure S1A). Moreover, the variation of the correlation strength among the breast cancer group was higher than that of the normal control group (Figure S1B), suggesting that the methylation profiles among individuals with breast cancer may display some heterogeneity.

Next, the differentially methylated regions (DMRs) across the entire genome were identified. Approximately 838,817 hypomethylated DMRs were identified in the breast cancer samples, and this result was approximately 16-fold greater than that of hypermethylated DMRs (Figure S1C). By investigating the genomic distribution of the DMRs, we found that the hyper- and/or hypomethylated DMRs were dispersed throughout multiple chromosomes, and their distributions were similar (Figure S1D). In summary, the breast cancer samples contained extensive methylome alterations.

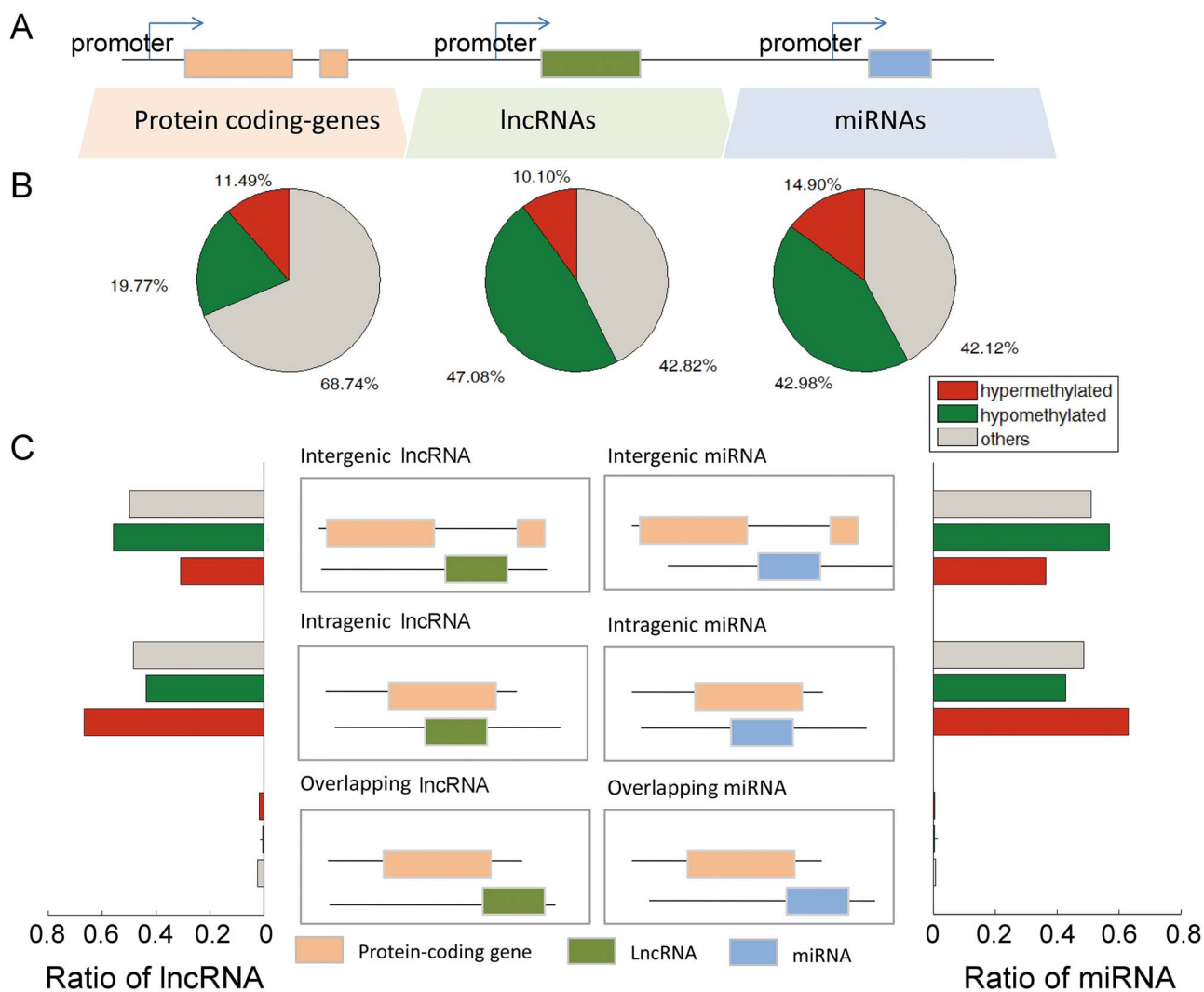
**Aberrant methylation is associated with ncRNAs in breast cancer.** ncRNAs (including miRNAs and lncRNAs) are important regulators of gene expression; however, the genome-wide aberrant methylation of ncRNAs in breast cancer has been poorly understood. By mapping the DMRs to the ncRNA and protein-coding genes, we observed wide methylation changes in the

promoters of ncRNA genes, which were more frequent than those of protein-coding genes (Figure 1A and 1B). Specifically, approximately 57.18% of the lncRNA promoters and 57.88% of the miRNA promoters were aberrantly methylated. By contrast, only 31.26% of the promoters of coding genes were aberrantly methylated in breast cancer, some of which were in known disease-related genes, such as HOXB2, FGF4 and TTK. By further investigating the direction of methylation dysregulation, we identified 1,359/278 hypermethylated lncRNA/miRNAs and 6,336/802 hypomethylated lncRNA/miRNAs (Table S1). Among these identified miRNAs, many have been demonstrated to be involved in the process of tumorigenesis, such as hsa-mir-29b<sup>16</sup>, hsa-mir-21<sup>24</sup> and hsa-mir-10b<sup>25</sup>.

Previous reports have suggested that different biotypes of ncRNAs perform distinct functions<sup>26,27</sup>. Thus, these ncRNAs were further classified into three biotypes based on their location with respect to protein-coding genes: intergenic, intragenic and overlapping ncRNAs. Intergenic lncRNAs comprised the majority of the aberrantly methylated ncRNA promoters (51.45% of all lncRNAs and 51.57% of all miRNAs) in the breast cancer samples (Figure 1C). Most of these intergenic lncRNAs had not previously been detected via microarray analysis, and our genome-wide analysis produced new candidates for further functional analysis. Specifically, approximately 31.79% of the hypermethylated lncRNAs and 55.84% of the hypomethylated lncRNAs were intergenic. Similar to the dysregulation patterns of lncRNAs, 36.33% of the hypermethylated miRNAs were intergenic (Figure 1C). Intragenic ncRNAs comprised another interesting type of ncRNAs. We found that most intragenic ncRNAs shared the same -or largely overlapping promoters with their host genes. Moreover, for the ncRNA-host promoter pairs displaying an overlap of less than 50%, 32.65% and 24.79% of the lncRNAs and miRNAs, respectively, exhibited the same directional change in methylation as their host genes. These results indicate that most intragenic ncRNAs tend to display identical methylation patterns to their host genes. However, it cannot be ignored that several ncRNAs display independent methylation patterns.

**Aberrant epigenetic regulation patterns of ncRNA promoters in the context of CGIs.** CGIs have been found to be functionally important in cancer based on DNA methylation studies. To determine whether CGIs tend to be aberrantly methylated in breast cancer, the DMRs were mapped to CGIs. As a result, we observed that 31.85% of the CGIs were hypermethylated in the breast cancer samples (Figure S2A). In addition to CGIs, CGI shores have been demonstrated to be frequently aberrantly methylated in colon cancer<sup>16</sup>. However, the investigation of the methylation patterns of CGI shores in breast cancer has been hampered by a lack of genome-wide profiling. Here, we found that approximately 40% of CGI shores were aberrantly methylated and the number of 5' shores and 3' shores displaying hyper- or hypomethylation appeared to be equal (Figure S2A). Moreover, we compared the methylation frequency distributions of hypomethylated CGIs with those of hypermethylated CGIs and observed opposing trends (Figure S2B and S2C). The hypermethylation frequency was extremely high on CGIs, whereas the hypomethylation frequency was substantially high on CGI shores. Specifically, for 88.69% of the hypermethylated CGIs, their corresponding 5' or 3' shore regions were also hypermethylated (Figure S2D). By contrast, the hypomethylated DMRs tended to reside on CGI shores but not on CGIs, and the 5' and 3'-CGI shores were rarely simultaneously hypomethylated (Figure S2E).

Recently, Weber et al. reported that gene promoters containing weak CGIs unbound to RNA polymerase II were frequently methylated<sup>28</sup>. Thus, the relevance of specific characteristics of DNA methylation, such as the location of aberrant methylation, or the association between CGI and CGI shore methylation, to the develop-



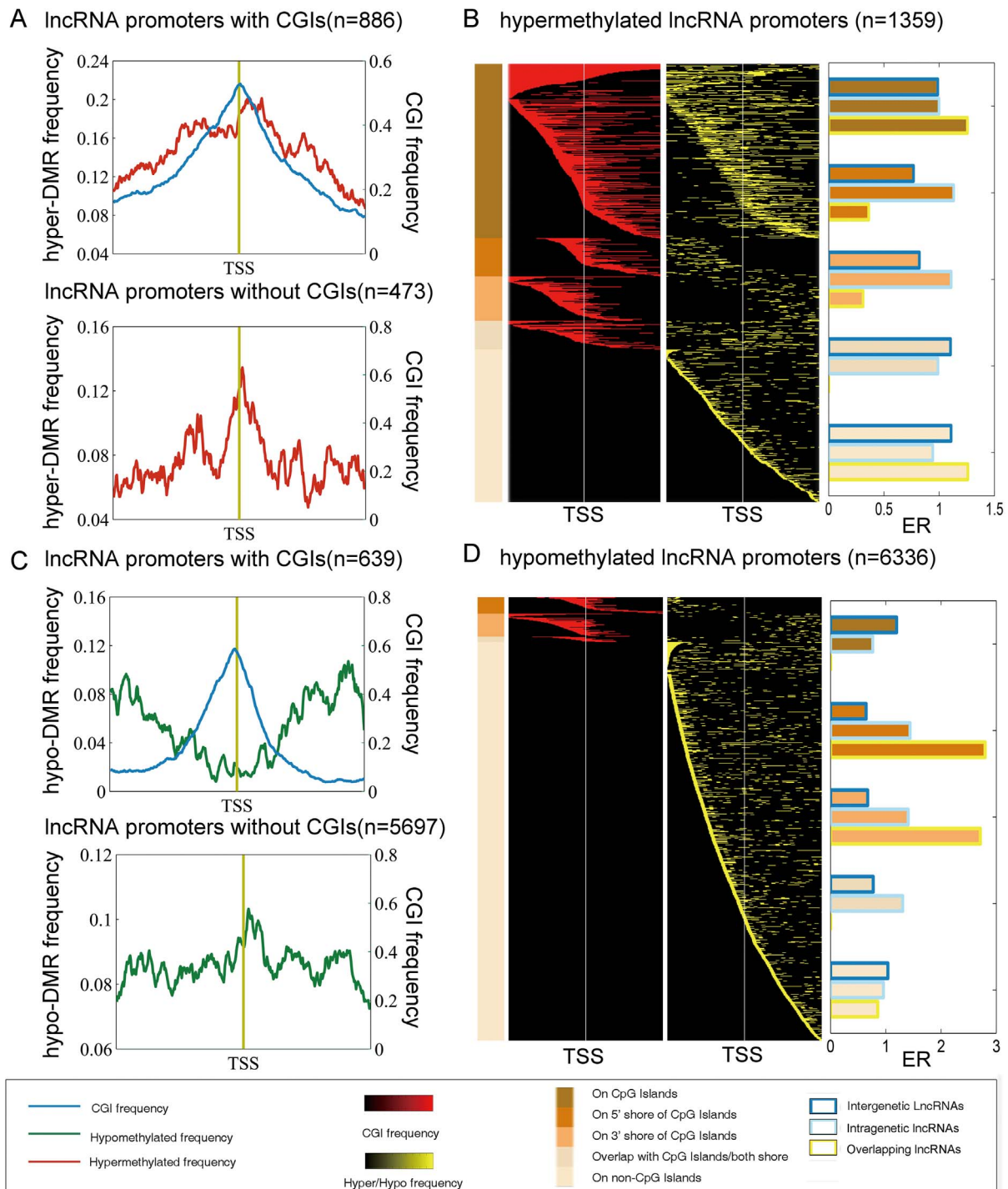
**Figure 1 | Characterization of the genome-wide methylation patterns of ncRNAs.** (A) Schematic representing the analyzed genomic features. (B) Pie charts representing the percentage of coding and non-coding promoters that overlapped with hyper- and hypo-methylated DMRs in breast cancer. The green pie charts represent the percentage of the regions covered by hypo-methylated DMRs, and the red pie charts represent the percentage of the regions covered by hypermethylated DMRs. (C) lncRNAs and miRNAs were classified into different subcategories based on their location relative to protein-coding genes. The bar graphs show that the majority of intragenic ncRNAs were overrepresented with respect to hypermethylation and that most intergenic ncRNAs were hypomethylated in the breast cancer samples (all p-values < 0.05, Fisher's exact test).

ment of cancer merits investigation. Next, we examined the aberrant methylation patterns of ncRNAs in the context of CGIs and identified several distinct aberrant methylation patterns on lncRNA promoters. Overall, 34.81% ( $n = 473$ ) of the hypermethylated lncRNA promoters lacked a CGI and displayed aberrant hypermethylation around the transcription start site (TSS) (Figure 2A). The remaining 65.19% ( $n = 886$ ) of the hypermethylated lncRNA promoters contained CGIs, and among these promoters, four distinct aberrant methylation patterns were classified (Figure 2B): (1) aberrant methylation predominantly confined to CGIs; (2) aberrant methylation located on the 5' shore of the CGIs; (3) aberrant methylation located on the 3' shore of the CGIs and (4) aberrant methylation that overlapped with the CGI shores and an overlap of less than 50% between the CGIs and the DMR. We found that aberrant hypermethylation was predominantly confined to CGIs or their 5' and 3' shores (Figure S3A and Table S2). Next, we calculated the average aberrant methylation frequency by dividing the promoters into 10 bp fragments and observed that the aberrant hypermethylation frequency was particularly high immediately downstream of the TSS in these hypermethylated promoters (Figure 2A). This region has generally been regarded

as the core promoter<sup>29</sup>, in which RNA polymerase II is recruited to the DNA and initiates transcription. Aberrant methylation in core promoter regions may lead to transcriptional inactivation/activation of cancer-related genes and plays an integral role in tumorigenesis. For the hypomethylated lncRNA promoters in the breast cancer samples, we found that nearly 89.91% of the hypomethylated lncRNA promoters lacked a CGI (Figure 2C and Figure S3A); the aberrant hypomethylation frequency peak was immediately downstream of the TSS (Figure 2C), similar to that for aberrant hypermethylation. However, the aberrant methylation frequency of the hypomethylated promoters containing CGIs was distinct from that of other promoters because these promoters displayed the lowest aberrant hypomethylation frequency surrounding the TSS (Figure 2C). Moreover, we found that intragenic lncRNAs comprised the majority of each pattern of aberrantly methylated lncRNA promoters. In addition, the intergenic lncRNAs were overrepresented among the hypomethylated lncRNAs (Figure 2B and 2D).

We also found that the global distribution of the aberrant methylation frequency of miRNAs was similar to that of lncRNAs (Figure 3





**Figure 2 | Aberrant methylation patterns surrounding the TSSs of the lncRNAs in breast cancer.** (A) Average aberrant hypermethylation frequency of 886 lncRNA promoters containing CGIs and 473 lncRNA promoters lacking a CGI. (B) Heat map of the CGI frequency and the aberrant hypermethylation frequency in breast cancer. A total of 1359 lncRNA promoters contained a hypermethylated region (yellow) were identified. Each row represents a unique promoter region at a 10-bp window size, covering the  $\pm 2$ -kb region flanking the transcription start sites. The location of a CGI (red) in the aberrantly methylated lncRNA promoters is shown in the first column. The promoters are ordered according to the location of methylation on a CGI, or adjacent to the CGI (CGI shore); the promoters that lacked a CGI are represented with different shades of brown on the left. ER, enrichment ratio. (C) Average aberrant hypomethylation frequency of 639 lncRNA promoters containing CGIs and 5697 lncRNAs lacking a CGI. (D) Heat map of the CGI frequency and the aberrant hypomethylation frequency in breast cancer. The ratio of lncRNA subtypes for each pattern is shown adjacent to the heat map.

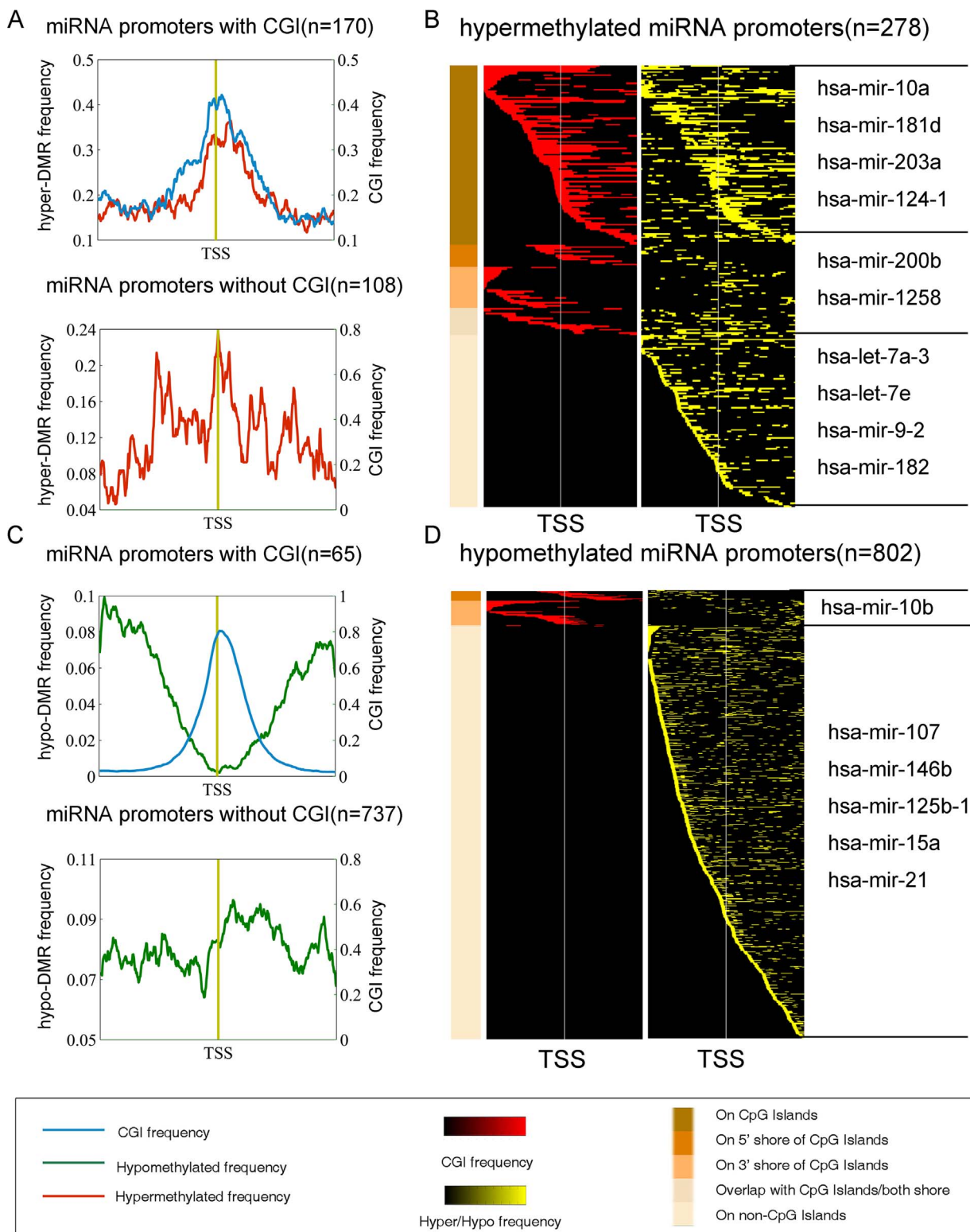
and Table S2). Specifically, the aberrantly hyper-methylated miRNA promoters were confined to CGIs, and the hypo-methylated miRNA promoters lacked CGIs (Figure 3B). Moreover, the aberrant methylation frequency, especially the hypermethylation frequency, was

increased surrounding the TSS, implying that miRNA promoters were frequent targets of aberrant DNA methylation in human breast cancer<sup>22</sup>. These results suggest a similar epigenetic mechanism between different types of ncRNAs.



**Regulation of ncRNA expression by DNA methylation.** To explore the functional consequences of aberrant methylation, we next examined the association between ncRNA methylation events and transcriptional changes. Based on the repressive effects of DNA

methylation, two scenarios of particular biological relevance were considered. In the first scenario, reduced expression of ncRNA is due to hypermethylation (silencing), and in the second scenario, increased expression of ncRNA is due to hypomethylation



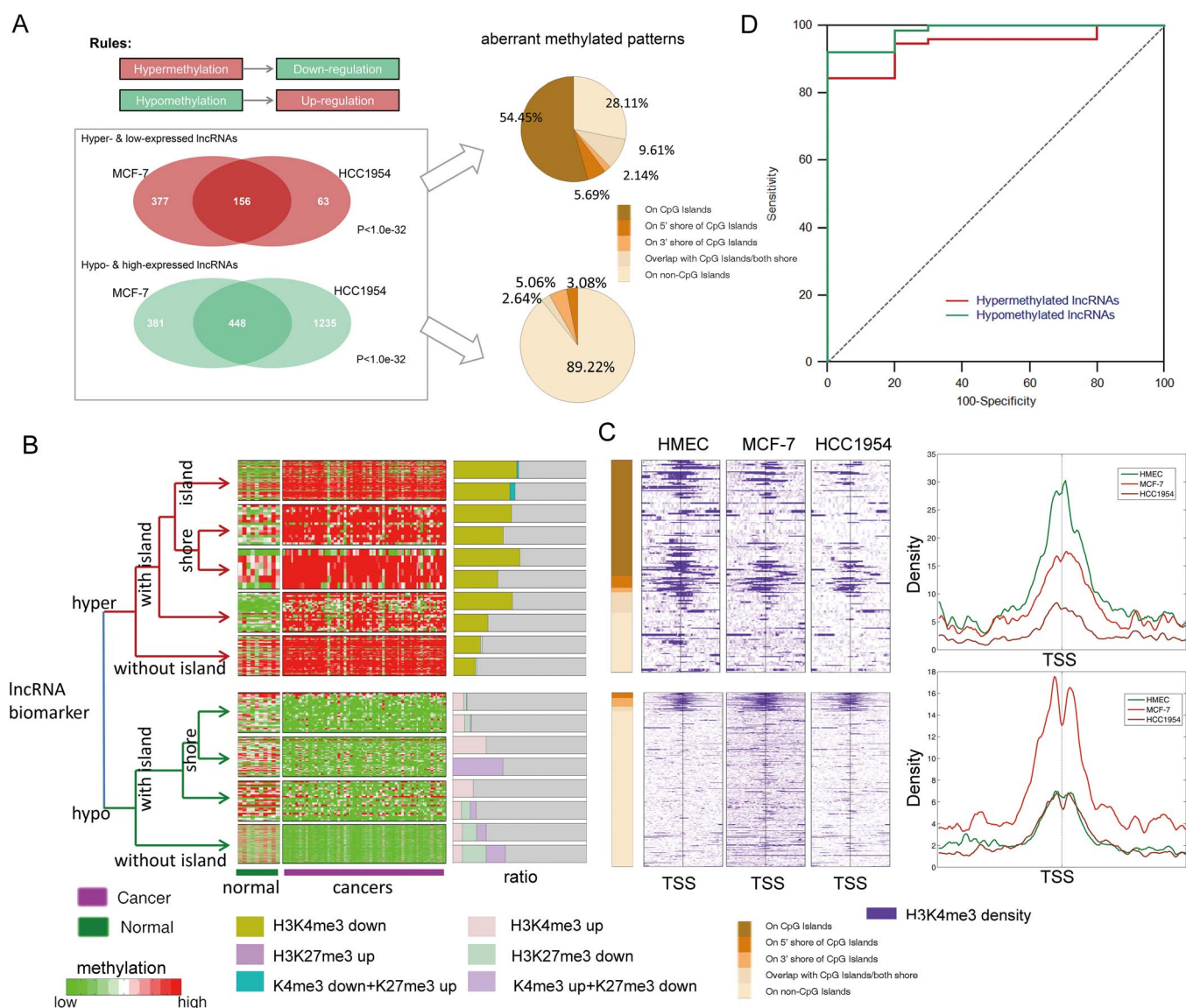
**Figure 3 | Aberrant methylation patterns surrounding the TSSs of miRNAs in the breast cancer samples.** (A) Average aberrant hypermethylation frequency of 170 promoters containing CGIs and that of 108 promoters that lacked a CGI. (B) Heat map of the CGI frequency and the aberrant hypermethylation frequency in breast cancer. Representative miRNAs in each group are shown on the right. (C) Average aberrant hypomethylation frequency of 65 promoters containing CGIs and that of 737 promoters that lacked a CGI. (D) Heat map of the CGI frequency and the aberrant hypomethylation frequency in breast cancer. Representative miRNAs in each group are shown on the right.



(activating). As a result, 156 and 448 candidate silencing and activating lncRNAs candidates, respectively, were identified in both cell lines (Figure 4A), which were distributed in the patterns described above. Moreover, consistent with previous studies of coding genes, the aberrantly hypermethylated lncRNAs containing CGIs were more likely to be associated with transcriptional repression. Approximately 54.45% of the hypermethylated lncRNAs containing CGIs were associated with transcriptional repression (Figure 4A). As both DNA methylation and histone modification are involved in establishing patterns of gene expression during the progression of cancer, we next explored whether aberrant methylation of lncRNAs was associated with altered histone modification (Text S1). Notably, approximately 50% of the hypermethylated lncRNAs displayed reduced H3K4me3 or elevated H3K27me3 levels, whereas, approximately 30% of the hypomethylated lncRNAs were associated with elevated H3K4me3 or reduced H3K27me3 levels (Figure 4B). Moreover, the aberrant DNA hypermethylation of ncRNAs was

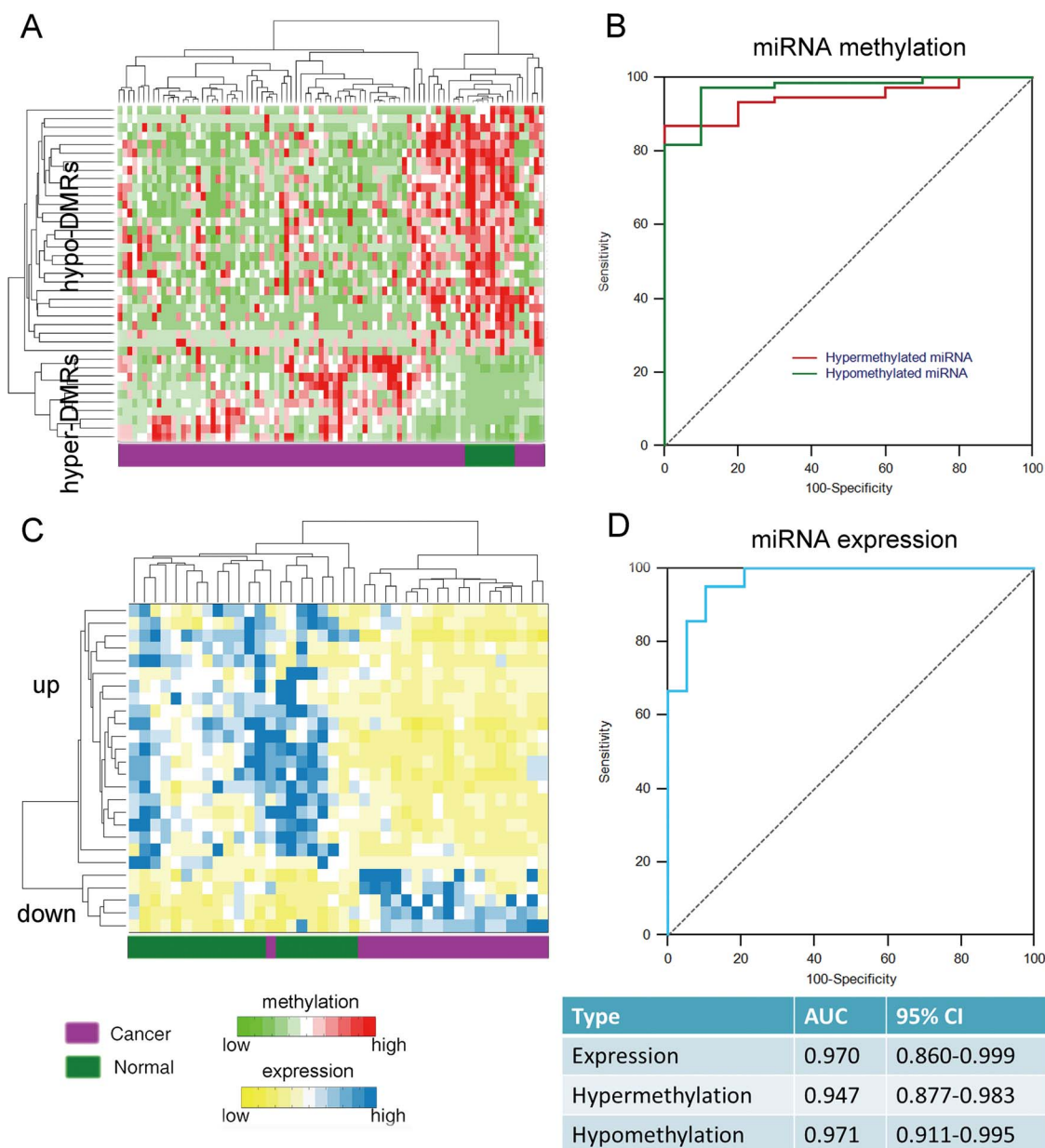
more likely to coincide with aberrant modification of H3K4me3 than that of H3K27me3. By contrast, ncRNA hypomethylation was associated with an alteration of the H3K27me3 levels. This result is consistent with previous observations that H3K27me3 and DNA methylation are mutually exclusive, especially on CGIs<sup>30</sup>. Specifically, we found that the incidence of histone modification changes was extremely high at the TSS, reinforcing the notion that aberrant epigenetic modifications were strongly biased toward the core promoter regions (Figure 4C).

The ncRNA or protein-coding genes display aberrant expression might serve as biomarkers for cancer diagnosis. Therefore, we examined whether these lncRNAs served as potential biomarkers for early diagnosis. The potential biomarkers that were examined consisted of the activating and silencing lncRNAs identified above. The results revealed that these activating and silencing lncRNAs distinguished breast cancer samples from healthy control samples, which might aid in the future development of diagnostic biomarkers for breast cancer.



**Figure 4 | Regulation of lncRNA expression by DNA methylation in breast cancer.** (A) The aberrantly methylated and expressed lncRNAs in the two cell lines significantly overlapped. (B) Clustering map of selected lncRNA biomarkers for CMS methylation. The lncRNAs were grouped based on their aberrant methylation patterns. The bar graph adjacent to the heat map shows the percentage of lncRNAs in that group covered by an aberrant histone modification. (C) Colored profiles of the H3K4me3 sequencing read densities from the HMEC, MCF-7 and HCC1954 cell lines. Each profile shows the 4-kb regions surrounding the aberrantly methylated lncRNAs. The average profiles are shown on the right of the colored profiles. (D) ROC map of the selected markers.





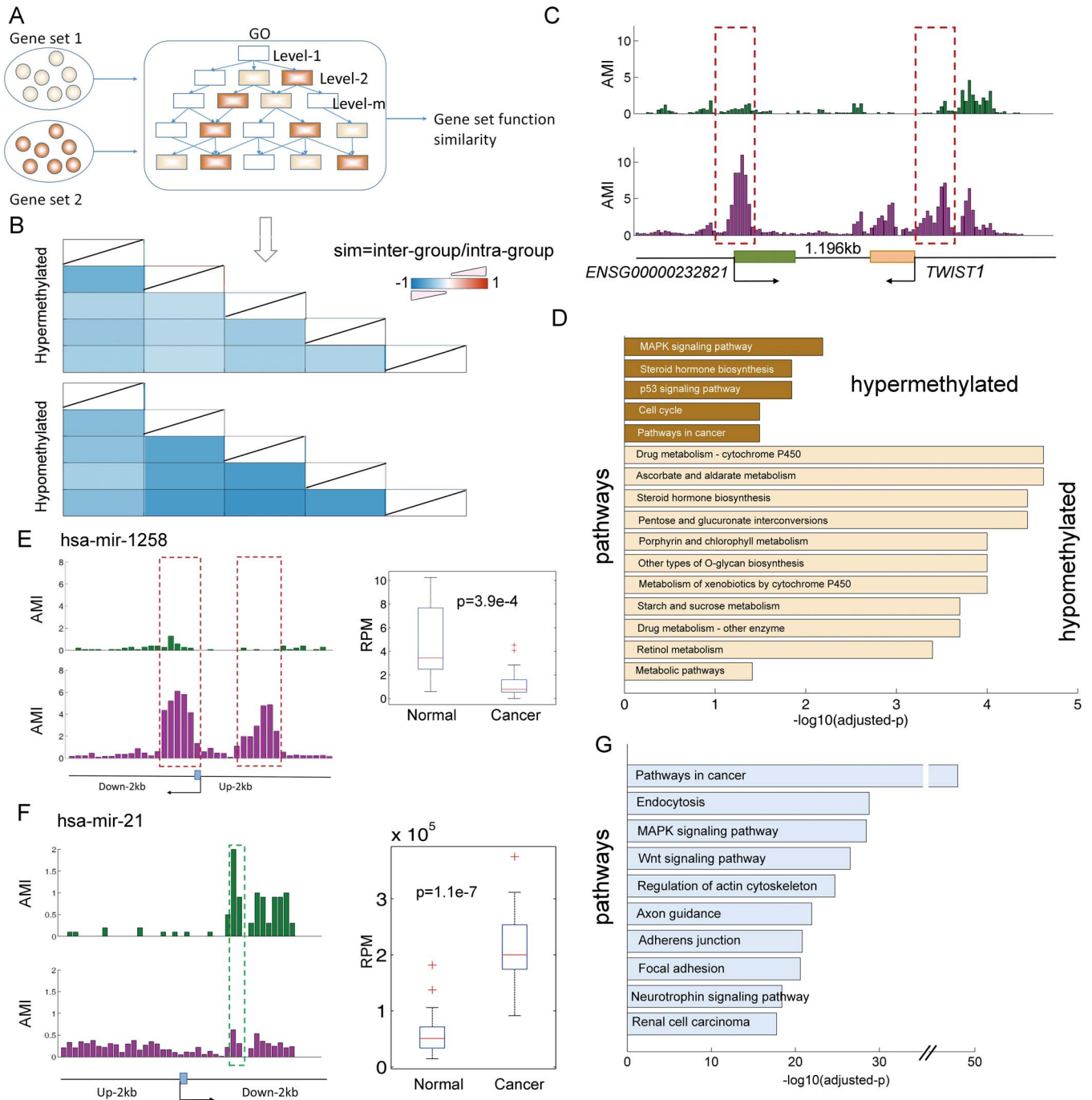
**Figure 5 | Regulation of miRNA expression by DNA methylation in breast cancer.** (A) and (C) Clustering map of the selected miRNA biomarkers in the CMS methylation dataset and TCGA expression dataset, respectively. (B) and (D) ROC analysis of these miRNA biomarkers.

The diagnostic value of these lncRNAs was evaluated based on area under the receiver operation characteristic (ROC) curve (AUC) analysis. The ROC curves of the hyper- and hypomethylated lncRNAs displayed AUC values of 0.944 and 0.983, respectively, using the CMS dataset (Figure 4D).

Furthermore, miRNA biomarkers were identified, including 26 (5 hypermethylated and 21 hypomethylated) miRNAs. Three of these five hypermethylated miRNAs, were confined to CGIs. Most of the hypomethylated miRNAs lacked a CGI in their promoters; aberrant methylation was located on the CGI shores of several miRNA promoters, such as hsa-mir-106a, hsa-mir-1292 and hsa-mir-3613. Based on hierarchical clustering analysis of the samples according to the methylation and expression of these 26 miRNAs (Figure 5A and 5C), most breast cancer tissues were distinguishable from adjacent benign tissues, with overall AUCs larger than 0.90 in the CMS and TCGA datasets (Figure 5B and 5D). Notably, most of these miRNAs were previously associated with breast cancer (Table S3).

#### Epigenetically dysregulated ncRNAs disrupt functions associated with breast cancer.

An increasing number of lncRNAs have been characterized; however, the functions of most lncRNA genes remain unknown. In addition, the functional prediction of lncRNAs is hampered by the lack of collateral information, such as molecular interaction data and expression profiles. The large-scale functions annotated to lncRNAs were primarily predicted based on co-expression and genomic adjacency<sup>31,32</sup>. However, despite similar expression patterns, groups of functionally related genes can be further distinguished at the chromatin level<sup>33</sup>. Thus, we hypothesized that genes displaying similar aberrant methylation patterns may perform similar functions. To test this hypothesis, we first investigated the aberrant methylation patterns of the protein-coding genes. We found that the protein-coding genes shared similar aberrant methylation patterns with the ncRNA genes (Figure S4). Next, we determined the functional similarity of the gene set between genes displaying distinct aberrant methylation patterns (Figure 6A and

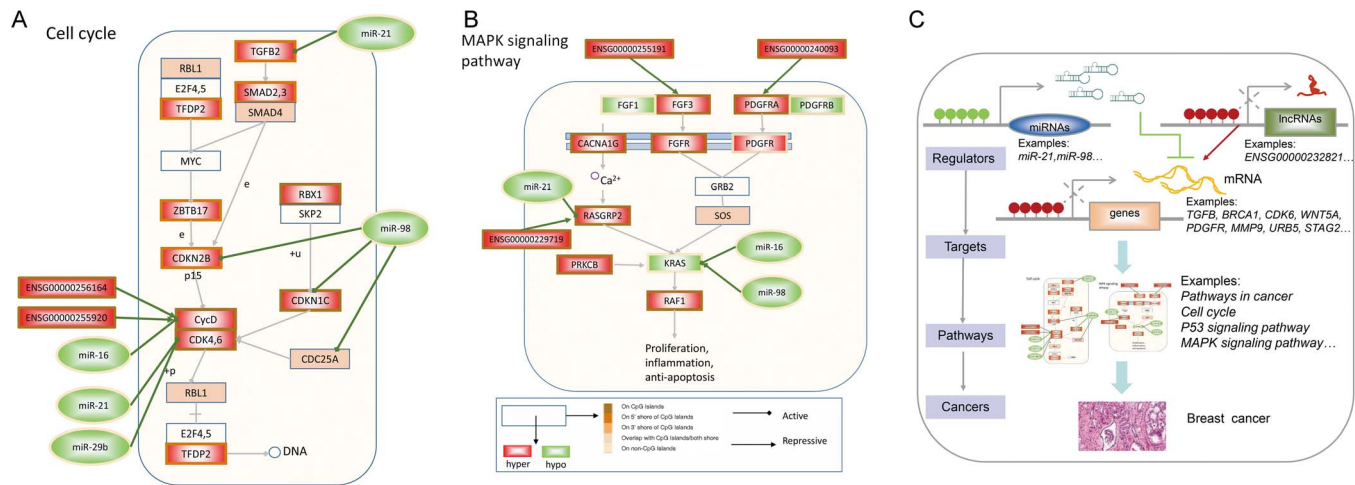


**Figure 6** | Aberrantly methylated ncRNAs widely disturb functions associated with breast cancer. (A) The system used to calculate the gene set functional similarity. (B) The genes that displayed similar aberrant methylation patterns exhibited with high functional similarity. The similarity score was normalized to the intra-classes. (C) Representative lncRNA and protein-coding genes. The average methylation intensities (AMIs) in the control and breast cancer samples are shown. Green, control samples; Pink, cancer samples. (D) The biological processes enriched by adjacent genes of lncRNA displaying similar aberrant methylation patterns. (E) The average methylation and expression levels of hsa-mir-1258 in the normal and breast cancer samples. RPM, reads per million. (F) The average methylation and expression levels of hsa-mir-21 in the control and breast cancer samples. (G) The top 10 pathways enriched by targets of miRNA biomarkers.

Text S1). Consistent with our hypothesis, we found that genes displaying similar aberrant methylation patterns exhibited greater functional similarity than genes displaying distinct methylation patterns (Figure 6B). In addition, the genes that display the same directional change in methylation exhibited greater functional similarity than those that displayed opposing directional changes in methylation (Figure S5).

As the general large-scale functional annotation of lncRNAs has been based on the ‘guilt-by-association’ principle, we applied this principle and aberrant methylation patterns to predict the probable functions of the aberrantly methylated and expressed lncRNAs identified above. For example, the lncRNA ENSG00000232821 displayed hypermethylation confined to its CGI. We found that the coding gene -TWIST1, which is adjacent to this lncRNA, also displayed





**Figure 7 | NcRNAs and coding genes cooperatively mediated pathway dysregulation in breast cancer.** (A) The cell cycle pathway was dysregulated by ncRNAs and coding genes. Only the representative component of this pathway is shown. The hypermethylated genes are marked in red, and the hypomethylated genes are marked in green. The aberrant methylation patterns are labeled on the edge of the rectangle. The experimentally validated miRNA-mediated regulations from TarBase and adjacent lncRNAs with common aberrant methylation patterns are linked. (B) The MAPK signalling pathway. Only the representative component of this pathway is shown. (C) A model illustrating the dysregulated cellular network intertwined by ncRNAs and coding genes, ncRNA and coding genes cooperatively mediate the pathway dysregulation.

hypermethylation confined to its CGI (Figure 6C). TWIST1 is an important transcription factor that has been implicated in cell lineage determination and differentiation. TWIST1 promoter methylation has been demonstrated to be significantly more prevalent in malignant than in healthy breast tissue. It is reasonable to infer that this lncRNA may also be involved in development of breast cancer. We found that the potential functions of this lncRNA based on the NONCODE database were the regulation of growth and development, suggesting that this lncRNA plays important roles in the development of cancers. Next, we aimed to explore the dysregulated pathways corresponding to different patterns of aberrantly methylated lncRNA biomarkers. We found that only the hypermethylated lncRNAs confined to CGIs and the hypomethylated lncRNAs lacking CGIs were enriched in pathways directly associated with the development and progression of breast cancers (Figure 6D and Table S4). The hypermethylated lncRNAs primarily affected the cell cycle and signaling pathways, whereas the hypomethylated lncRNAs predominantly disturbed most of metabolic pathways (Table S4). Cancer cells require metabolism to sustain their existence, including the support of several functions, such as cell maintenance, proliferation and motility. Common to all of these activities is the demand for energy<sup>34</sup>. These results indicated that lncRNA hypomethylation may be associated with carcinogenesis by dysregulation of the signaling and metabolic pathways.

It is well understood that miRNAs perform their important functions via targets; functional enrichment analysis revealed that the predicted targets of these 26 miRNA biomarkers, were implicated in functions that play direct and important roles in tumor growth and metastasis (Figure 6G and Table S5). For example, hsa-miR-29b was highly over-expressed in breast cancer, and directly enhancing the hsa-miR-29b mediated impairment of apoptosis and increasing tumor cell migration and invasion by targeting the tumor suppressor PTEN<sup>35</sup>. In the present study, we found that although the promoter of miR-29b lacked a CGI, it was hypomethylated, providing an epigenetic perspective of the aberrant expression of hsa-miR-29b in cancers. Another example is the tumor suppressor- hsa-miR-1258, which was hypermethylated on the 3' CGI shore and was under-expressed in breast cancer samples in the present study (Figure 6E). Heparanase (HPSE) is a potent pro-tumorigenic, pro-angiogenic, and pro-metastatic enzyme that is over-expressed in brain metastatic breast cancer (BMBC). Studies have shown that hsa-miR-1258 inhi-

bits the expression and activity of HSPE in BMBC cells, and stable expression of hsa-miR-1258 in BMBC cells inhibits HSPE-induced cell invasion in vitro and brain metastasis in an experimental model<sup>36</sup>. In addition, high hsa-miR-21 expression was associated with mastectomy, larger tumor size, higher tumor stage, higher tumor grade, estrogen receptor status. Increasing studies have identified many downstream targets of hsa-miR-21, such as WNT5A, PTEN and BRCA1, in various types of cancers<sup>24</sup>. Consistent with previous studies, hsa-miR-21 was significantly highly expressed in our analyzed breast cancer samples ( $p = 1.1e-7$ ,  $t$  test). However, the mechanism underlying the high expression of hsa-miR-21 remains largely unknown. We found that the promoter of this miRNA lacks a CGI and is hypomethylated in breast cancer samples (Figure 6F), suggesting a possible epigenetic role in the regulation of its activity in breast cancer. Moreover, we found that some of the targets of this miRNA, such as BRCA1, CDK6 and WNT5A, were hypermethylated. Mutation of the tumor suppressor gene BRCA1 is an important contributor to the hereditary breast cancer; however, BRCA1 mutations have not been detected in certain types of breast cancer, suggesting that DNA methylation and/or miRNA-mediated repression of BRCA1 may participate in the development of breast cancers<sup>37</sup>. Our analysis indicated that hypomethylation of miRNA regulators and hypermethylation of the BRCA1 gene may be another repression mechanism of this tumor suppressor gene.

**NcRNAs and coding genes coordinately mediate pathway dysregulation in breast cancer.** Tumorigenesis is a complex dynamic biological process that involves multiple steps of genetic and epigenetic alterations<sup>38</sup>. However, our understanding of this complex network intertwined by coding genes and ncRNAs in cancer biology remains at an early stage. We found that aberrant methylation of either lncRNA or miRNA can perturb specific common pathways, such as the cell cycle, and MAPK signaling pathways, indicating that ncRNAs mediate the dysregulation of these pathways in a coordinated manner. As one of the most important cellular processes, the cell cycle is precisely regulated in all organisms. Dysregulation of the cell cycle can lead to catastrophic cellular events, e.g., premature apoptosis or abnormal proliferation, which are causes of cancers<sup>39</sup>. We found that the lncRNA and miRNA regulated key components of this pathway. One such example is the CDK6 gene, which is responsible for modulating



the activities of growth-suppressing Rb family proteins. Evidence has shown that normal human mammary epithelial cells (HMECs) exhibit a high level of CDK6 activity, but all breast tumor-derived cell lines exhibit lower CDK6 activity levels, with several having little or no CDK6 activity<sup>40</sup>. Our analyses demonstrated that DNA methylation may be one mechanism by which these cell cycle-associated genes are repressed (Figure 7A). Moreover, we found that this gene was targeted by two hypomethylated miRNAs, hsa-miR-21 and hsa-miR-29b. The hypomethylation of these two key regulators may further repress the activity of CDK6, implying complementary effects of DNA methylation- and miRNA-mediated regulation. Another example is the hypermethylated gene- CCND2 (Figure 7A), which was regulated by a hypomethylated miRNA (hsa-miR-16). This cyclin forms a complex with and functions as a regulatory subunit of CDK4 or CDK6, the activity of which is required for the G1/S transition in the cell cycle<sup>41</sup>. In addition, two hypermethylated lncRNAs share a similar aberrant methylation pattern to this gene (hypermethylated on CGIs), implying that these two lncRNAs play key roles in cell cycle regulation.

Another dysregulated pathway is the MAPK signaling pathway, which controls fundamental cellular processes, such as growth, proliferation, differentiation, migration and apoptosis<sup>42</sup>. It was revealed in our study that most genes in this pathway were aberrantly methylated in breast cancer (Figure 7B). Some of these genes, such as FGFR1, KRAS and PRKCB, were reported to be involved in tumorigenesis. Correlations between the expression of FGFRs and breast cancer progression have been observed<sup>43</sup>, and our study indicated that DNA hypomethylation may be one mechanism responsible for the repression of FGFR genes in breast cancer. The Ras gene is another important gene in breast cancer and belongs to the Ras oncogene family, which has been very extensively studied and has been found to be involved in pathological processes such as cancer and development<sup>44</sup>. The experimental observations accumulated over many years suggest that somatic mutations are the typical genetic alterations that affect Ras. However, the mutation event for the RAS gene does not appear to be involved in the etiology of breast cancer in some patients<sup>45</sup>. We found that the KRAS gene was hypomethylated and was regulated by two miRNAs (hsa-miR-16 and hsa-miR-98). Moreover, these two miRNAs, which have been found to be implicated in the breast cancer development<sup>46</sup>, were aberrantly methylated. Therefore, the KRAS gene is strictly regulated by DNA methylation and miRNAs.

Taken together, our findings indicated that aberrant methylation of ncRNA regulators and target mRNAs, might result in the aberrant expression of genes and, subsequently, might cooperatively mediate pathway dysregulation, leading to the development and progression of breast cancer (Figure 7C).

## Discussion

In this study, we characterized genome-wide methylation patterns in breast cancer based on published high-throughput sequencing methylation data. Our study reveals that greater number of ncRNA promoters were aberrantly methylated, including lncRNAs and miRNAs. Moreover, we summarized five aberrant methylation patterns of ncRNA promoters and found that aberrant methylation not only on CGIs, but also on 5' and 3' CGI shores. Because of a lack of genome-wide methylation profiling, investigating the methylation patterns of CGI shores in breast cancer is difficult. Here, we found that the methylation patterns of both CGI shores in breast cancer are consistent with previous results for protein coding genes in colorectal cancer<sup>47</sup> and prostate cancer<sup>48</sup>. Therefore, although coding RNAs and nc-RNAs function distinctly, they share similar aberrant methylation patterns.

Despite a growing consensus that long intergenic ncRNAs (lincRNAs) are modulators of cancer, the understanding of DNA

methylation patterns of lincRNAs in cancer remains limited. In our study, lincRNAs were found to comprise the majority of the aberrantly methylated lncRNA promoters in the breast cancer samples. As intergenic CGIs are typically methylated in actively transcribed genes, it would be interesting to determine what proportion of these lincRNAs is located in an actively transcribed gene. H3K4me3 is a prominent histone mark associated with active genes, and we found that approximately 59.62% of the hypermethylated lincRNAs were occupied by H3K4me3 signals in HMECs. By contrast, only 13.11% of the hypomethylated lincRNAs were occupied by H3K4me3 signals in HMECs, indicating that these lincRNAs might be silenced in normal cells. Moreover, we found that most hypermethylated lincRNAs were associated with decrease H3K4me3 density, and that the hypomethylated lincRNAs were associated with elevated H3K4me3 density. These results indicate that DNA methylation and histone modifications are two mechanisms that regulate lncRNA expression.

Many studies have shown that the expression of ncRNAs can be modified by genetic variants, such as copy number variants (CNVs) and single nucleotide polymorphisms (SNPs). Here, we found that approximately 11.8% and 9.71% of the hypermethylated lincRNAs and miRNAs, respectively were located in regions of recurring deletions<sup>49</sup>. This result is consistent with recent observations that although individual genetic alterations appear to affect the gene expression and methylation levels, these effects are rare<sup>50</sup>. Our observations demonstrated that ncRNA expression regulated by not only genetic but also epigenetic factors, which is consistent with recent observations<sup>51,52</sup>. Next, we identified a panel of ncRNA biomarkers that effectively discriminated between cancer and control samples. Among the lncRNA biomarkers identified in our study, 64.90% lincRNAs were also differentially expressed in the comparison of 84 breast cancer samples and 51 non-malignant breast tissue samples<sup>53</sup>. We found that the clustering of the expression of these lincRNAs also distinguished cancerous from noncancerous tissues (Figure S6). Finally, functional analysis indicated that aberrant methylation of ncRNAs widely disturbs processes associated with the development and progression of breast cancer.

Polycomb-targeted genes are frequent targets of aberrant DNA methylation in cancers, and some miRNA genes have been shown to be targeted by polycomb proteins. Therefore, we examined whether polycomb-targeted ncRNAs are more likely to be targets of aberrant DNA methylation in breast cancer. To this end, we used the repressed regions of a study by Ernst et al. as polycomb-targeted regions<sup>54</sup>. We found a significant overlap between the ncRNA promoters that were occupied by the polycomb and those that were aberrantly hypermethylated in cancer cells. Comparative analysis revealed that the ncRNA promoters that were targeted by polycomb were more likely to be hypermethylated in cancer. These observations provided additional evidence for the association between polycomb-mediated repression and aberrant DNA methylation in breast cancer. The high proportion of polycomb targets among ncRNA promoters may, to a certain extent, explain the high proportion of ncRNA promoters displaying aberrant methylation in cancer. However, it is possible that no direct functional association exists between these two phenomena. Such ncRNAs are unmethylated in normal cells, are repressed by polycomb proteins, and acquire DNA methylation as an alternative silencing mechanism. These ncRNAs, perform tumour-suppressive functions, are inducible in normal cells, and are repressed by polycomb. Hypothetically, the epigenetic switch to DNA methylation-mediated repression reduces the plasticity of the regulatory program of these ncRNAs, mandating the silencing of the key ncRNA regulators and contributing to the abnormal growth potential of the cell.

In summary, our data demonstrates the suitability of MBDCap-seq to investigate cancer methylomes and identifies novel epigenetically dysregulated ncRNAs (lncRNAs and miRNAs). This study pre-



sents the aberrant methylation patterns of ncRNAs, providing a highly valuable resource for the investigation of the epigenetic regulation of breast cancer.

## Methods

**Genome-wide DNA methylation profiles.** To explore the DNA methylation pattern of ncRNAs in breast cancer, DNA methylation data generated using the MBDCap-seq protocol were obtained from CMS<sup>23</sup>, consisting of 77 cancer samples and 10 control samples. The genome-wide methylation intensity values were directly downloaded and quantified as the numbers of reads uniquely mapped to each 100-bp genomic bin. Next, the normalization was performed based on the linear method<sup>55</sup>.

**Genomic coordinates of ncRNA and protein-coding genes.** We downloaded the human lncRNA annotation from GENCODE (version 18), and the primary miRNA genomic coordinates were obtained from the miRBase database (version 20.0)<sup>12</sup>. After converting the genomic location to hg18 using the LiftOver tool in the UCSC browser, 13,459 lncRNAs and 1866 pri-miRNAs were analyzed. In addition, the genomic coordinates of the protein-coding gene were downloaded from the UCSC browser, and the RefSeq table was used. Moreover, the ncRNA and mRNA promoters were defined as the  $\pm 2$  kb region surrounding the TSS of each lncRNA/mRNAs or the start site of each pri-miRNA<sup>56</sup>.

**Genome-wide expression of ncRNAs.** The expression data assessed using RNA-seq for two breast cancer cell lines (MCF-7 and HCC1954) and one control cell line (HMEC) were considered, and the raw reads data were downloaded from the ENCODE project<sup>57</sup> and the study by Hon et al.<sup>1</sup>. After mapping and aligning the RNA-seq data using the TopHat and Cuffdiff program according to the default parameters<sup>58</sup>, the expression levels of the lncRNAs were measured as 'fragments per kilobase of exon model per million mapped reads' (FPKM).

The miRNA expression profiles of the breast cancer tissue samples were downloaded from The Cancer Genome Atlas (TCGA) (Table S6), which were measured using miRNA-Seq. The reads per million (RPM) mapped to each miRNA were calculated, and a total of 1046 pre-miRNAs were analyzed.

**Annotation of genomic features.** The coordinates of the CGIs were primarily obtained from the UCSC database. Similar to previous studies<sup>47</sup>, the 5' and 3' CGI shores were defined as the 2 kb on either side of each CGI.

**Identification of the differentially methylated ncRNAs.** An ncRNA or protein-coding gene was considered to be differentially methylated only if at least one DMRs resided in or overlapped with its promoter region. Here, we required that the overlap region include at least 50% of the DMR. Furthermore, we detected both the CGI and CGI shore regions displaying differential methylation by using the same procedure.

To detect the DMRs, differentially methylated bins were first identified. A 100-bp genomic bin was considered to display significantly altered methylation levels between breast cancer and control samples when following two criteria were met: (i) 2-fold or greater change in the methylation level and (ii) adjusted  $P < 0.01$ , which was calculated using Wilcoxon rank-sum test followed by multiple testing correction using the BH method. Furthermore, consecutive hyper-/hypo-methylated bins with no gap were merged into DMRs.

**Identification of the patterns of aberrant methylation.** The aberrant ncRNA methylation patterns were further investigated in the context of CGIs, and five methylation patterns were classified: (1) aberrant methylation predominantly confined to the CGIs, in which the overlap region within the promoter was greater than 50% of either region; (2) aberrant methylation was strictly located 5' of the CGIs; (3) aberrant methylation was strictly located 3' of the CGIs; (4) aberrant methylation that overlapped with the CGI shores or an overlap of less than 50% between the CGIs and the DMR; and (5) aberrantly methylated promoters lacking CGIs.

**Aberrant expression of ncRNAs regulated by DNA methylation.** To investigate the change in expression of each ncRNA caused by its aberrant methylation, differential expression analysis was performed. The differentially expressed miRNAs were detected using student's t-test followed by multiple testing correction using the BH method (adjusted p-values at 5% thresholds). The lncRNAs displaying greater than two-fold changes in expression in the cell lines were considered to be differentially expressed. Considering the repressive effects of DNA methylation, we further required opposing directional changes between methylation and expression.

**Validation of the ncRNA biomarkers.** The quality of the lncRNA or miRNA biomarkers was evaluated using plotting ROC curves at various thresholds of the Z-score-transformed methylation or expression levels. First, we performed Z-score transformation on the methylation or expression levels across the samples for each lncRNA or miRNA. Then, we summarized the Z-scores for these ncRNAs into one score for each sample. The validation procedure was conducted using the MedCalc program.

1. Hon, G. C. *et al.* Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome research* **22**, 246–258, doi:10.1101/gr.125872.111 (2012).

2. Hervouet, E., Cartron, P. F., Jouvenot, M. & Delage-Mourroux, R. Epigenetic regulation of estrogen signaling in breast cancer. *Epigenetics* **8**, 237–245, doi:10.4161/epi.23790 (2013).
3. Hill, V. K. *et al.* Genome-wide DNA methylation profiling of CpG islands in breast cancer identifies novel genes associated with tumorigenicity. *Cancer research* **71**, 2988–2999, doi:10.1158/0008-5472.CAN-10-4026 (2011).
4. Fackler, M. J. *et al.* Genome-wide methylation analysis identifies genes specific to breast cancer hormone receptor status and risk of recurrence. *Cancer research* **71**, 6195–6207, doi:10.1158/0008-5472.CAN-11-1630 (2011).
5. Cui, P. *et al.* A comparison between ribo-minus RNA-sequencing and polyA-selected RNA-sequencing. *Genomics* **96**, 259–265, doi:10.1016/j.ygeno.2010.07.010 (2010).
6. Pandey, R. R. *et al.* Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* **32**, 232–246, doi:10.1016/j.molcel.2008.08.022 (2008).
7. Flintoft, L. Non-coding RNA: Structure and function for lncRNAs. *Nat Rev Genet* **14**, 598, doi:10.1038/nrg3561 (2013).
8. Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and functions of long noncoding RNAs. *Cell* **136**, 629–641, doi:10.1016/j.cell.2009.02.006 (2009).
9. Nie, Y. *et al.* Long non-coding RNA HOTAIR is an independent prognostic marker for nasopharyngeal carcinoma progression and survival. *Cancer science* **104**, 458–464, doi:10.1111/cas.12092 (2013).
10. Augoff, K., McCue, B., Plow, E. F. & Sossey-Alaoui, K. miR-31 and its host gene lncRNA LOC554202 are regulated by promoter hypermethylation in triple-negative breast cancer. *Molecular cancer* **11**, 5, doi:10.1186/1476-4598-11-5 (2012).
11. Mercer, T. R. & Mattick, J. S. Structure and function of long noncoding RNAs in epigenetic regulation. *Nat Struct Mol Biol* **20**, 300–307, doi:10.1038/nsmb.2480 (2013).
12. Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A. & Enright, A. J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic acids research* **34**, D140–144, doi:10.1093/nar/gkj112 (2006).
13. Iorio, M. V. & Croce, C. M. MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review. *EMBO Mol Med* **4**, 143–159, doi:10.1002/emmm.201100209 (2012).
14. Wu, W. K. *et al.* MicroRNA dysregulation in gastric cancer: a new player enters the game. *Oncogene* **29**, 5761–5771, doi:10.1038/onc.2010.352 (2010).
15. Croce, C. M. Causes and consequences of microRNA dysregulation in cancer. *Nat Rev Genet* **10**, 704–714, doi:10.1038/nrg2634 (2009).
16. Suzuki, H., Maruyama, R., Yamamoto, E. & Kai, M. DNA methylation and microRNA dysregulation in cancer. *Mol Oncol* **6**, 567–578, doi:10.1016/j.molonc.2012.07.007 (2012).
17. Baer, C., Claus, R. & Plass, C. Genome-wide epigenetic regulation of miRNAs in cancer. *Cancer research* **73**, 473–477, doi:10.1158/0008-5472.CAN-12-3731 (2013).
18. Weber, B., Stresmann, C., Brueckner, B. & Lyko, F. Methylation of human microRNA genes in normal and neoplastic cells. *Cell Cycle* **6**, 1001–1005 (2007).
19. Shen, J. *et al.* Genome-wide aberrant DNA methylation of microRNA host genes in hepatocellular carcinoma. *Epigenetics* **7**, 1230–1237, doi:10.4161/epi.22140 (2012).
20. Zhang, Z. *et al.* Epigenetic Silencing of miR-203 Upregulates SNAI2 and Contributes to the Invasiveness of Malignant Breast Cancer Cells. *Genes Cancer* **2**, 782–791, doi:10.1177/1947601911429743 (2011).
21. Li, D. *et al.* Analysis of MiR-195 and MiR-497 expression, regulation and role in breast cancer. *Clin Cancer Res* **17**, 1722–1730, doi:10.1158/1078-0432.CCR-10-1800 (2011).
22. Vrba, L., Munoz-Rodriguez, J. L., Stampfer, M. R. & Futscher, B. W. miRNA gene promoters are frequent targets of aberrant DNA methylation in human breast cancer. *PLoS One* **8**, e54398, doi:10.1371/journal.pone.0054398 (2013).
23. Gu, F. *et al.* CMS: a web-based system for visualization and analysis of genome-wide methylation data of human cancers. *PLoS ONE* **8**, e60980, doi:10.1371/journal.pone.0060980 (2013).
24. Yan, L. X. *et al.* MicroRNA miR-21 overexpression in human breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis. *RNA* **14**, 2348–2360, doi:10.1261/rna.1034808 (2008).
25. Ma, L., Teruya-Feldstein, J. & Weinberg, R. A. Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature* **449**, 682–688, doi:10.1038/nature06174 (2007).
26. Ponjavic, J., Oliver, P. L., Lunter, G. & Ponting, C. P. Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain. *PLoS genetics* **5**, e1000617, doi:10.1371/journal.pgen.1000617 (2009).
27. Jia, H. *et al.* Genome-wide computational identification and manual annotation of human long noncoding RNA genes. *RNA* **16**, 1478–1487, doi:10.1261/rna.1951310 (2010).
28. Koga, Y. *et al.* Genome-wide screen of promoter methylation identifies novel markers in melanoma. *Genome research* **19**, 1462–1470, doi:10.1101/gr.091447.109 (2009).
29. Bajic, V. B., Choudhary, V. & Hock, C. K. Content analysis of the core promoter region of human genes. *In Silico Biol* **4**, 109–125 (2004).





30. Brinkman, A. B. *et al.* Sequential CHIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome research* **22**, 1128–1138, doi:10.1101/gr.133728.111 (2012).
31. Liao, Q. *et al.* Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. *Nucleic acids research* **39**, 3864–3878, doi:10.1093/nar/gkq1348 (2011).
32. Guo, X. *et al.* Long non-coding RNAs function annotation: a global prediction method based on bi-colored networks. *Nucleic acids research* **41**, e35, doi:10.1093/nar/gks967 (2013).
33. Wamstad, J. A. *et al.* Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell* **151**, 206–220, doi:10.1016/j.cell.2012.07.035 (2012).
34. Dolfi, S. C. *et al.* The metabolic demands of cancer cells are coupled to their size and protein synthesis rates. *Cancer Metab* **1**, 20, doi:10.1186/2049-3002-1-20 (2013).
35. Wang, C., Bian, Z., Wei, D. & Zhang, J. G. miR-29b regulates migration of human breast cancer cells. *Molecular and cellular biochemistry* **352**, 197–207, doi:10.1007/s11010-011-0755-z (2011).
36. Zhang, L., Sullivan, P. S., Goodman, J. C., Gunaratne, P. H. & Marchetti, D. MicroRNA-1258 suppresses breast cancer brain metastasis by targeting heparanase. *Cancer research* **71**, 645–654, doi:10.1158/0008-5472.CAN-10-1910 (2011).
37. Rice, J. C., Ozcelik, H., Maxeiner, P., Andrulis, I. & Futscher, B. W. Methylation of the BRCA1 promoter is associated with decreased BRCA1 mRNA levels in clinical breast cancer specimens. *Carcinogenesis* **21**, 1761–1765 (2000).
38. Cao, W. *et al.* Integrated analysis of long noncoding RNA and coding RNA expression in esophageal squamous cell carcinoma. *International journal of genomics* **2013**, 480534, doi:10.1155/2013/480534 (2013).
39. Cheng, C., Ung, M., Grant, G. D. & Whitfield, M. L. Transcription factor binding profiles reveal cyclic expression of human protein-coding genes and non-coding RNAs. *PLoS computational biology* **9**, e1003132, doi:10.1371/journal.pcbi.1003132 (2013).
40. Lucas, J. J., Domenico, J. & Gelfand, E. W. Cyclin-dependent kinase 6 inhibits proliferation of human mammary epithelial cells. *Molecular cancer research: MCR* **2**, 105–114 (2004).
41. Pic-Taylor, A., Darieva, Z., Morgan, B. A. & Sharrocks, A. D. Regulation of cell cycle-specific gene expression through cyclin-dependent kinase-mediated phosphorylation of the forkhead transcription factor Fkh2p. *Molecular and cellular biology* **24**, 10036–10046, doi:10.1128/MCB.24.22.10036-10046.2004 (2004).
42. Dhillon, A. S., Hagan, S., Rath, O. & Kolch, W. MAP kinase signalling pathways in cancer. *Oncogene* **26**, 3279–3290, doi:10.1038/sj.onc.1210421 (2007).
43. Hynes, N. E. Tyrosine kinase signalling in breast cancer. *Breast cancer research: BCR* **2**, 154–157 (2000).
44. Fernandez-Medarde, A. & Santos, E. Ras in cancer and developmental diseases. *Genes Cancer* **2**, 344–358, doi:10.1177/1947601911411084 (2011).
45. Sanchez-Munoz, A. *et al.* Lack of evidence for KRAS oncogenic mutations in triple-negative breast cancer. *BMC cancer* **10**, 136, doi:10.1186/1471-2407-10-136 (2010).
46. Cho, W. C. OncomiRs: the discovery and progress of microRNAs in cancers. *Molecular cancer* **6**, 60, doi:10.1186/1476-4598-6-60 (2007).
47. Irizarry, R. A. *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature genetics* **41**, 178–186, doi:10.1038/ng.298 (2009).
48. Kim, J. H. *et al.* Deep sequencing reveals distinct patterns of DNA methylation in prostate cancer. *Genome research* **21**, 1028–1041, doi:10.1101/gr.119347.110 (2011).
49. Beroukhi, R. *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905, doi:10.1038/nature08822 (2010).
50. Gibbs, J. R. *et al.* Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS genetics* **6**, e1000952, doi:10.1371/journal.pgen.1000952 (2010).
51. Jacobsen, A. *et al.* Analysis of microRNA-target interactions across diverse cancer types. *Nat Struct Mol Biol* **20**, 1325–1332, doi:10.1038/nsmb.2678 (2013).
52. Aure, M. R. *et al.* Individual and combined effects of DNA methylation and copy number alterations on miRNA expression in breast tumors. *Genome biology* **14**, R126, doi:10.1186/gb-2013-14-11-r126 (2013).
53. Varley, K. E. *et al.* Recurrent read-through fusion transcripts in breast cancer. *Breast cancer research and treatment* **146**, 287–297, doi:10.1007/s10549-014-3019-2 (2014).
54. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49, doi:10.1038/nature09906 (2011).
55. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome research* **22**, 1775–1789, doi:10.1101/gr.132159.111 (2012).
56. Wu, S. C., Kallin, E. M. & Zhang, Y. Role of H3K27 methylation in the regulation of lncRNA expression. *Cell research* **20**, 1109–1116, doi:10.1038/cr.2010.114 (2010).
57. Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74, doi:10.1038/nature11247 (2012).
58. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* **28**, 511–515, doi:10.1038/nbt.1621 (2010).

## Acknowledgments

This work was supported by the National High Technology Research and Development Program of China [863 Program, Grant No. 2014AA021102], the National Natural Science Foundation of China [Grant Nos. 91439117, 61473106, and 61203264], the China Postdoctoral Science Foundation [Grant No.2014T70364 and LBH-Z14134] and Weihai Yu Youth Science Fund Project of Harbin Medical University.

## Author contributions

X.L., J.X. and Y.L. participated in the design and coordination of the study. Y.L., Y.Z. and S.L. acquired and analyzed the data. Y.L., S.L. and J.C. carried out the function analysis. Y.W. and Y.L. participated in the figures preparation. J.L. and Y.L. prepared the supplemental files. Y.L. and J.X. wrote the manuscript. All authors reviewed the manuscript. All authors have read and approved the manuscript and its contents, and are aware of responsibilities connected to authorship.

## Additional information

**Supplementary information** accompanies this paper at <http://www.nature.com/scientificreports>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Li, Y. *et al.* Genome-wide DNA methylome analysis reveals epigenetically dysregulated non-coding RNAs in human breast cancer. *Sci. Rep.* **5**, 8790; DOI:10.1038/srep08790 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>