Locus-specific enrichment analysis of 5-hydroxymethylcytosine reveales novel genes associated with breast carcinogenesis

Running Title

Landscape of 5-methylcytosine and 5-hydroxymethylcytosine of breast cancer

Authors

Deepa Ramasamy[1], Arunagiri Kuha Deva Magendhra Rao[1], Meenakumari Balaiah[1], Arvinden Vittal Rangan[1], Shirley Sundersingh[2], Sridevi Veluswami[3], Rajkumar Thangarajan[1], Samson Mani[1*]

Affiliations

1. Department of Molecular Oncology, Cancer Institute (WIA), 38, Sardar Patel Road, Chennai-600036, Tamilnadu, India.

2. Department of Oncopathology, Cancer Institute (WIA), 38, Sardar Patel Road, Chennai- 600036, Tamilnadu, India.

3. Department of Surgical Oncology, Cancer Institute (WIA), 38, Sardar Patel Road, Chennai-600036, Tamilnadu, India.

***Corresponding author:** Dr. Samson Mani, Ph.D.
Professor
Department of Molecular Oncology,
Cancer Institute (WIA),
38, Sardar Patel Road, Adyar,
Chennai 600036, India.
samsonn.m@gmail.com
Telephone: 044-22350131 Extn: 196

Abstract

Background: An imbalance in DNA methylation is a hallmark epigenetic alteration in cancer. The conversion of 5-methylcytosine (5-mC) to 5-hydroxymethyl cytosine (5-hmC), which causes the imbalance, results in aberrant gene expression. The precise functional role of 5-hydroxymethylcytosine in breast cancer remains elusive. In this study, we describe the landscape of 5-mC and 5-hmC and their association with breast cancer development.

Results: We found a distinguishable global loss of 5-hmC in the localized and invasive types of breast cancer, which correlate strongly with *TET* expression. Genome-wide analysis revealed a unique 5-mC and 5-hmC signature in breast cancer. The differentially methylated regions (DMRs) were primarily concentrated in the proximal regulatory regions such as the promoters and UTRs, while the differentially hydroxymethylated regions (DhMRs) were densely packed in the distal regulatory regions such as the intergenic regions (>-5 kb from TSSs). Our results indicate 4809 DMRs and 4841 DhMRs associated with breast cancer. Validation of nine 5-hmC enriched loci in a distinct set of breast cancer and normal samples, positively correlated with their corresponding gene expression. The novel 5-hmC candidates such as *TXNL1*, *CNIH3*, and *BNIPL* implicate a pro-oncogenic role in breast cancer. Therefore, 5-hmC modified regions could be used as promising diagnostic and therapeutic markers for breast cancer.

Conclusion: Global loss of 5-hmC is associated with down-regulation of the *TET 1* and *TET3* genes. Genome-wide profiling has revealed a profound imbalance in the region-specific distribution of 5-mC and 5-hmC in breast cancer. Predominant 5-hmC modifications are localized at distal gene regulatory sites. Novel 5-hmC candidates associated with breast cancer have been identified. Hence, these results provide new insights in the loci-specific accumulation of 5-mC and 5-hmC which are aberrantly methylated and demethylated in breast cancer.

MAIN TEXT

Introduction

DNA methylation imbalance is one of the hallmark epigenetic events in cancer. Cytosine DNA methylation (5-methylcytosine or 5-mC) occurs at the gene promoter and is often associated with gene repression, while its oxidized form 5-hydroxymethylcytosine (5-hmC) relaxes the repression (*1*, *2*). The oxidation of 5-mC to 5-hmC is catalyzed by the *TET* family of genes *TET1*, *TET2*, and *TET3* (*3–7*). Based on tissue specificity, 5-hmC levels can vary between 0.1% and 1% of the human genome (*8*). The increase in 5-hmC is strongly associated with transcriptional activation (*9*). Effective binding of methylation readers such as *MBD3* and *MeCP2* preferentially to 5-hmC results in active transcriptional assembly and activity (*10*, *11*). It turned out that 5-hmC is the stable epigenetic modification involved in the transcription machinery and not just serving as an intermediate in the demethylation process. The imbalance between 5-mC and 5-hmC is of recent interest as both are associated with gene expression and lead to carcinogenesis.

A global reduction of 5-hmC was evident in several cancers (*8*, *12–16*). Studies on melanoma, pancreatic cancer, lung cancer, and prostate cancer suggest that aberrant 5-mC and 5-hmC levels may predispose to tumor progression (*17–24*). However, there is limited evidence for 5-mC and 5-hmC dynamics in breast cancer. We hypothesize an imbalance among the genomic 5-mC and 5-hmC levels that contributes to breast carcinogenesis. Previous reports affirm that 5-hmC levels depends on tissue-specific *TET* expression. Particularly, the genes *TET1* and *TET2* downregulation has been reported to alter the 5-hmC levels (*25*). A recent study on breast cancer showed the altered 5-hmC profiles and their association with lymph node metastases (*26*). In breast cancer, the locus-specific deposition of 5-hmC and its functional role in the control of gene expression are poorly understood. Emerging enrichment approaches can identify 5-mC and 5-hmC genomic regions with single-base resolution and describe the differentially methylated regions (DMRs) and differentially hydroxymethylated regions (DhMRs) in cancer (*27*, *28*). Determining the 5-hmC modified genomic regions in breast cancer will be of promising diagnostic and therapeutic markers.

In this study we find that global methylation and hydroxymethylation levels were drastically reduced in breast cancer tissues. The global 5-hmC reduction was associated with the downregulation of the *TET1* and *TET3* genes. The genome-wide analysis revealed differentially methylated and differentially hydroxymethylated breast cancer loci. We also identified a strong correlation between the 5-hmC alterations and the gene expression changes. Altogether, the study

provides a comprehensive genome-wide distribution of the 5-mC and 5-hmC and also the imbalance in the DNA methylation machinery that leads to breast cancer development.

Results

Loss of 5-hmC is Associated with *TET 1 and TET3* Downregulation in Breast Cancer

Global levels of 5-hmC, 5-mC were first quantified in breast cancer [invasive ductal carcinoma (IDC) and ductal carcinoma *in situ* (DCIS)] paired normal (PN), and apparent normal (AN) tissues. We found a significant decrease in 5-hmC levels in IDC *vs* PN (FC = –2.58, $p = 0.0003$) and DCIS *vs* AN (FC = –2.08, $p = 0.0324$) (Figure 1a). Although global 5-mC levels were reduced in IDC *vs* PN (FC = –2.28, $p = 0.0014$), there was no significant difference between the DCIS *vs* AN group ($p = 0.5467$) (Figure 1b). The results imply that the "global loss of 5-hmC" is a characteristic epigenetic alteration of localized and invasive breast cancer. The differential expression of *TET* genes leads to the altered 5-hmC levels seen in breast cancer. Therefore, we quantified the *TET* gene expression levels and found that the genes *TET2* (FC = +2.02, $p = 0.0317$) (Figure 1d), and *TET3* (FC = +2.0, $p = 0.0159$) (Figure 1e) were upregulated in DCIS *vs* AN group. However, *TET1* (FC = –2.08, $p = 0.0266$) (Figure 1c) and *TET3* (FC = –2.0, $p = 0.026$) (Figure 1e) genes were downregulated in the IDC *vs* PN and IDC *vs* AN groups. Spearman's rank test showed no significant correlation of *TET* genes with global 5-hmC levels in PN tissues (Figure 1f- h), while it showed a significant positive correlation of *TET1* ($r = 0.544$, $p = 0.05$) (Figure 1i) and *TET3* ($r = 0.5662$, $p = 0.0437$) (Figure 1k) genes with the global loss of 5-hmC in breast cancer but not with the *TET2* gene (Figure 1j). In addition to *TET1*, we report here that *TET3* is also associated with a global loss of 5-hmC in breast cancer.

Relative Abundance of 5-hmC in Breast Cancer and their enrichment at distal regulatory sites

We performed the enrichment of genomic 5-mC and 5-hmC specific regions followed by high-throughput sequencing and achieved ~14 million reads from 5-hmC-enriched libraries and ~21 million reads from 5-mC-enriched libraries. Almost 99.7% of the 5-hmC reads and ~98.94% of 5-mC reads were mapped effectively against the reference genome. Principal component analysis (PCA) and hierarchical clustering analysis (HCA) showed a clear pattern of segregation from the tumor [IDC and DCIS] to normal samples [PN and AN] (SF 1a-d). MACS-2 peak calling of the mapped reads resulted in a total of 3.3 million 5-hmC-enriched peak sets and 4.7 million 5-mC-enriched peak sets. Differential peak calling analysis between breast cancer and normal groups

identified 4809 differential methylated regions (DMRs) ($p<0.01$, FDR<0.05) (Figure 2a) and 4841 differential hydroxymethylated regions (DhMRs) ($p<0.01$, FDR<0.05) (Figure 2b) (Sf 1a-b). The distribution of peaks across the chromosomes showed a higher peak intensity of DhMR over DMR (Figure 2c-d) (window size: 1×10^{-6}). A significantly higher peak intensity of DhMR indicates a potential difference in loci-specific 5-hmC levels between breast tumor and paired normal samples.

The peaks characterized by their genomic features indicated that both DMR and DhMR were moderately found in the gene body (28.97% and 37.31%, respectively), particularly in the intronic regions but not in the exons. The DMR profile was high in the promoter (28.57%) regions. However, only a mere 8.82% enrichment of DhMR was found in the promoter region compared to massive accumulation in the distal intergenic regions **(**43.11%**) [**(Figure 2e) (SF 2a-b and ST 1a-b)]. Therefore, we show here that the distribution in the gene body does not invariably differ between the two modifications, but DMR is typically enriched at the proximal regulatory sites (promoter), while DhMR at the distal regulatory sites (intergenic region). We found the enrichment of 5-mC and 5-hmC were significantly distinguishable between tumor and normal tissues. Further, the enriched peak sets of the DhMRs and DMRs were analyzed for the transcription factor binding sites. We found that the accumulation in the promoter regions of DMRs in the interval 1500 bp upstream and 1500 bp downstream of the TSSs (read count frequency>6.5×10^{-5}) was higher than that of DhMRs (read count frequency <3×10^{-5}) (SF 2c). While in DhMRs, the accumulation was observed in the distal intergenic regions upstream 5 kb from the TSSs (read count frequency>6.5×10^{-5}) [(Figure 2f-g) (SF 2d)]. We also used LOLA-Web to identify the locus overlap between the 5-hmC sites and the regulatory sites in the distal intergenic regions (i.e., regions>5 kb from the TSSs). Tumor and PN peak sets of 5-hmC were tested against the ENCODE data set with the reference genome hg19 and normalized with preloaded Tiles1000.hg19.bed. Tumor-specific 5-hmC peak sets are significantly associated with the enhancer sites of the breast cancer cell line MCF-7 regions such as H3K4me1, H3K4me3, H3K14ac, and H3K9ac (log (p-value) > 300). Also, enhancer sites were overlapped with loci-specific 5-hmC levels of the candidate genes in the breast cancer cell lines such as MDA-MD-468, MDA-MB-231 and MCF-7 regions but not in the normal luminal cell line MCF-10A (SF 3a-f). Hence, the results suggest that DhMRs were widespread in the distal regulatory regions, while DMRs accumulated in the proximal regulatory regions of the breast cancer genome.

Locus-Specific Imbalance of DhMRs and DMRs in Breast Cancer

To determine the exact loci and the differential distribution of 5-hmC and 5-mC accumulation in breast cancer, we identified 35 hyper-hmC loci (Sf 2a), and 30 hypo-hmC loci (Sf 2b). The hyper-hmC loci included coding genes (*GALC*, *BNIPL*, *TXNL1*, *CNIH3*, etc.), lncRNA (*LINC00535*, *LINC00662*, and *PTPRN2* lncRNA), and microRNA (*MIR4278*, *MIR1204*, *MIR944*, and *MIR921*). We found 26 coding genes (*ZBTB16*, *SP8*, *THRB*, *HIC2*, etc.,) and only four non-coding loci (*MIR4417*, *MIR3612*, *LINC00911*, and *LINC00417*) among the hypo-hmC-specific regions. A total of 57 hyper-mC loci inclusive of 53 coding genes (*CCDC181*, *SIM2*, *ID4*, etc.) and 4 non-coding genes (*LINC01257*, *LOC728989*, *MIR5087*, and *MIR183*) were obtained. The hypo-mC loci consisted of 24 coding genes (*OPCML*, *MKI67*, and *SPOCK1*) and 6 were non-coding (*MIR548AR*, *LINC02347*, *MIR744*, *MIR3612*, etc.) (Figure 3a) (Sf. 2c-d). Further, we validated five hyper-hmC and four hypo-hmC loci in breast cancer and normal tissues. The results confirmed the 5-hmC gain of *TXNL1* (FC=4, p=0.0102) (Figure 3b-c), *CNIH3* (FC=2, p=0.0242) (Figure 3d), while for *BNIPL*, *A4GALT* and *CBLN4* no statistical significance was observed, although they follow the same trend (Figure 3e-g). On the other hand, hypo-hmC candidates *CHODL* showed a two-fold loss of 5-hmC (p = 0.0416) (Figure 3h). Other loci such as *ZBTB16*, *HIC2*, and *SP8* showed a trend towards 5-hmC loss in tumors but did not show any statistical significance (Figure 3i-k). Our validation analysis confirmed that the gain of 5-hmC in *TXNL1*, *BNIPL*, *CNIH3* and loss of 5-hmC in *CHODL*, *ZBTB16*, *SP8*, *HIC2* in breast cancer samples. The gene function prediction by g: Profiler indicated 5-mC and 5-hmC genes were related to cell cycle regulation, cell cycle inhibition, and cell proliferative signals that could potentially affect breast cancer development and progression (Sf 3a-d). The gene ontology and KEGG pathway enrichment map of DhMRs and DMRs also revealed the association of breast cancer-specific 5-hmC and 5-mC with transcriptional machinery (SF 4a-d). Altered levels of 5-mC and 5-hmC eventually control and determine the progression of breast cancer development. Extensive functional analysis of the identified loci will elucidate the significance of the 5-mC and 5-hmC imbalance in breast cancer development.

Association of 5-mC and 5-hmC modifications with Gene Expression

We investigated the aberrant levels of methylation and hydroxymethylation in the locus-specific aspect and its impact on the regulation of gene expression using TCGA-breast cancer data set (UALCAN) (Table 1). We found the hyper-methylation of *CCDC181* (beta value > 0.5, $p < 0.05$), *ID4* (beta value > 0.5, $p < 0.05$) and hypo-methylation of *MKI67*, *OPCML*, and *SPOCK1* (beta value < 0.3, $p < 0.05$). Correspondingly, *CCDC181* (normalized TPM count = –0.094, $p < 0.1$)

(Figure 4a), *OR4F29* (normalized TPM count = –0.005, $p < 0.1$) (Figure 4b), and *ID4* (normalized TPM count = –58.839, $p < 0.001$) (Figure 4c), were downregulated in tumor samples ($n = 1094$) while, *MKI67* (normalized TPM count = 9.94, $p < 0.001$) and *SPOCK1* (normalized TPM count = 3.26, $p < 0.001$) were found to be over-expressed in tumor samples but not OPCML (Figure 4d-f). The inverse correlation confirmed that hypermethylation leads to the suppression of gene expression and hypomethylation leads to overexpression of the gene. Further, we found that hyper-hmC candidates, *BNIPL* (normalized TPM count = 5.344, $p <0.05$), *CNIH3* (normalized TPM count = 0.778, $p <0.005$), and *TXNL1* (normalized TPM count = 14.549, $p < 0.001$) to be upregulated and the hypo-hmC candidates such as *ZBTB16* (normalized TPM count = –13.933, $p < 0.001$), *HIC2* (normalized TPM count = –0.436, $p < 0.001$), *CHODL* (normalized TPM count = –1.093, $p < 0.001$), *THRB* (normalized TPM count = –13.792, $p < 0.001$) and *RAPGEF2* (normalized TPM count = –9.296, $p < 0.001$) were significantly downregulated (Figure 4g-p). Several loci relax the repressive methylation marks by increasing the 5-hmC levels leading to gene activation. In this study, we found that three candidate genes, *TXNL1*, *CNIH3*, and *BNIPL*, showed an increase in 5-hmC associated with gene overexpression. The direct proportionality between 5-hmC and gene expression reinstate that gain of 5-hmC activates gene transcription. We further investigated the influence of gene expression and overall survival of the candidate genes with hyper-hmC specifications. The results indicated that overexpression of genes such as *TXNL1* ($p = 0.042$) (SF 5a), *CNIH3* ($p = 0.26$) (SF 5b) and *BNIPL* ($p = 0.0001$) (SF 5c) was associated with poor overall survival in breast cancer patients.

Discussion

The present study illustrated the methylation and hydroxymethylation landscape of the breast cancer genome. Initial findings showed that the downregulation of the *TET* genes causes a global 5-hmC reduction in breast cancer. The genome-wide profiling revealed a higher 5-mC accumulation around the TSSs from –1.5k to +1.5k, while the 5-hmC accumulated in the intergenic regions (>–5 kb away from TSSs). Thus, DMRs have mainly been associated with proximal gene regulation and DhMRs with distal regulation. We found an intergenic and gene body gain of 5-hmC associated with gene overexpression and a loss of 5-hmC towards downregulation of the corresponding genes. The study results show that the imbalance between 5-mC and 5-hmC is a novel phenomenon orchestrating the epigenetic machinery of breast cancer.

Previous studies also reported global 5-mC and 5-hmC loss in various cancers, including breast cancer (*17*, *23*, *29*, *30*). The tissue-specificity and the alterations of *TET* gene expression in

advancing cancer stages determine the 5-hmC levels and the demethylation process (*14*, *18*, *24*, *31*, *32*). The cytoplasmic mislocalization of *TET1* in ER/PR-negative subtypes of IDC and DCIS was directly proportional to the global reduction in 5-hmC levels (*25*). We show here, that the global reduction in the 5-hmC content in IDC is not only dependent on *TET1*, but also on *TET3* genes. The CXXC domains of TET1 and TET3 enhance the DNA binding efficiency at DNA demethylation sites (*33*). The expression and nuclear import of TET1 and TET3 facilitate an active oxidation process. Our finding shows the downregulation of TET3 also contribute to 5hmC loss in breast cancer tissues.

Genome-wide profiling showed the effects of promoter methylation in breast cancer. The advancement in enrichment strategies and next-generation sequencing has resulted in the contemporary analysis of 5-mC and 5-hmC describing their genomic signatures in breast cancer (*34*–*36*). In our study, we found the enrichment of 5-mC and 5-hmC significantly distinguishable between tumor and normal tissues. Higher hydroxymethylation levels were observed in the promoter and UTRs of DCIS tissues. On the other hand, the invasive type showed a higher accumulation in the gene body and intergenic regions than in the promoter regions. Hence, we speculate that accumulation of 5hmC in the preliminary stage of breast cancer occurs mainly at the proximal regulatory regions, reducing the suppression caused by 5-mC. In the locally advanced breast tumors, the enrichment at the distal intergenic regions implicates an enhancer like activity of 5-hmC. Previous studies also reported that 5-hmC could potentially act as an enhancer or super-enhancer elements ~5–10 kb and >20 kb away from the TSSs (*37*, *38*). The overlapping histone markers from the ENCODE roadmap project and 5-hmC sites emphasized the association of active enhancer sites in the breast cancer genome. The positive association of histone activation and 5-hmC gain suggest a synergistically enhanced gene.

Several loci relax the repressive methylation marks by increasing the 5-hmC levels leading to gene activation. In our results, we found *TXNL1* as a novel 5-hmC candidate gene in breast cancer with an increased 5-hmC level and corresponding gene overexpression. Survival analysis also indicated the overexpression of *TXNL1* and *BNIPL* in breast cancer being significantly associated with poor overall survival. Previous studies reported that induction of oxidative stress led to the overexpression of *TXNL1* associated with the downregulation of the DNA repair protein XRCC1, an accumulation of DNA damage, and *BCL-2* regulation (*39*). Further functional analysis on 5-hmC gain at *TXNL1* and other loci would be warranted to elucidate the mechanistic insight of 5-hmC in the transcriptional activation and breast cancer development.

The present study opens a new paradigm on the imbalance of 5-mC and 5-hmC in breast cancer. The study offers a detailed perspective on an epigenomic instability substantiated by the loss of 5-mC and 5-hmC in breast tumors. Global loss of 5-hmC is associated with *TET1* and *TET3* downregulation. Genome-wide profiling has revealed a profound imbalance in the region-specific distribution of 5-mC and 5-hmC in breast cancer. Predominant 5-hmC modifications localized at distal gene regulatory sites implicating a transcription enhancing function. The novel 5-hmC candidates identified in the study can be promising diagnostic and therapeutic markers for breast cancer.

Materials and Methods

Clinical Specimen

Breast cancer tissues of stages IIA–IV and paired normal (PN) tissue were obtained from the Tumor Bank, Cancer Institute (WIA), Chennai, India. Tissue samples were collected from the patient undergoing direct surgery for invasive ductal carcinoma (IDC) or ductal carcinoma *in situ* (DCIS). Tumor tissues (n = 15) were histopathologically confirmed to consist of >70% tumor cells, and paired non-cancerous tissue (n = 15) free of tumor cells was excised away from the tumor margin. Similarly, DCIS (n = 5) were obtained from patients undergoing a wide-excision biopsy, and absolute normal samples (n = 5) were collected from patients undergoing wide-excision biopsy for non-tumorous conditions like fibrosis or adenosis and histopathologically confirmed to be free of any tumor cells. An additional set of tumour samples (n = 30) and non-cancerous (n = 6) tissues were also collected for the validation study. Informed consent for participation and sampling was obtained from all patients. The study was approved by the Cancer Institute (WIA), Institutional Ethics Committee (IEC/2016/05).

Isolation of Genomic DNA

About 25 mg of tissue was homogenized and DNA was isolated using the Nucleospin Tissue DNA Kit (Macherey Nagel, GmbH) according to the manufacturer's instructions. The isolated DNA was quantified with Nanodrop ND-2000 and stored at –20 ° C until further use.

Estimation of Global Levels of 5-hmC and 5-mC

Genomic DNA (100 ng) was used for the estimation of global 5-hmC and 5-mC levels by ELISA using the Quest 5-hmC ELISA kit and the 5-mC DNA ELISA kit (Zymo research Inc, USA) according to the manufacturer's instructions.

RNA Isolation and *TET* Expression Assay

Briefly, tissues were homogenized and RNA was isolated using Nucleospin® RNA Isolation Kit (Macherey Nagel, GmbH). RNA was quantitated using Nanodrop ND-2000 and cDNA was synthesized from 500 ng of total RNA using a Quantitect® reverse transcription kit (Qiagen, USA). Gene expression analysis of *TET* 1, 2, and 3 was performed using TaqMan probes (Sf 4), TaqMan™ Universal Master Mix II, no UNG (Applied Biosystems, USA) and the Quant studio 12Kflex system (Applied Biosystems, USA).

Enrichment of 5-mC Modified DNA Regions and Library Preparation

Genomic DNA (1 μg) was fragmented to 300–600 bp after 25 cycles of 30 s of pulsed sonication in Bioruptor (Diagenode, Belgium). Fragmented DNA was end-repaired and adapter ligation was performed using the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs, USA) for Illumina. Furthermore, the fragments with a length of 400–500 bp (300 bp insert + 120 bp adapter) were size-selected using AMPure beads (Beckman Coulter, USA). Immunoprecipitation of methylated DNA was performed using the MagMeDIP kit (Diagenode, Belgium), and enriched DNA was purified using the I Pure kit (Diagenode, Belgium) according to the manufacturer's protocol. Purified DNA was indexed and amplified using the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs, USA).

Enrichment of 5-hmC Modified DNA Regions and Library Preparation

For the enrichment of 5-hmC-modified DNA, a reduced representation hydroxymethylation profiling by RRHP kit (Zymo Research Inc, USA) was carried out. The genomic DNA (1 μg) was digested using the *Msp1* enzyme and ligated with p5 and p7 adapters. The adapter-ligated fragments were glycosylated with UDP-glucose and T4-glycosyltransferase and digested again with *Msp1* to cleave adapters from non-glycosylated fragments. The fragments were size selected 400–500 bp (300 bp insert + 120 bp adapter) using AMPure XP beads (Beckman Coulter, USA) and amplified as libraries using RRHP™ 5-hmC Library Prep Kit (Zymo Research. Inc, USA).

Sequencing and Data Analysis

Enriched libraries were sequenced by 150 × 2 paired ends to generate 25 million reads in the Illumina Nextseq 500. The FASTQ files were quality checked by FastQC. Trimmomatic was used to trim the adapter sequence and the low-quality reads (Phred > 30) were discarded. The processed fastq files were then aligned to the reference genome (hg19) using the BWA mem algorithm. Aligned files were then converted into sorted bam files using samtools. The peaks were called with MACS2 and the peak files were used to find overlapping peaks in multiple files, and their raw counts were extracted with the DiffBind R package.

DMR Analysis

DMR analyses were carried out using the DESeq2 R package and the likelihood ratio test (LRT) was used to determine DMR with the padj value cut-off of ≤0.01. Furthermore, the DMR regions were filtered based on the number of peaks called in biological replicate with at least 10 | 10 | 5 | 5 (T | PN | DCIS | AN). Filtered DMR was annotated using the ChIPseeker R package with the UCSC hg19 known gene sets. The Z-Score was calculated from normalized counts and the heatmap was plotted using the Complex Heat map R package. The sorted bam files were indexed and the coverage profile was calculated using the Deep Tools bam Coverage with RPKM normalization and a bin size of 20 bp. The resulted bigwig files were used to visualize peak regions with IGV.

DhMR Analysis

For the DhMR analysis, RPKM values were extracted from the DiffBind R package and a global rank-invariant set normalization was carried out using the rank-invariant function from the Lumi R package. The Kruskal–Wallis test was performed and the *p*-value was adjusted with FDR. Cut-off of padj ≤0.1 was set to define DhMR. In addition, the DhMR peaks were filtered with the same criteria as DMR. The chromosomal distribution of DMR and DhMR was analyzed with the Karyoplot R-Package with a *p*-value of 0.05. The ideogram was examined for the DMR and DhMR regions with an FDR of 0.05.

Pathway Enrichment and Gene Ontology Analysis

Pathway enrichment and gene ontology analysis were performed with the g: Profiler and Cluster Profiler R package. For the *p*-adjusted corrected method, FDR with a *p*-value cut-off of 0.05 was used. Dot blots were generated based on the criteria mentioned above both for DMRs and DhMRs.

Validation of Loci-Specific 5-hmC Enriched Regions Using qPCR Assays

Briefly, the tumor (n = 30) and normal tissue (n = 6) were homogenized and DNA was isolated using the Nucleospin Tissue DNA Kit (Macherey Nagel, GmBH) according to the manufacturer's instructions. The genomic DNA was subjected to a 5-hmC-specific enrichment with EpiMark 5-hmC and 5-mC analysis kit (New England Biolabs, USA). The processed DNA samples were analyzed with qPCR using loci-specific primers (Sf 4). Gene expression, methylation analysis, and survival analysis were carried out with the UALCAN (*40*) web tool for 5-mC- and 5-hmC-specific candidate genes.

Statistical analysis

The correlation analysis was performed between the expression levels of *TET* enzymes and the global levels of 5-mC and 5-hmC distribution using GraphPad Prism v 7.0a (GraphPad Software, La Jolla, CA, USA). Spearman's rank correlation test was performed with a confidence interval of 95% for generating the correlation plot. Linear regression lines were generated based on the R-value of the entities. Wilcoxon signed-rank test was used for the nonparametric paired analysis. Mann–Whitney U test was performed for the nonparametric unpaired analysis. The p-value of <0.05 is considered a significant outcome.

References

1. M. Kulis, M. Esteller, DNA Methylation and Cancer. *Advances in Genetics*. **70**, 27–56 (2010).
2. L. D. Moore, T. Le, G. Fan, DNA methylation and its basic function. *Neuropsychopharmacology*. **38** (2013), , doi:10.1038/npp.2012.112.
3. M. Münzel, D. Globisch, T. Carell, 5-hydroxymethylcytosine, the sixth base of the genome. *Angewandte Chemie - International Edition*. **50** (2011), , doi:10.1002/anie.201101547.
4. C. Dahl, K. Grønbæk, P. Guldberg, Advances in DNA methylation: 5-hydroxymethylcytosine revisited. *Clinica Chimica Acta*. **412** (2011), , doi:10.1016/j.cca.2011.02.013.
5. M. Bachman, S. Uribe-Lewis, X. Yang, M. Williams, A. Murrell, S. Balasubramanian, 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nature Chemistry 2014 6:12*. **6**, 1049–1055 (2014).
6. L. Scourzic, E. Mouly, O. A. Bernard, TET proteins and the control of cytosine demethylation in cancer. *Genome Medicine*. **7** (2015), doi:10.1186/s13073-015-0134-6.
7. K. D. Rasmussen, K. Helin, Role of TET enzymes in DNA methylation, development, and cancer. *Genes and Development*. **30** (2016), , doi:10.1101/gad.276568.115.
8. W. Li, M. Liu, Distribution of 5-hydroxymethylcytosine in different human tissues. *Journal of Nucleic Acids*. **2011** (2011), doi:10.4061/2011/870726.

9. S. G. Jin, X. Wu, A. X. Li, G. P. Pfeifer, Genomic mapping of 5-hydroxymethylcytosine in the human brain. *Nucleic Acids Research*. **39** (2011), doi:10.1093/nar/gkr120.
10. O. Yildirim, R. Li, J. H. Hung, P. B. Chen, X. Dong, L. S. Ee, Z. Weng, O. J. Rando, T. G. Fazzio, Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell*. **147**, 1498–1510 (2011).
11. M. M, A. P, D. S, K. S, H. N, MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell*. **151**, 1417–1430 (2012).
12. H. MC, C. A, M. AK, E. DM, G. J, P. LG, T. A, A. P, I.-D. C, N. WG, N. GJ, D. M. AM, Y. S, Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers. *Oncotarget*. **2**, 627–637 (2011).
13. M. C. Haffner, A. Chaux, A. K. Meeker, D. M. Esopi, J. Gerber, L. G. Pellakuru, A. Toubaji, P. Argani, C. Iacobuzio-Donahue, W. G. Nelson, G. J. Netto, A. M. de Marzo, S. Yegnasubramanian, Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers. *Oncotarget*. **2** (2011), doi:10.18632/oncotarget.316.
14. C. Liu, L. Liu, X. Chen, J. Shen, J. Shan, Y. Xu, Z. Yang, L. Wu, F. Xia, P. Bie, Y. Cui, X. wu Bian, C. Qian, Decrease of 5-Hydroxymethylcytosine Is Associated with Progression of Hepatocellular Carcinoma through Downregulation of TET1. *PLoS ONE*. **8** (2013), doi:10.1371/JOURNAL.PONE.0062828.
15. A. Murata, Y. Baba, T. Ishimoto, K. Miyake, K. Kosumi, K. Harada, J. Kurashige, S. Iwagami, Y. Sakamoto, Y. Miyamoto, N. Yoshida, M. Yamamoto, S. Oda, M. Watanabe, M. Nakao, H. Baba, TET family proteins and 5-hydroxymethylcytosine in esophageal squamous cell carcinoma. *Oncotarget*. **6**, 23372–23382 (2015).
16. Z. Chen, X. Shi, L. Guo, Y. Li, M. Luo, J. He, Decreased 5-hydroxymethylcytosine levels correlate with cancer progression and poor survival: A systematic review and meta-analysis. *Oncotarget*. **8**, 1944–1952 (2017).
17. F. AF, B. GF, S. MI, U. RG, T. EG, G. MG, C. A, L. V, S. P, P. RF, B. T, T. JR, C. I, M. P, M. C, F. C, R. L, A. A, O. I, C. D. S, R.-G. P, G. A. JI, M. M, M. B, D. G, B. F, F. MF, Loss of 5hmC identifies a new type of aberrant DNA hypermethylation in glioma. *Hum Mol Genet*. **27**, 3046–3059 (2018).
18. E. M, L. S, X. H, B. A, B. A, W. F, G. R, H. D, H. S, T. A, V. D, H. A, D. Z, S. A, A. AG, B. C, J. S, V. C, S. M, S. Z, M. T, S. A, M. S, O. E, TET2 Drives 5hmc Marking of GATA6 and Epigenetically Defines Pancreatic Ductal Adenocarcinoma Transcriptional Subtypes. *Gastroenterology*. **161**, 653-668.e16 (2021).
19. J. Zhang, X. Han, C. Gao, Y. Xing, Z. Qi, R. Liu, Y. Wang, X. Zhang, Y. G. Yang, X. Li, B. Sun, X. Tian, 5-Hydroxymethylome in Circulating Cell-free DNA as A Potential Biomarker for Non-small-cell Lung Cancer. *Genomics, Proteomics and Bioinformatics*. **16**, 187–199 (2018).
20. M. Wielscher, W. Liou, W. Pulverer, C. F. Singer, C. Rappaport-Fuerhauser, D. Kandioler, G. Egger, A. Weinhäusel, Cytosine 5-hydroxymethylation of the LZTS1 gene is reduced in breast cancer. *Translational Oncology*. **6**, 715–721 (2013).
21. O. M. Wilkins, K. C. Johnson, E. A. Houseman, J. E. King, C. J. Marsit, B. C. Christensen, Genome-wide characterization of cytosine-specific 5-hydroxymethylation in normal breast tissue. *Epigenetics*. **15**, 398–418 (2020).
22. C. J, C. L, Z. Z, Z. X, L. X, L. W, S. G, G. Y, G. P, Y. Y, K. A, X. L, D. R, Z. Y, Y. X, W. J, Z. T, Y. D, H. X, S. C, Q. S, S. F, S. H, Z. W, Z. J, N. J, Z. C, S. EK, Z. X, C. BC, L. WY, H. C, W. H, Z. W, F. J, Genome-wide mapping of 5-hydroxymethylcytosines in circulating cell-free DNA as a non-invasive approach for early detection of hepatocellular carcinoma. *Gut*. **68**, 2195–2205 (2019).

23. G. Ficz, J. G. Gribben, Loss of 5-hydroxymethylcytosine in cancer: Cause or consequence? *Genomics*. **104** (2014), pp. 352–357.
24. C. A, T. JR, G. MG, U. RG, B. GF, S. M, L. V, G.-T. E, S.-O. P, P. RF, B. T, M. C, C.-T. MD, S.-S.-M. I, M. M, M. B, A. A, C. MD, F. AF, F. MF, Epigenetic downregulation of TET3 reduces genome-wide 5hmC levels and promotes glioblastoma tumorigenesis. *Int J Cancer*. **146**, 373–387 (2020).
25. K. W. Tsai, G. C. Li, C. H. Chen, M. H. Yeh, J. S. Huang, H. H. Tseng, T. Y. Fu, H. H. Liou, H. W. Pan, S. F. Huang, C. C. Chen, H. Y. Chang, L. P. Ger, H. T. Chang, Reduction of global 5-hydroxymethylcytosine is a poor prognostic factor in breast cancer patients, especially for an ER/PR-negative subtype. *Breast Cancer Research and Treatment*. **153**, 219–234 (2015).
26. W. SL, Z. X, C. M, H. C, Q. J, L. Q, Y. F, S. L, Y. X, C. X, J. J, C. M, L. Y, W. HW, L. ZY, W. X, N. Y, T. WM, J. F, Genome-wide 5-Hydroxymethylcytosine Profiling Analysis Identifies MAP7D1 as A Novel Regulator of Lymph Node Metastasis in Breast Cancer. *Genomics Proteomics Bioinformatics* (2021), doi:10.1016/J.GPB.2019.05.005.
27. D. Beck, M. ben Maamar, M. K. Skinner, Genome-wide CpG density and DNA methylation analysis method (MeDIP, RRBS, and WGBS) comparisons. *Epigenetics* (2021), doi:10.1080/15592294.2021.1924970.
28. J. Peng, B. Xia, C. Yi, Single-base resolution analysis of DNA epigenome via high-throughput sequencing. *Science China Life Sciences*. **59**, 219–226 (2016).
29. L. CG, X. Y, C. C, W. F, L. A, D. K, X. W, T. L, H. Y, Z. Q, L. CW, H. D, L. BQ, K. S, Y. Y, N. J, K. AJ, F. R, L. C, D. LM, S. RA, T. JF, K. H, H. Y, Z. LI, M. MC, K. UB, S. T, W. BA, M. GF, S. YG, Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. *Cell*. **150**, 1135–1146 (2012).
30. O. N, V. HG, M. K, K. T, Loss of 5-Hydroxymethylcytosine is an Epigenetic Hallmark of Thyroid Carcinomas with TERT Promoter Mutations. *Endocr Pathol*. **31**, 359–366 (2020).
31. A. Carella, J. R. Tejedor, M. G. García, R. G. Urdinguio, G. F. Bayón, M. Sierra, V. López, E. García-Toraño, P. Santamarina-Ojeda, R. F. Pérez, T. Bigot, C. Mangas, M. D. Corte-Torres, I. Sáenz-de-Santa-María, M. Mollejo, B. Meléndez, A. Astudillo, M. D. Chiara, A. F. Fernández, M. F. Fraga, Epigenetic downregulation of TET3 reduces genome-wide 5hmC levels and promotes glioblastoma tumorigenesis. *International Journal of Cancer*. **146** (2020), doi:10.1002/ijc.32520.
32. Z. LY, L. PL, W. TZ, Z. XC, Prognostic values of 5-hmC, 5-mC and TET2 in epithelial ovarian cancer. *Arch Gynecol Obstet*. **292**, 891–897 (2015).
33. J. An, A. Rao, M. Ko, TET family dioxygenases and DNA demethylation in stem cells and cancers. *Experimental & Molecular Medicine 2017 49:4*. **49**, e323–e323 (2017).
34. O. Taiwo, G. A. Wilson, T. Morris, S. Seisenberger, W. Reik, D. Pearce, S. Beck, L. M. Butcher, Methylome analysis using MeDIP-seq with low DNA concentrations. *Nature Protocols*. **7** (2012), doi:10.1038/nprot.2012.012.
35. D. Li, B. Zhang, X. Xing, T. Wang, Combining MeDIP-seq and MRE-seq to investigate genome-wide CpG methylation. *Methods*. **72** (2015), doi:10.1016/j.ymeth.2014.10.032.
36. C. Clark, P. Palta, C. J. Joyce, C. Scott, E. Grundberg, P. Deloukas, A. Palotie, A. J. Coffey, A Comparison of the Whole Genome Approach of MeDIP-Seq to the Targeted Approach of the Infinium HumanMethylation450 BeadChip® for Methylome Profiling. *PLoS ONE*. **7** (2012), doi:10.1371/journal.pone.0050233.
37. C. XL, N. J, K. J, D. U, W.-S. DC, C. F, E. CK, S. L, N. Y, D. Z, Z. CWT, B. A, P. J, W. J, B. AV, Z. Z, S. G, T. R, A. P, K. J, G. A, B. M, Z. W, L. S, H. C, A human tissue map of 5-hydroxymethylcytosines exhibits tissue specificity through gene and enhancer modulation. *Nat Commun*. **11** (2020), doi:10.1038/S41467-020-20001-W.

38. W. Y, C. X, Z. Y, W. Y, L. Y, X. C, Genome-wide DNA methylation and hydroxymethylation analysis reveal human menstrual blood-derived stem cells inhibit hepatocellular carcinoma growth through oncogenic pathway suppression via regulating 5-hmC in enhancer elements. *Stem Cell Res Ther*. **10** (2019), doi:10.1186/S13287-019-1243-8.
39. W. Xu, S. Wang, Q. Chen, Y. Zhang, P. Ni, X. Wu, J. Zhang, F. Qiang, A. Li, O. D. Re, S. Xu, M. Wang, R. Zhang, J. Zhou, TXNL1-XRCC1 pathway regulates cisplatin-induced cell death and contributes to resistance in human gastric cancer. *Cell Death & Disease 2014 5:2*. **5**, e1055–e1055 (2014).
40. C. DS, B. B, B. SAH, C. CJ, P.-R. I, C. BVSK, V. S, UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia*. **19**, 649–658 (2017).

Acknowledgments: Not applicable

Funding: The study was fully funded by a financial grant sanctioned by the Science and Engineering Research Board, Department of Science and Technology (DST), Government of India (EMR/2015/001319), and the student fellowship was sanctioned by the DST-INSPIRE fellowship (IF190144), Government of India.

Ethical approval and consent to participate

This study was approved by the Ethics Committees of Cancer Institute (WIA), Regional Cancer Centre, Chennai – 600036 (IEC/2016/05). The study was conducted according to the principles expressed in the Declaration of Helsinki.

Author contributions:

Deepa Ramasamy: Data curation, formal analysis, methodology, validation, visualization, writing—original draft; **Arunagiri Kuha Deva Magendhra Rao**: Formal analysis, methodology, investigation, writing—review and editing; **Meenakumari Balaiah**: Methodology; **Arvinden Vittal Rangan**: Methodology; **Shirley Sundersingh**: Resources; **Sridevi Veluswami**: Resources; **Rajkumar Thangarajan**: Writing—review and editing, supervision; **Samson Mani**: Conceptualization, funding acquisition, project administration, investigation, supervision, writing—review and editing.

Competing interests: The authors declare that they have no conflict of interest.

Consent for publication: All authors provide their consent for the publication of the manuscript

Data and materials availability: Raw sequencing data are available in Sequence Read Archive Hosted by National Centre for Biotechnology Information (NCBI) search database with accession number PRJNA769519

Figures and Tables

Figure 1

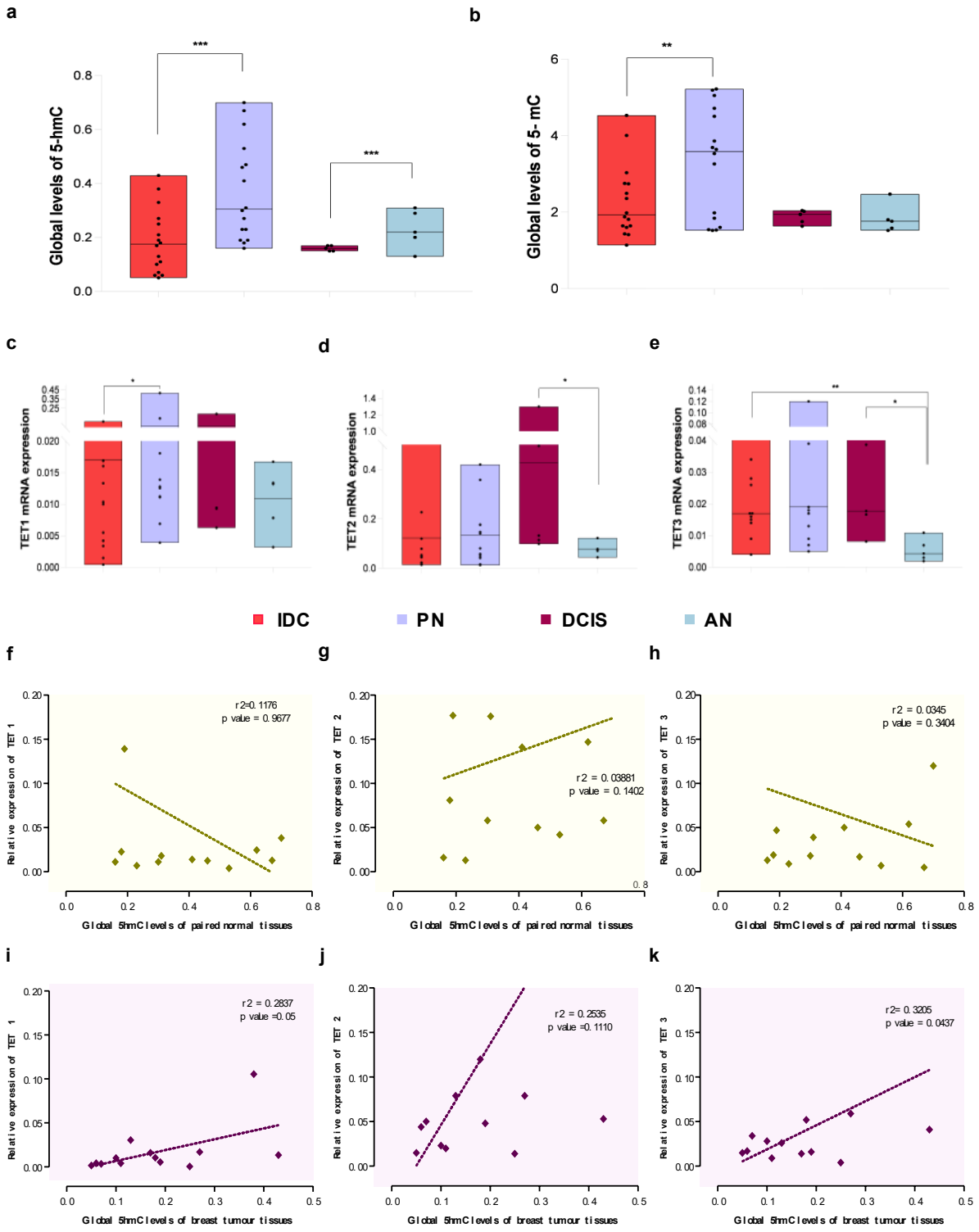


Figure 1: Global 5-mC, 5-hmC and *TET* gene expression levels in breast cancer. (a) Global levels of 5-hmC in IDC (n = 15), PN (n = 15), DCIS (n = 5), and AN (n = 5) tissues (b) Global levels of 5-mC in IDC (n = 15), PN (n = 15), DCIS (n = 5), and AN (n = 5) tissues (c) Gene expression

analysis of *TET1* among the IDC, PN, AN, and DCIS samples (d) Gene expression analysis of *TET2* among the IDC, PN, AN, and DCIS samples (e) Gene expression analysis of *TET3* among the IDC, PN, AN, and DCIS samples (f) Correlation analysis of global 5-hmC levels of PN samples with relative mRNA expression of *TET1* gene (g) Correlation analysis of global 5-hmC levels of PN samples with relative mRNA expression of *TET2* gene (h) Correlation analysis of global 5-hmC levels of PN samples with relative mRNA expression of *TET3* gene (i) Correlation analysis of global 5-hmC levels of breast tumour samples with relative mRNA expression of *TET1* gene (j) Correlation analysis of global 5-hmC levels of breast tumour samples with relative mRNA expression of *TET2* gene (k) Correlation analysis of global 5-hmC levels of breast tumour samples with relative mRNA expression of *TET3* gene. [Wilcoxon signed-rank test was applied to test the statistical significance of paired analysis, Mann–Whitney U test was used to evaluate unpaired or grouped analysis (***$p < 0.0001$; **$p < 0.001$, and *$p < 0.05$) (Abbreviations: PN- paired normal; IDC- invasive ductal carcinoma; AN- apparent normal breast tissues; DCIS- ductal carcinoma in situ)].

Figure 2

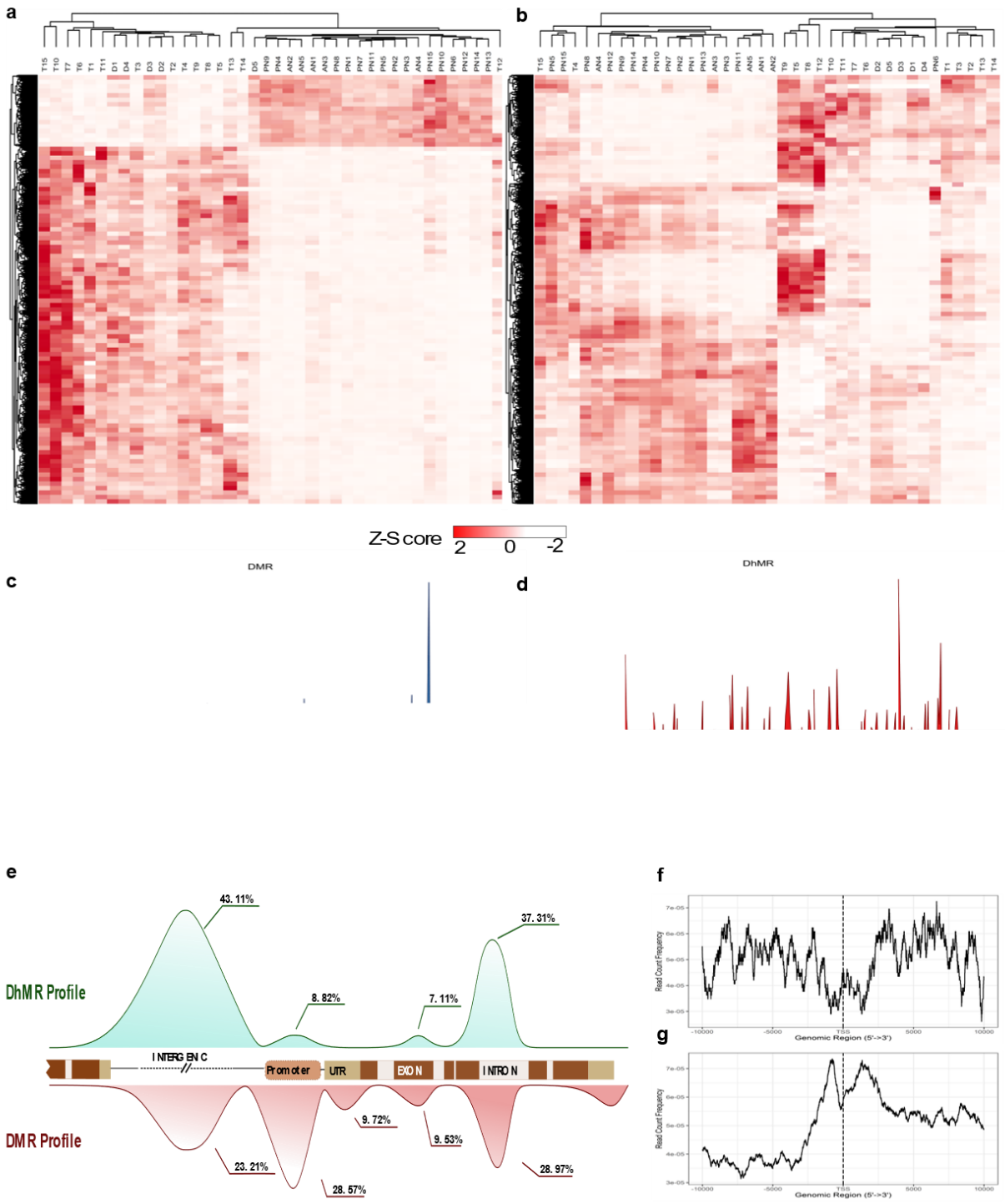


Figure 2: Genome-wide distribution of 5-mC and 5-hmC in breast cancer. (a) Heatmap representing DMRs in breast cancer (b) Heatmap showing DhMRs in breast cancer (Z-score ranges from –2 (white) to +2 (red)) (c) Chromosomal distribution of differentially methylated regions (DMRs) in breast cancer (d) Chromosomal distribution of differentially hydroxymethylated regions (DhMRs) in breast cancer (e) Genomic features of DMRs and DhMRs in breast cancer (f) Relative

peak count frequency of DhMR from Transcription Start Sites (TSSs) (g) Relative peak count frequency of DMR from TSSs

Figure3

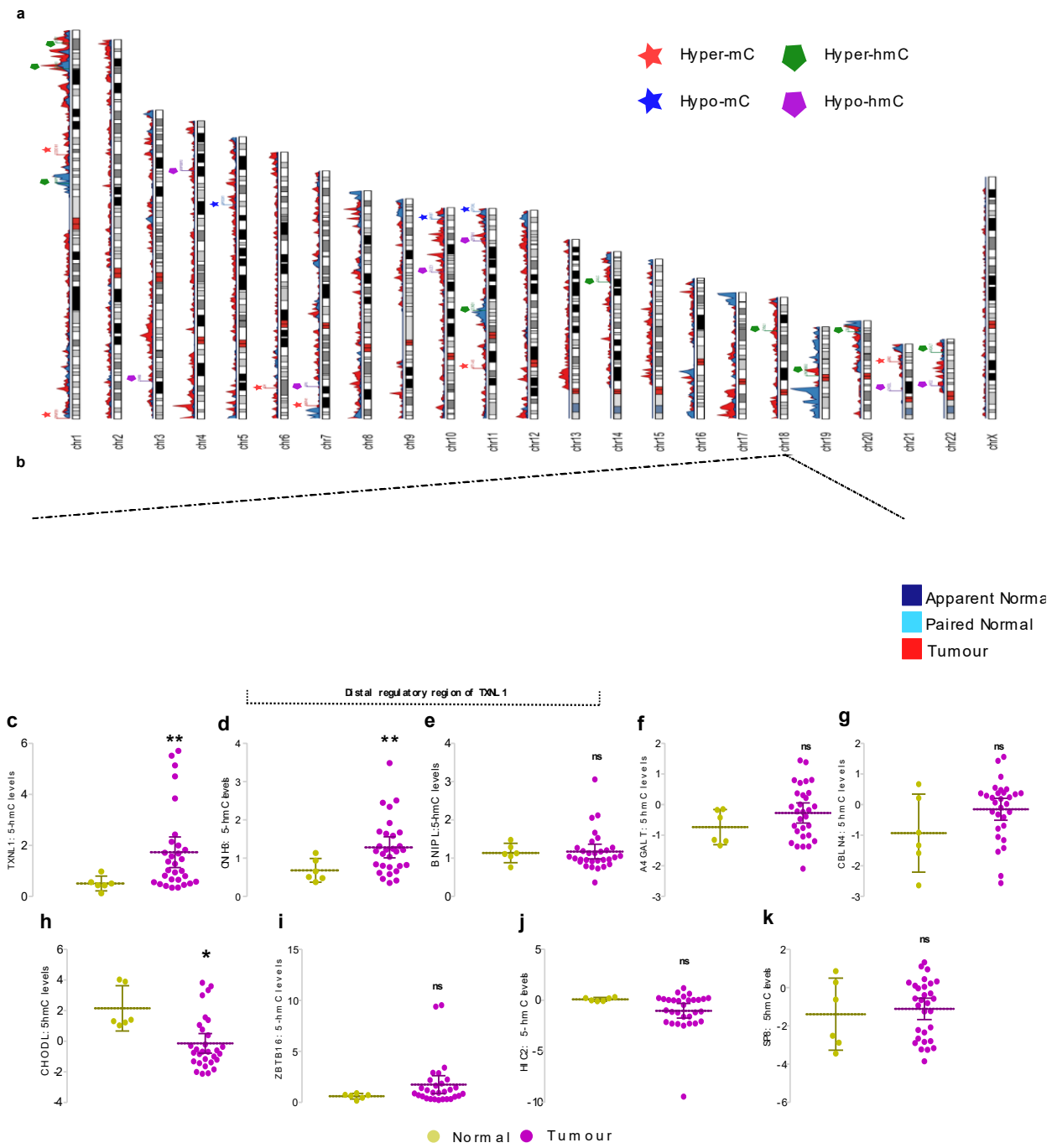


Figure 3: Validation of 5-hmC specific loci in breast tumour and normal samples. (a) Ideogram representing DMRs (blue) and DhMRs (red) across all chromosomes and the candidate loci of hyper-mC, hypo-mC, hyper-hmC, and hypo-hmC groups (b) Integrative genome viewer representing the gain of 5-hmC in the distal regulatory region of TXNL1 in tumour, paired normal and apparent normal samples (c) Validation of 5-hmC levels of TXNL1 (d) Validation of 5-hmC levels of CNIH3 (e)Validation of 5-hmC levels of BNIPL (f) Validation of 5-hmC levels of A4GALT (g) Validation of 5-hmC levels of CBLN4 (h) Validation of 5-hmC levels of CHODL (i)

Validation of 5-hmC levels of ZBTB16 (j)Validation of 5-hmC levels of HIC2 (k) Validation of 5-hmC levels of SP8. The *p*-value of <0.05 was considered statistically significant (***$p < 0.0001$; **$p < 0.001$, and *$p < 0.05$).

Figure 4

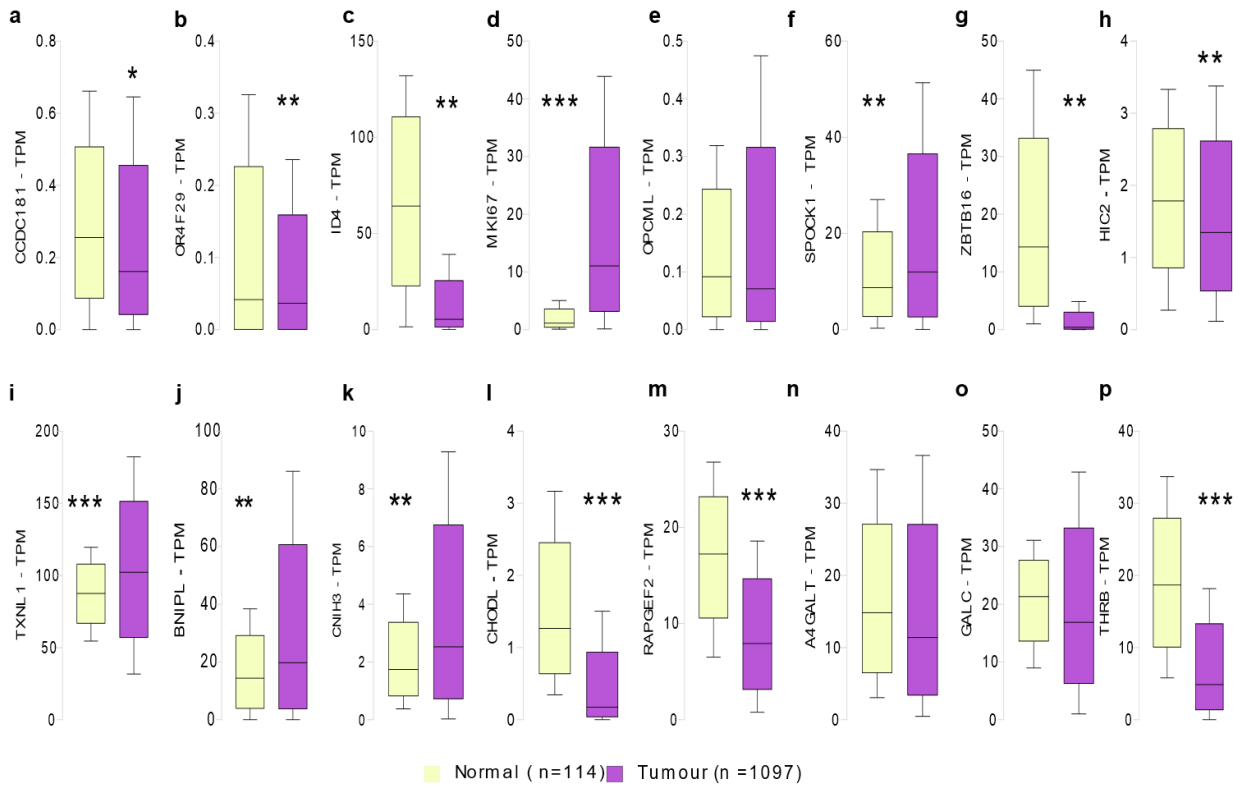


Figure 4: Gene expression analysis of 5-mC and 5-hmC specific loci in breast tumour and normal samples (a-f) Gene expression analysis of hyper- and hypo-mC candidates (g-p); Gene expression analysis of hyper- and hypo-hmC candidates using UALCAN webtool (***$p < 0.0001$; **$p < 0.001$, and *$p < 0.05$).

Table 1: Hyper-mC/hmC and hypo-mC/hmC candidates for the validation and gene expression analysis.

Hyper-hmC regions				
Gene symbol	Entrez Gene Name	Annotated regions	Log2 Fold change	P-value
GALC	Galactosylceramidase	Distal intergenic	4.915287863	0.0073
TXNL1	Thioredoxin like -1	Distal intergenic	1.719914693	0.00217
BNIPL	BCL2 interacting protein-like	Promoter	1.563365748	0.00108
CNIH3	Cornichon family AMPA receptor auxiliary protein 3	Intron	1.541702124	0.00187
A4GALT	alpha 1,4-galactosyltransferase (P blood group)	Promoter	1.520339559	0.00848
CBLN4	cerebellin 4 precursor	Distal intergenic	1.499534815	0.00075
Hypo-hmC regions				
Gene symbol	Entrez Gene Name	Annotated regions	Log2 Fold change	P-value
RAPGEF2	Rap guanine nucleotide exchange factor 2 [Homo sapiens (human)]	Distal intergenic	-1.569277122	0.00587
CHODL	Chondrolectin	Intron	-1.570929568	0.00015
ZBTB16	Zinc finger and BTB domain containing 16	Intron	-1.270011867	0.00753
HIC2	HIC ZBTB transcriptional repressor 2	Intron	-1.50172022	0.00355
SP8	Sp8 transcription factor	Distal intergenic	-1.311713468	5.3E-05
Hypermethylated regions				
Gene symbol	Entrez Gene Name	Annotated regions	Log2 Fold change	P-value
CCDC181	coiled-coil domain containing 181	Promoter	3.864111418	3.1E-14
SIM2	SIM bHLH transcription factor 2	Promoter	3.049891817	2.6E-31
OR4F29	olfactory receptor family 4 subfamily F member 29	Distal intergenic	3.000663491	0.00014
NXPH1	neurexophilin 1	Intron	2.809643256	7.1E-36
WT1-AS	WT1 antisense RNA	Promoter	2.791763836	1E-25
ID4	inhibitor of DNA binding 4, HLH protein	Promoter	2.788606821	3.1E-12
Hypo methylated regions				
Gene symbol	Entrez Gene Name	Annotated regions	Log2 Fold change	P-value
MKI67	marker of proliferation Ki-67	Distal intergenic	-1.265945306	5.9E-11
OPCML	opioid binding protein/cell adhesion molecule like	Intron	-1.626970741	1.9E-06
SPOCK1	SPARC (osteonectin), cwcv and kazal like domains proteoglycan 1	Intron	-1.412627239	3.4E-06

Supplementary Materials for

Locus specific enrichment analysis of 5-hydroxymethylcytosine reveales novel genes associated with breast carcinogenesis

Deepa Ramasamy[1], Arunagiri Kuha Deva Magendhra Rao[1], Meenakumari Balaiah[1], Arvinden Vittal Rangan[1], Shirley Sundersingh[2], Sridevi Veluswami[3], Rajkumar Thangarajan[1], Samson Mani[1*]

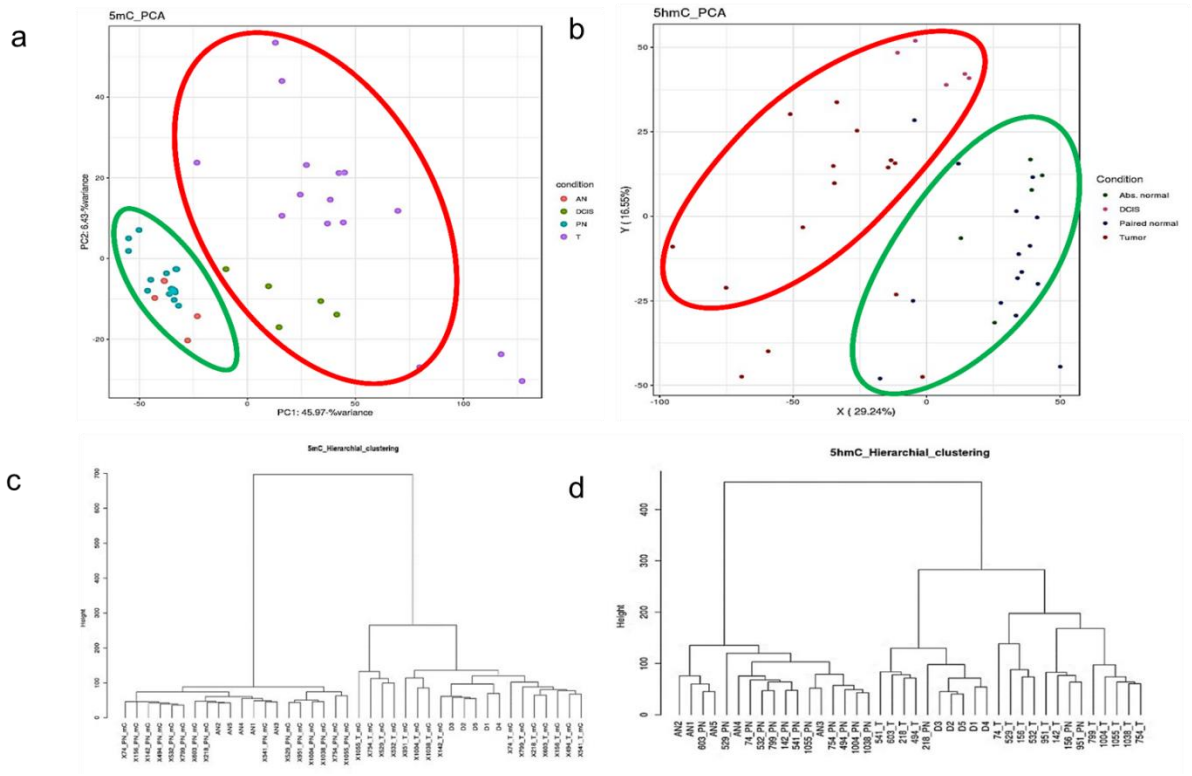
*Corresponding author. Email: samsonn.m@gmail.com

This PDF file includes:

Supplementary Figures SF1 to SF5
Supplementary Table ST1
Supplementary files Sf1 to Sf4

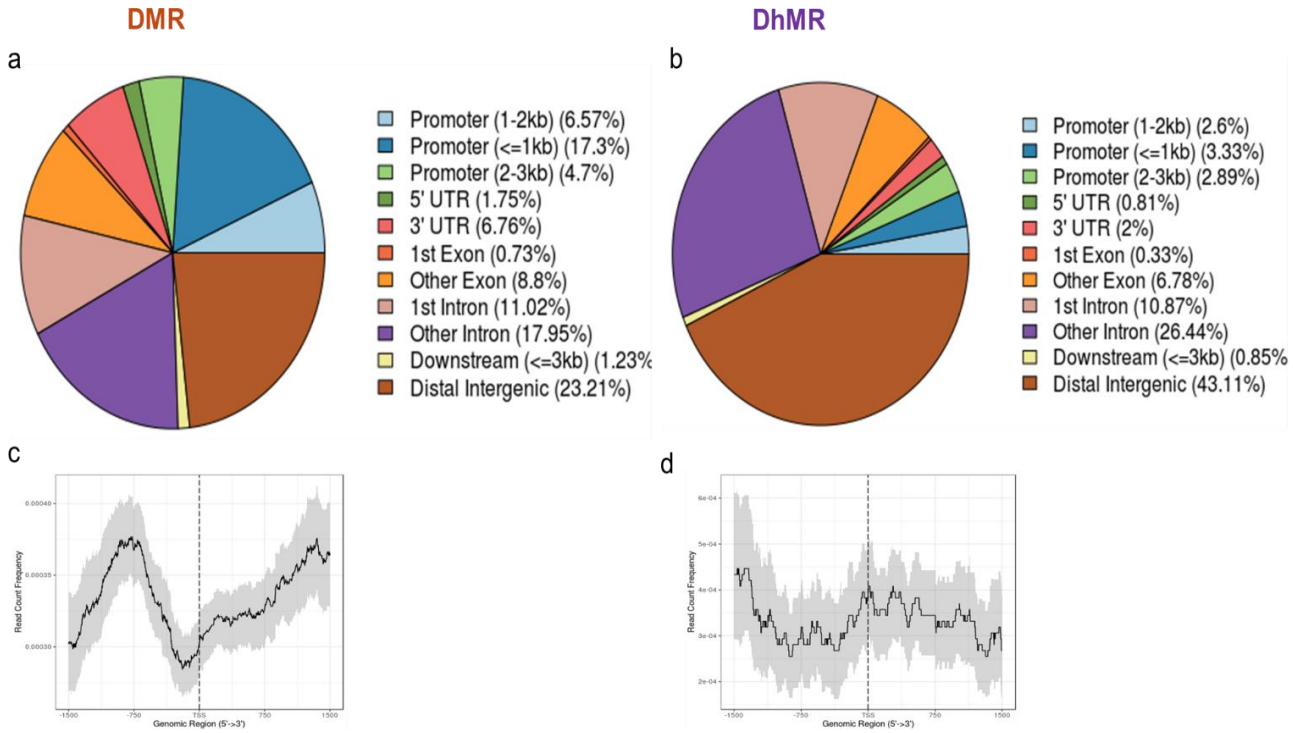
Supplementary Figures:

SF 1. PCA and HCA of 5-mC and 5-hmC enriched regions of breast tumour and normal tissues



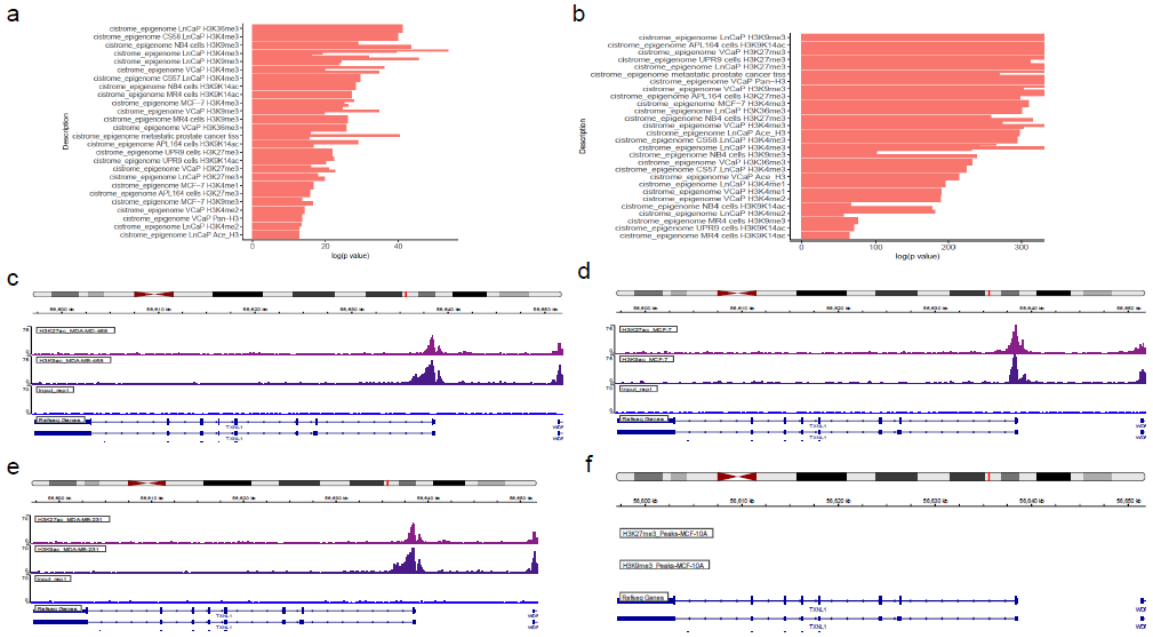
SF 1: (a) Principal component analysis of 5-mC (b) Principal component analysis of 5-hmC (c) Hierarchical clustering analysis of 5-mC in PN, IDC, DCIS, and AN samples (d) Hierarchical clustering analysis of 5-hmC in PN, IDC, AN and DCIS samples.

SF 2: Genome-wide distribution of DMR and DhMR in breast cancer



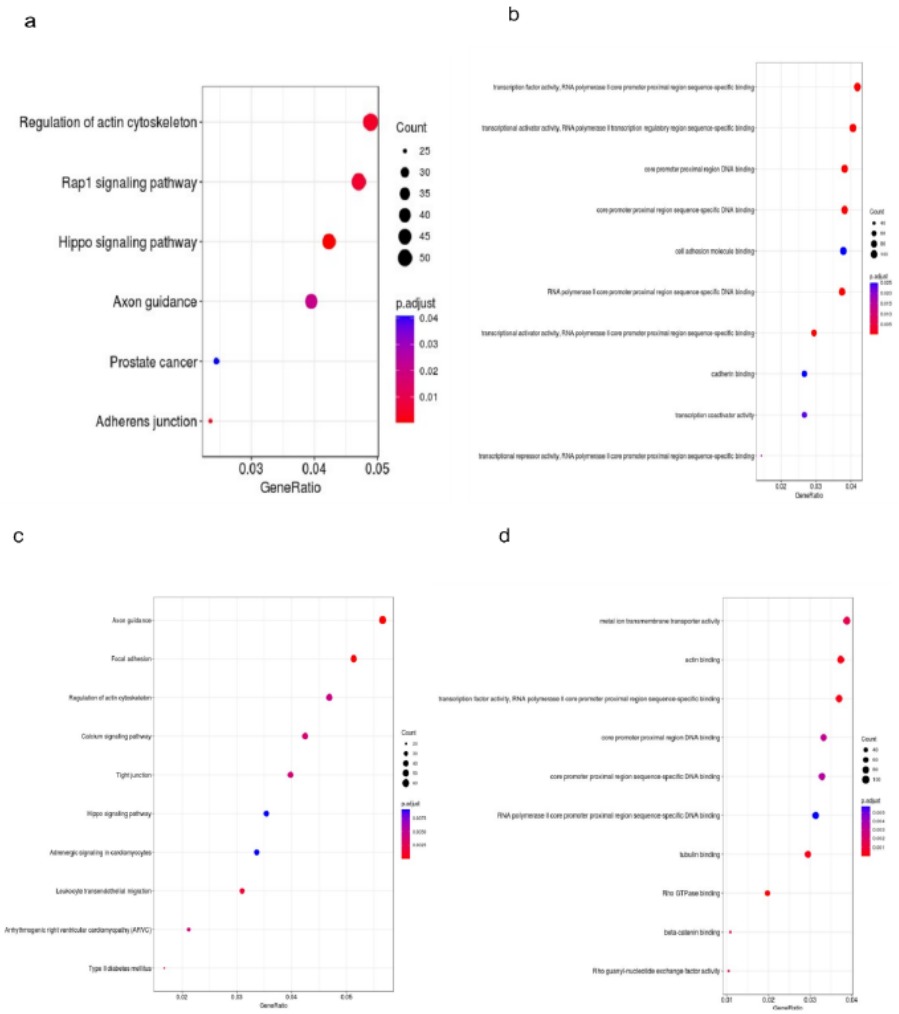
SF 2: (a) Genome-wide distribution of DMRs (b) Genome-wide distribution of DhMRs (c) DMRs peak intensity of TFBS in the region of 1500 bp upstream and downstream of TSSs (d) DhMRs peak intensity of TFBS in the region of 1500 bp upstream and downstream of TSSs.

SF 3. Locus Overlap analysis and ChIP seq analysis of DhMR



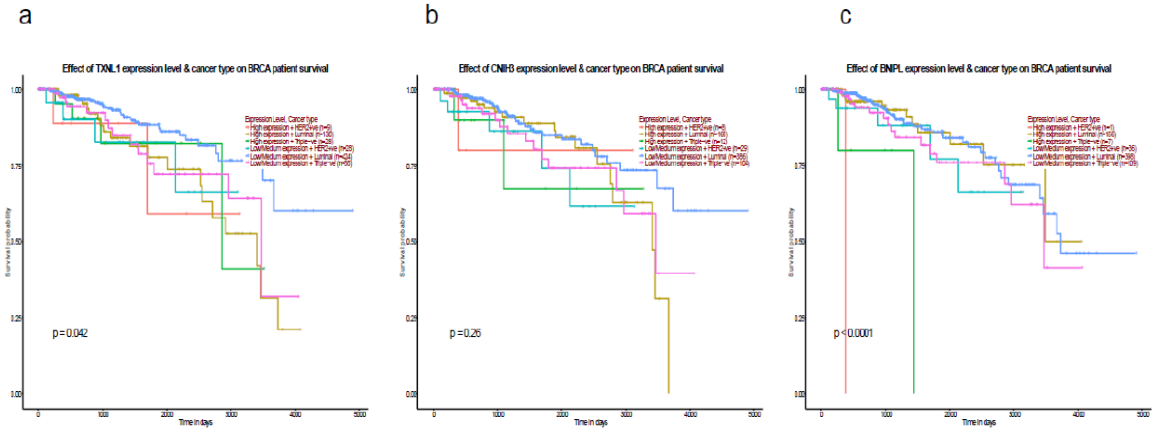
SF 3: (a-b) Locus overlap analysis of DhMRs using LOLA webtool (c-f) Chip-seq analysis using UALCAN webtool.

SF 4. KEGG pathway enrichment and GO term analysis of DMR and DhMR of breast cancer



SF 4: (a) KEGG pathway enrichment map of DMRs (a) GO term pathway enrichment map of DMRs (c) KEGG pathway enrichment map of DhMRs (d) GO term pathway enrichment map of DhMRs.

SF 5. Survival analysis of hyper-hmC candidates



SF 5: (a)Survival analysis of TXNL1 (b) Survival analysis of CNIH3 (c) Survival analysis of BNIPL.

Supplementary Table:

ST. 1 (1a) Genomic features of 5-mC (1b) Genomic features of 5-hmC (paired normal (PN), invasive ductal carcinoma (IDC), ductal carcinoma in situ (DCIS) and apparent normal (AN) samples).

Supplementary Files:

Sf 1: (1a) Significant differentially hydroxymethylated regions (DhMRs) of breast tumor and normal samples (1b) Significant differentially methylated regions (DMRs) of breast tumor and normal samples.

Sf 2: Significant hyper- and hypo-5-mC and 5-hmC regions (2a) Hyper-hmCgene list (2b) hypo-hmC gene list (2c) hyper-mC gene list (2d) hypo-mC gene list.

Sf 3: Gene specific analysis using gprofiler (3a) Hyper-hmC genes dataset (3b) Hypo-hmC genes dataset (3c) Hyper-mC genes dataset (3d) Hypo-mC genes dataset.

Sf 4: qPCR primers used in the study.