

RESEARCH

Open Access



# Genome-wide characterization of dynamic DNA 5-hydroxymethylcytosine and TET2-related DNA demethylation during breast tumorigenesis

Shuang-Ling Wu<sup>1,2†</sup>, Lin Yang<sup>2†</sup>, Changcai Huang<sup>3†</sup>, Qing Li<sup>2</sup>, Chunhui Ma<sup>2</sup>, Fang Yuan<sup>4</sup>, Yinglin Zhou<sup>5</sup>, Xiaoyue Wang<sup>6</sup>, Wei-Min Tong<sup>2,7</sup>, Yamei Niu<sup>2,7\*</sup> and Feng Jin<sup>1\*</sup>

## Abstract

**Background** Breast tumorigenesis is a complex and multistep process accompanied by both genetic and epigenetic dysregulation. In contrast to the extensive studies on DNA epigenetic modifications 5-hydroxymethylcytosine (5hmC) and 5-methylcytosine (5mC) in malignant breast tumors, their roles in the early phases of breast tumorigenesis remain ambiguous.

**Results** DNA 5hmC and 5mC exhibited a consistent and significant decrease from usual ductal hyperplasia to atypical ductal hyperplasia and subsequently to ductal carcinoma in situ (DCIS). However, 5hmC showed a modest increase in invasive ductal breast cancer compared to DCIS. Genomic analyses showed that the changes in 5hmC and 5mC levels occurred around the transcription start sites (TSSs), and the modification levels were strongly correlated with gene expression levels. Meanwhile, it was found that differentially hydroxymethylated regions (DhMRs) and differentially methylated regions (DMRs) were overlapped in the early phases and accompanied by the enrichment of active histone marks. In addition, TET2-related DNA demethylation was found to be involved in breast tumorigenesis, and four transcription factor binding sites (TFs: ESR1, FOXA1, GATA3, FOS) were enriched in TET2-related DhMRs/DMRs. Intriguingly, we also identified a certain number of common DhMRs between tumor samples and cell-free DNA (cfDNA).

**Conclusions** Our study reveals that dynamic changes in DNA 5hmC and 5mC play a vital role in propelling breast tumorigenesis. Both TFs and active histone marks are involved in TET2-related DNA demethylation. Concurrent changes in 5hmC signals in primary breast tumors and cfDNA may play a promising role in breast cancer screening.

**Keywords** Breast tumorigenesis, DNA 5-hydroxymethylcytosine, 5-methylcytosine, DNA methylation, DNA demethylation, TET2, cfDNA

<sup>†</sup>Shuang-Ling Wu, Lin Yang and Changcai Huang contributed equally to this work.

\*Correspondence:

Yamei Niu

niuym@ibms.pumc.edu.cn

Feng Jin

jinfeng@cmu.edu.cn

Full list of author information is available at the end of the article



## Background

Breast cancer is the most prevalent cancer worldwide [1, 2]. Epidemiological and pathological studies have shown that breast tumorigenesis is a stepwise process, staging from usual ductal hyperplasia (UDH) to atypical ductal/lobular hyperplasia (ADH/ALH), ductal/lobular carcinoma in situ (DCIS/LCIS) and invasive ductal breast cancer (IDC) [3–6]. In general, patients diagnosed with ADH/ALH or LCIS have a 4–10 fold increased risk for IDC [7–9]. Although advancements in imaging techniques have improved the diagnosis of breast tumors, there is still a need for efficient biomarkers to distinguish the tumors with elevated risk of malignant transformation [10]. Therefore, identifying the molecular determinants underlying the progression of breast precancerous lesions holds immense promise for the development of early diagnostic and predictive biomarkers.

The maintenance of dynamic equilibrium between DNA methylation and demethylation is crucial in numerous physiological processes [11, 12]. In contrast, extensive studies have demonstrated that cancer cells exhibit aberrant DNA methylation and 5-hydroxymethylation patterns [13–17]. To date, both 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) signatures have emerged as pivotal epigenetic regulators in cancer initiation and progression. Intriguingly, in addition to primary tumors, mounting evidence suggests that 5hmC in cell-free DNA (cfDNA) is tissue-specific and holds promise as a biomarker for cancer detection [18–21].

Previously, epigenetic modulations, encompassing DNA methylation, have been recognized as pivotal contributors to breast tumorigenesis [22]. Recent investigations have expanded beyond the global changes of DNA 5hmC in breast cancer to explore the genomic landscape of 5hmC in breast tissues and breast cancers (DCIS/IDC) [23, 24]. Notably, breast cancer and precursor lesions exhibit a significant reduction in the abundance of 5hmC compared to normal tissues [25–28]. Moreover, genomic 5hmC exhibits dynamic changes in various stages of lymph node metastasis in breast cancer [29]. However, although the involvement of DNA 5hmC and 5mC in breast tumors has been widely demonstrated [30], it remains unclear whether dysregulated DNA 5hmC and 5mC are involved in the progression of breast precancerous lesions. Furthermore, the majority of methylated loci identified in previous reports are primarily based on methylation-specific PCR [31–33] and Human Methylation450 microarray [34–37]. Thus, our understanding of the genome-wide characteristics of DNA 5mC and 5hmC in breast precancerous lesions and the role of active DNA demethylation in breast tumorigenesis is still lagging behind.

In this study, we comprehensively analyzed the dynamic landscapes of DNA 5hmC, spanning from early precancerous lesions to malignant tumors, and investigated the intricate crosstalk between 5hmC and 5mC in breast precancerous lesions. Furthermore, through in silico analyses, we unveiled the potential role of TET2 in collaboration with transcription factors (TFs) in influencing dynamic DNA demethylation and propelling breast tumorigenesis. Finally, we identified hydroxymethylated regions in cfDNA that hold promise for application in breast cancer screening.

## Methods

### Human breast tumor samples' collection

Samples of precancerous lesions including usual ductal hyperplasia (UDH), atypical ductal hyperplasia (ADH), and ductal carcinoma in situ (DCIS) as well as invasive ductal breast cancer (IDC) were obtained from the patients who received surgery in the First Affiliated Hospital of China Medical University (Table 1). Detailed information of clinical samples used for hMeDIP-seq/MeDIP-seq/RNA-seq in this study were listed in Table 2.

### Immunohistochemical staining analysis

4- $\mu$ m paraffin-embedded sections were employed for immunohistochemical staining (IHC). After de-paraffinization and antigen retrieval, tissue sections were incubated with primary antibodies at 4 °C overnight and secondary antibodies at 37 °C for 2 h. After that, all the slides were counterstained with hematoxylin. The antibodies used in this study include anti-5hmC (Active motif, 39769, 1:2000), and ImmPRESSTM horse anti-Rabbit IgG (Vector, MP-7401).

All the images were acquired using the TissueFAXS cell analysis system (TissueGnostics, Austria). In each slide, we randomly selected more than three regions (> 1 mm<sup>2</sup>) that were enriched with ducts for quantitative analysis. According to the degree of cell aggregation and the

**Table 1** Samples information and number of clinical samples used in this study

Sample numbers	IHC <sup>a</sup>	LC-MS/MS <sup>b</sup>	hMeDIP-seq <sup>b</sup>	MeDIP-seq <sup>b</sup>	RNA-seq <sup>b</sup>
<b>Sample type</b>					
UDH	104	0	3	3	2
ADH	68	0	3	3	2
DCIS	48	9	3	3	2
IDC	81	39	4	0	0

<sup>a</sup> Formalin-fixed paraffin-embedded sample

<sup>b</sup> Fresh frozen sample

**Table 2** Information of patient characteristics that used for genomic sequencing

Sample	Age	Tumor size (cm)	Pathological diagnosis	Molecular subtype	TNM
UDH1	52	2.1	Usual ductal hyperplasia	–	–
UDH2	39	2.0	Usual ductal hyperplasia	–	–
UDH3	44	3.2	Usual ductal hyperplasia	–	–
ADH1	51	1.3	Atypical ductal hyperplasia	–	–
ADH2	53	1.8	Atypical ductal hyperplasia	–	–
ADH3	45	2.0	Atypical ductal hyperplasia	–	–
DCIS1	46	4.8	Intermediate grade ductal carcinoma in situ	ER(50%) PR(50%+) C-erB-2(1+) Ki67(5%)	TisNOMO
DCIS2	64	1.8	Intermediate grade ductal carcinoma in situ	ER(90%) PR(90%+) C-erB-2(1+) Ki67(< 10%)	TisNOMO
DCIS3	53	3.5	Intermediate grade ductal carcinoma in situ	ER(90%) PR(70%+) C-erB-2(1+) Ki67(10%)	TisNOMO
IDC1	63	1.8	Invasive ductal breast cancer	ER(80%) PR(70%+) C-erB-2(2+) Ki67(10%) Fish(no amplification)	T1NOMO
IDC2	62	1.9	Invasive ductal breast cancer	ER(90%) PR(70%+) C-erB-2(1+) Ki67(40%)	T1NOMO
IDC3	52	1.5	Invasive ductal breast cancer	ER(75%) PR(75%+) C-erB-2(1+) Ki67(20%)	T1NOMO
IDC4	61	1.5	Invasive ductal breast cancer	ER(35%) PR(–) C-erB-2(–) Ki67(25%)	T1NOMO

length/width ratio of each cell, the nuclei of luminal cells and cancer cell in each duct were identified and marked. DAB staining intensity of each cell was measured by using HistoQuest software. The score value in each region was calculated as the average staining intensity in all the selected cells. The mean score value of all the selected cells was calculated and recorded as relative level of DNA 5hmC in each sample.

#### Histological identification and manual macro-dissection

All the clinical tumor samples were collected after pathological diagnosis. Fresh frozen samples were sliced to 8- $\mu$ m tissue sections firstly and then subjected to hematoxylin and eosin staining (H & E). Histological identification was performed for each slide under two pathologists' screen separately. Ultimately, tumor lesions diagnosed with definite UDH, ADH, DCIS and IDC without too much infiltrating lymphocytes (roughly < 10%) were subjected for hMeDIP-seq/MeDIP-seq. For macro-dissection, samples with 8- $\mu$ m slides were used for pathological identification, while continuously adjacent 30- $\mu$ m fresh frozen tissue sections were subjected for H & E staining and manual macro-dissection using stereomicroscope [29]. Macro-dissected ducts/cells were collected and then used for DNA and RNA extraction. Additionally, regarding the samples used for RNA extraction, all the reagents and consumables were pretreated with RNase removal, and all the procedures were performed at 4 °C.

#### LC-MS/MS analysis

The 5hmC and 5mC content in cells were quantified by the LC-MS/MS as described previously [38]. Briefly, genomic DNA obtained from 48 samples were firstly

digested into single nucleosides with DNA Degradase Plus (Zymo Research, E2021). Subsequently, the nucleosides and labeled products were analyzed with Thermo Scientific Dionex Ultimate 3000 HPLC coupled with a Triple Quad™ 5500 mass spectrometer with an ESI source.

#### DNA extraction and hMeDIP-seq/MeDIP-seq

A total of 3.0  $\mu$ g of intact genomic DNA per sample was used for Hydroxymethylated/Methylated DNA immunoprecipitation sequencing (hMeDIP/ MeDIP-seq). 1.5  $\mu$ g of genomic DNA mixed with 5hmC or 5mC spike-in DNA control (5hmC: ZYMO Research, D5405-3, 1:20000; 5mC: Wise gene, S001, 1:200) were fragmented (100–250 bp) and ligated with Illumina barcode adapter. In this study, input DNA was used as a control to determine the enrichment ratio of 5hmC/5mC-modified DNA. Immunoprecipitation reaction was performed by mixing DNA with 5hmC/5mC antibody (5hmC: Active motif, 39769; 5mC: Abcam, ab10805) and protein A/G beads for 2 h. The immunoprecipitated 5hmC/5mC-containing DNA fragments were purified using QIAGEN Mini Elute PCR purification kit (QIAGEN, 28004). All the immunoprecipitated products and input DNA were subjected to amplification, size selection (275–475 bp), purification (QIAGEN, 28704) and quality control test. All the samples (UDH: 3 cases, ADH: 3 cases, DCIS: 3 cases) were subjected to next-generation sequencing on Illumina Hiseq X-Ten system. The hMeDIP-seq data of early-stage invasive ductal breast cancer were described in our previous report (GSA: CRA001593) [29].

### Reads mapping

First, raw reads were processed with Trimmomatic (Version 0.33) to remove sequencing adaptors and low-quality bases by using default parameters [39]. The clean reads were mapped to hg19 genome by Bowtie2 (Version 2.3.2) with default parameters [40]. Then Samtools (Version 1.9) [41] were used to remove duplicated and unpaired reads.

### Peak calling and annotation

Whole-genome scanning of hydroxymethylated/methylated region (hMR/MR) was conducted by using MACS2 (Version 2.1.1) [42]. Differentially hydroxymethylated/methylated regions (DhMR/DMR) were identified using Diffbind (Version 3.8) package in R with the parameters  $P$  value  $< 0.05$  and  $|\text{Log}_2(\text{foldchange})| > 1$  [43, 44]. To determine hydroxymethylated/methylated genes (hMG/MG) and differentially hydroxymethylated/methylated genes (DhMG/DMG), hMRs, MRs, DhMRs and DMRs were annotated to genomic regions and corresponding genes with ChIPseeker R package (Version 1.36.0) [45].

### Continually hyper/hypo-methylated peaks during tumorigenesis

We first implemented differential hydroxymethylation analysis upon samples between adjacent stages. Regions showing continual 5hmC accumulation ( $P$  value  $< 0.05$  and  $|\text{Log}_2(\text{foldchange})| > 1$ ) were identified as continually hyper-hydroxymethylated regions, with the same criteria being applied for hypo-hydroxymethylation. For the sake of visualization, we averaged the signal values in biological replicates of each stage and normalized across all phases with the highest value as 1.

### KEGG pathway enrichment and GESA analyses

KEGG pathway enrichment analysis of the selected genes was carried out using the DAVID tool [39]. The cut-off value of FDR value for the significantly enriched pathways was 0.05. Meanwhile, Breast cancer related gene sets were selected from MSigDB and were analyzed through Gene set enrichment analyses (GSEA), the criteria were  $\text{NES} > 1$ ,  $Q$  value  $< 0.05$ .

### RNA extraction and RNA-seq

Total RNA was extracted from paired macro-dissected samples of hMeDIP-seq/MeDIP-seq using TriReagent (Sigma, T92424). Truseq RNA library preparation was used for UDH1, UDH2, DCIS1, and DCIS2 samples; Ribo-Minus RNA library preparation was used for ADH2 and ADH3 samples. Subsequently, RNA

sequencing on Illumina HiSeq X-Ten system were performed as described previously [40].

### Analysis of RNA-seq data

For the comparison of RNA abundance between adjacent stages during breast tumorigenesis, we downloaded RNA-seq data from the GEO database (GSE47462) [42]. Normalized read counts were input to the Limma package (Version 3.56.2) [44] for differential expression analysis.

To further explore the effect of DNA epigenetic modifications on transcriptional regulation, the raw data of RNA-seq with part of our own paired samples were firstly subjected to QC analyses with the FastQC tool, and then mapped to hg19 genome by bowtie2 (Version 2.3.2) with default parameters. The expression level of each RNA was quantified with Fragments Per Kilobase of transcript per Million mapped reads (FPKM).

### Correlation analysis between RNA expression and DNA modification level

To identify the correlation between RNA expression and DNA methylation/ hydroxymethylation levels, we divided RNAs into three equally-sized groups according to the tri-sectional quantiles of their expression levels in every single sample, and then made the average plots of all 5hmC/5mC peaks along the DNAs encoding of RNAs in each group.

### ChIP-seq analysis of TET2, histone marks, and TFs

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) data of TET2 were obtained from GEO database (GSE153251) [46], and genomic coordinates were transferred from hg38 to hg19 version with UCSC liftover. To obtain the complete TET2-binding genomic regions, we integrated ChIP-seq data from three experiments conducting ChIP-seq analyses on MCF7 cells without any perturbation. DhMRs/DMRs that overlapped with TET2-binding regions were identified as TET2-binding DhMR/DMRs. ChIP-seq data of TFs (ESR1, GATA3, FOXA1, FOS, FOSL2, FOXM1, JUNB) and histone marks (H3K27ac, H3K27me3, H3K36me3, H3K4me1, H3K4me3, H3K9me3) were downloaded from ENCODE [47] in bigwig format, and then analyzed with Deeptools (Version 3.5.1).

### Motif analysis

The motif of TET2-binding DhMRs/DMRs in each phase was identified with Homer (Version 4.11.1) [48] with default parameters. For visualizing the binding motifs of TET2-binding regions overlapped with TFs, the  $P$  values from motif enrichment analysis were used in the heatmap.

## Visualization

The average read counts per million distribution of the gene were displayed from 3 kb upstream of transcription start sites (TSSs) to 3 kb downstream of transcription ending sites (TESs) using Deeptools [49]. Clustering and heatmap plotting of hMR was conducted by Pheatmap package (Version 1.0.10) in R. KEGG enrichment plots, volcano plots, and boxplots in this paper were all made with Ggplot2 (Version 3.1.0) [50] R package.

Regarding the average plots showing the enrichment of different proteins (histone marks and transcriptional factors) flanking specific genomic regions (like DhMRs/DMRs), we used Multibigwig summary function in Deeptools to first summarize the enrichment of binding signals of the corresponding protein in equally binned regions and then plot the enrichment value in each bin.

## DhMRs analysis of cfDNA

Short-gun sequencing of 5-hydroxymethylated cfDNA from blood samples of both healthy controls and breast cancer patients were downloaded from GEO database (GSE81314) [50]. The DhMRs in cfDNA were identified under the same criteria as that in primary tumor samples. The overlap of DhMRs in cfDNA and primary tumors were quantified by the FindOverlaps function in the GenomicRanges R package [51].

## Statistical analysis

Unpaired student's *t*-test analysis was applied for statistical analysis in Fig. 1c, 1e, 2b, 3a, g, h, 5a, b, S1c, S2c, S4c, S5a, S5b. Paired student's *t*-test analysis was applied for statistical analysis in Fig. 1d. All the data were presented as mean  $\pm$  SEM,  $P < 0.05$  was set as statistically significant.

## Results

### Dynamic changes of DNA 5-hydroxymethylcytosine across different stages of breast precancerous lesions

To investigate the level of DNA 5hmC in breast tumors at different stages, we conducted immunohistochemical staining analysis of 5hmC using samples from patients with UDH, ADH, DCIS, and early-stage IDC (T1N0M0) (Fig. 1a). As breast precursor lesions primarily arise from the luminal epithelial compartment of Terminal Duct

Lobular Units (TDLUs), we focused on breast ductal epithelial cells to assess the staining intensity of 5hmC (Fig. 1b). A consistent and significant decrease in 5hmC levels was observed as lesions progressed from UDH to ADH and subsequently to DCIS. However, there was a modest increase in 5hmC levels from DCIS to IDC. It's worth noting that this trend held across various patient samples (Fig. 1c). Interestingly, when comparing samples of UDH, ADH, and DCIS within the same histological sections, we found a marked reduction of 5hmC abundance in the tumor cells of advanced stage (Fig. 1d). To validate these findings, we performed a quantitative analysis of 5hmC level using liquid chromatography-mass spectrometry (LC-MS/MS), and confirmed an increasing trend in 5hmC levels from DCIS to IDC (Fig. 1e). Taken together, the dynamic changes of 5hmC in different stages of breast tumors imply that DNA 5hmC is involved in breast tumorigenesis and may have a crucial role in the transformation of precancerous lesions to invasive cancers.

## Genome-wide reprogramming of DNA

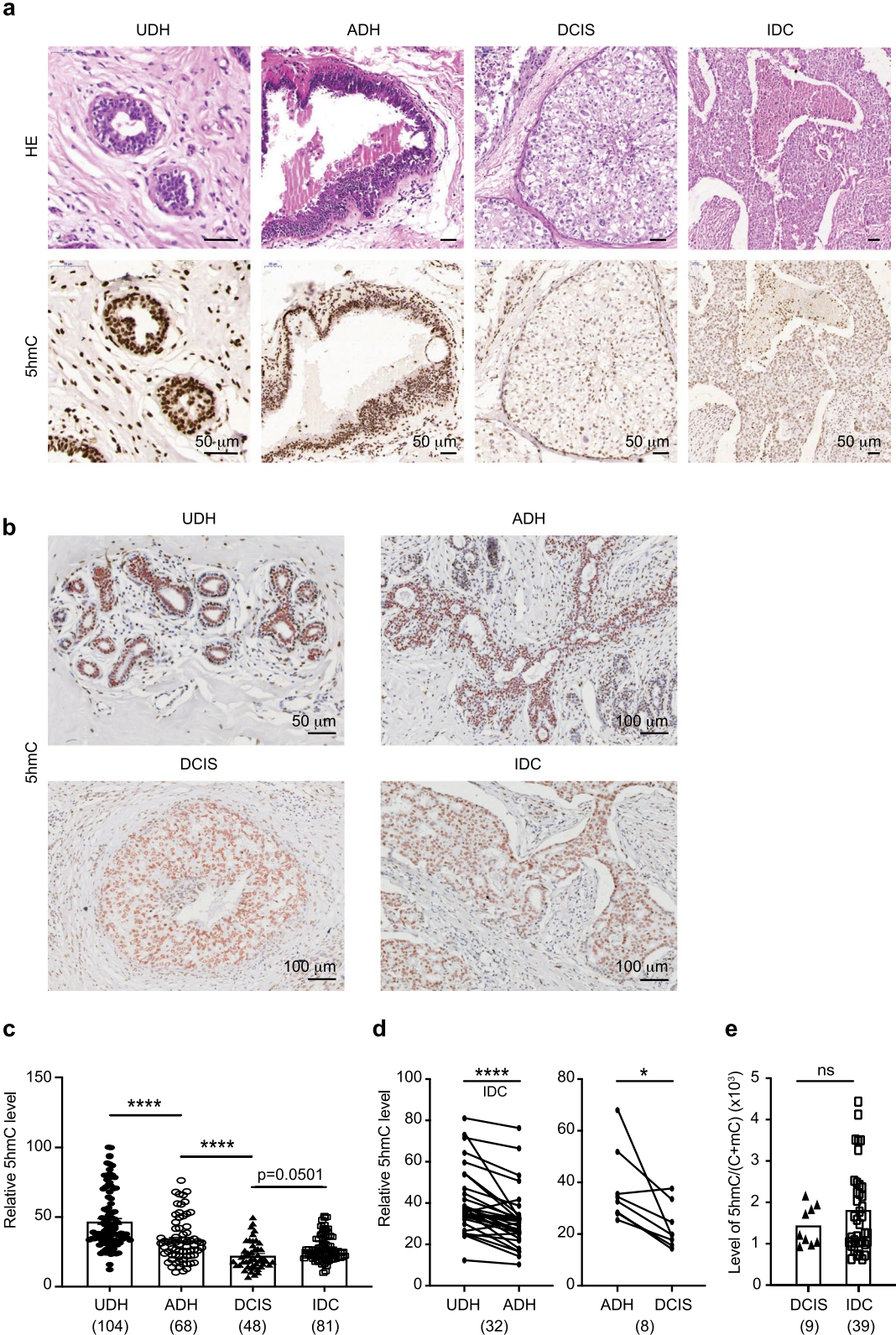
### 5-hydroxymethylcytosine during breast tumorigenesis

To examine the dynamic alterations in 5hmC associated with the progression of breast tumorigenesis at the genome-wide level, four distinct types of breast tumors (UDH, ADH, DCIS, and IDC) were utilized for hMeDIP-seq analysis (Tables 1, 2 and 3). Considering that breast cancer is a heterogeneous disease and mainly originates from epithelium, we specifically conducted macro-dissection to obtain abnormal ducts and cancer cells for 5hmC analysis (Fig. 2a and Additional file 1).

The genome-wide profiling revealed that 5hmC peaks were predominantly distributed in intronic and distal intergenic regions across all stages. However, as lesions progressed from UDH to DCIS, a gradual reduction in the proportion of 5hmC peaks situated in promoters, exons, 5'UTRs, and 3'UTRs was observed, followed by an increase in IDC (Additional file 2: Fig. S1a). Furthermore, a marked enrichment of 5hmC peaks was observed in the vicinity of transcription start sites (TSSs) in both UDH and IDC samples. In contrast, 5hmC modifications were evenly distributed across gene bodies in ADH and

(See figure on next page.)

**Fig. 1** DNA 5hmC modification exhibits dynamic change in the process of breast tumorigenesis. **a** Representative images of H & E staining and DNA 5hmC immunohistochemical staining (IHC) in UDH, ADH, DCIS, and IDC tissues. **b** Representative images showing the ducts of breast precancerous lesions (UDH, ADH, DCIS, and IDC) identified for quantitative analyses of 5hmC staining. The nuclear epithelial cells in the targeted ducts were denoted in red circles. **c** Quantitative comparison of DNA 5hmC levels based on IHC images (**a**) across UDH, ADH, DCIS, and IDC samples. The numbers of samples in each group were marked in the parentheses. **d** Pairwise comparison of DNA 5hmC levels between UDH and ADH, ADH and DCIS, respectively. **e** Quantification of 5hmC levels obtained by LC-MS/MS analysis of each sample in DCIS and IDC groups. Scale bars represent 50  $\mu$ m and 100  $\mu$ m. *P* values were calculated by using unpaired *t*-test in **c** and **e**; paired *t*-test was used in **d**. \* $P < 0.05$ , \*\*\*\* $P < 0.0001$ , ns: no significant. UDH, usual ductal hyperplasia; ADH, atypical ductal hyperplasia; DCIS, ductal carcinoma in situ; IDC, invasive ductal breast cancer



**Fig. 1** (See legend on previous page.)

DCIS samples (Additional file 2: Fig. S1b). In terms of modification levels, a steady decline in 5hmC levels was observed from UDH to DCIS, followed by an increase from DCIS to IDC (Fig. 2b), which is consistent with our previous observations. Furthermore, similar trends in the alteration of 5hmC levels were observed in specific genomic regions across these four stages (Additional file 2: Fig. S1c).

Building upon these dynamic 5hmC changes, pairwise comparisons between adjacent stages were performed to identify differentially hydroxymethylated regions (DhMRs,  $P$  value  $< 0.05$  and  $|\text{Log}_2(\text{foldchange})| > 1$ ) in the three phases (Fig. 2c, Table 4, and Additional file 3: Table S3). Regarding genomic distribution, an increase in the proportion of hyper-DhMRs in promoters, exons, 5'UTRs, and 3'UTRs was observed from phase I to phase III, whereas the proportion of hypo-DhMRs in these regions decreased (Fig. 2d). Moreover, hypo-DhMRs in phase I and hyper-DhMRs in phases II–III were significantly enriched around TSS, while hypo-DhMRs in phases II–III displayed a bimodal distribution near TSS (Fig. 2e). Subsequently, it was observed that the number of genes harboring hyper-DhMRs (hyper-DhMGs) increased during breast tumorigenesis, while the number of genes harboring hypo-DhMGs decreased (Fig. 2f). KEGG enrichment analysis and GSEA analysis of DhMGs revealed that both hyper- and hypo-DhMGs were closely associated with cancer-related pathways (Additional file 2: Fig. S1d–S1e and Table S1). Notably, 2307 genes harboring 2938 DhMRs exhibited continual 5hmC gain (Fig. 2g), and these genes were enriched in pathways such as cancer, RAS, RAP1, MAPK signaling pathways, and axon guidance (Fig. 2h). Conversely, 2036 DhMRs, annotated to 1501 genes, exhibited continual 5hmC loss (Fig. 2i). The corresponding hypo-DhMGs were significantly enriched in cancer-related pathways, including cell cytoskeleton, cell adhesion, and Hippo signaling pathway (Fig. 2j). These comprehensive genome-wide profiling analyses of breast lesions have unveiled the dynamic changes in DNA 5hmC throughout the four stages of breast tumorigenesis. The distribution of DhMRs around TSS suggests their potential roles in transcriptional

regulation. Additionally, genes exhibiting altered 5hmC modifications may play a crucial role in the development of early-stage breast cancer, particularly those associated with continual 5hmC changes. These findings underscore the significance of 5hmC as a potential epigenetic regulator in breast tumorigenesis.

#### Coincidence of dynamic 5-hydroxymethylcytosine with active DNA demethylation in the early stage of breast precancerous lesions

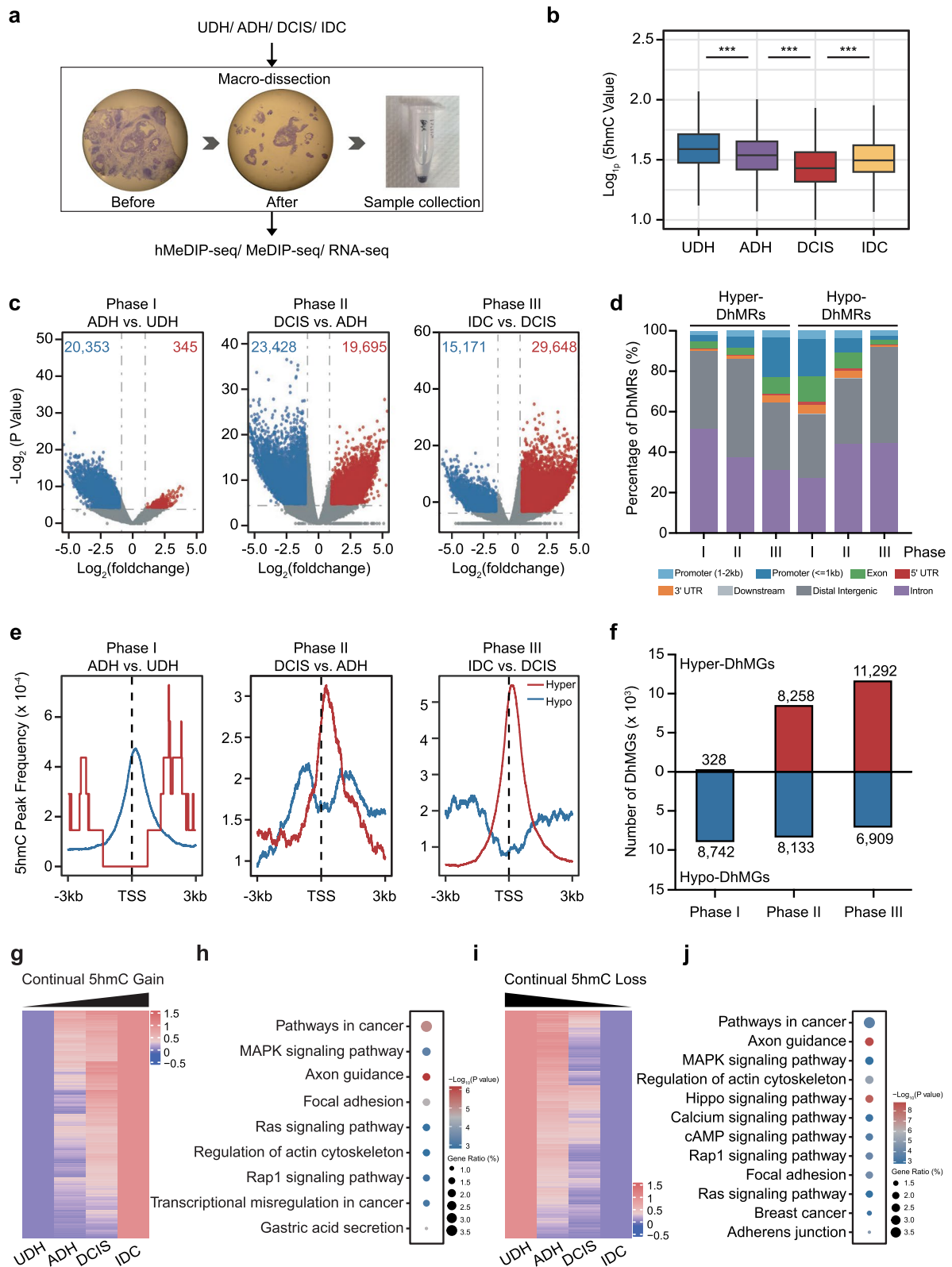
In addition to the role of 5hmC, 5mC has also been recognized as a significant contributor to cancer progression, including breast cancer [52]. However, the mechanism by which DNA 5hmC and 5mC orchestrate in different phases of breast tumorigenesis to promote tumor progression remains unclear. Recent microarray-based analyses have indicated that in spite of limited changes in 5mC between DCIS and IDC, there are more differentially methylated genes (DMGs) between normal breast tissue and DCIS [34]. Consequently, we conducted genome-wide 5mC profiling analyses of UDH, ADH, and DCIS (Tables 1, 2 and 3) to delve into the dynamic changes of DNA 5mC during breast tumorigenesis.

Unlike 5hmC (Additional file 2: Fig. S1a), the percentage of 5mC peaks located in each regulatory elements remained consistent across different stages of breast tumors (Additional file 2: Fig. S2a). Besides, the genomic pattern of 5mC was similar among all samples, characterized by a depletion around TSS (Additional file 2: Fig. S2b). In addition, it was found that the global level of 5mC firstly increased from UDH to ADH, followed by a decrease from ADH to DCIS (Fig. 3a). The trend of global 5mC levels was mirrored in the promoter, downstream, intergenic, and intron regions (Additional file 2: Fig. S2c). In contrast, the 5mC levels in the exon, 5'UTR, and 3'UTR regions were higher in UDH and DCIS compared to ADH (Additional file 2: Fig. S2c).

Subsequently, differentially methylated regions (DMRs) that were either reinforced (hyper-DMRs) or diminished (hypo-DMRs) as the lesions progressed, were identified by comparative analysis of 5mC profiles between adjacent stages. Notably, 11,507 hypo-DMRs

(See figure on next page.)

**Fig. 2** Genome-wide characteristics and dynamic changes of DNA 5hmC in different stages of breast tumors. **a** Process of breast tumor's macro-dissection. **b** Boxplots showing relative DNA 5hmC levels of UDH, ADH, DCIS, and IDC samples. **c** Volcano plots highlighting DhMRs ( $P$  value  $< 0.05$  and  $|\text{Log}_2(\text{foldchange})| > 1$ , hyper: red; hypo: blue) in each phase of breast tumorigenesis. **d** Distribution of hyper-DhMRs and hypo-DhMRs across genomic regions in each phase of breast tumorigenesis. **e** Distribution of DhMRs in TSS-surrounding regions (TSS  $\pm 3$  kb), the red line indicates hyper-DhMRs in the later stage, and the blue line indicates hypo-DhMRs in the later stage. **f** Numbers of genes possessing DhMRs (DhMGs) in each phase of breast tumorigenesis. **g** Heatmap showing the regions that exhibit continual increase of 5hmC from UDH to IDC. **h** KEGG enrichment analysis of the genes accompanied with continual hyper-DhMRs. **i** Heatmap showing the regions that exhibit continual decrease of 5hmC from UDH to IDC. **j** KEGG enrichment analysis of the genes accompanied with continual hypo-DhMRs. \*\*\* $P < 0.001$ . DhMRs: differentially hydroxymethylated regions; TSS: transcription starting site



**Fig. 2** (See legend on previous page.)



and 157 hyper-DMRs were identified in phase I, as well as 5,112 hypo-DMRs and 14,112 hyper-DMRs in phase II (Fig. 3b, Table 4, and Additional file 4: Table S4). Concurrently, an increase in the percentage of DMRs (both hyper-DMRs and hypo-DMRs) located in promoters was observed in phase II compared to phase I (Fig. 3c). Interestingly, both hyper- and hypo-DMRs in phase II exhibited a preference for enrichment around TSS, whereas such feature was not observed in phase I (Fig. 3d). Furthermore, KEGG and GSEA analyses revealed that the genes harboring either hypo- or hyper-DMRs were enriched in pathways closely associated with cancer progression (Additional file 2: Fig. S2d–S2f and Table S2).

Given that 5hmC is established through the oxidation of 5mC, we further investigated the correlation between 5hmC- and 5mC-modified genes during the aforementioned phases based on pairwise comparisons of DhMRs and DMRs in each phase. Considering the enrichment of DhMRs and DMRs around TSS, their positional relations were initially explored. In phase I, the majority of DMRs were enriched in the regions centered around DhMRs, while DhMRs were evenly distributed from DMRs to their downstream regions (Fig. 3e). Conversely, in phase II, it was observed that most DMRs were centered around the DhMRs, and vice versa (Fig. 3f). Subsequently, 332 and 1,269 regions displaying significant changes in both 5hmC and 5mC simultaneously were identified ( $P$  value  $< 0.05$  and  $|\text{Log}_2(\text{foldchange})| > 1$ ), respectively (Additional file 2: Fig. S2g). Of note, it was observed that the changes in 5hmC within DMRs were more prominent compared to those outside DMRs (Fig. 3g). In contrast, the changes in 5mC across DhMRs were comparable to those outside DhMRs (Fig. 3h). These results suggest that DhMRs and DMRs overlap each other during breast tumorigenesis. Additionally, the effect of 5mC changes on DhMR patterns appears to be more pronounced than 5hmC changes on DMR patterns, shedding light on the complex interplay between 5hmC- and 5mC-modified genes in breast tumorigenesis.

### Synchronization of active DNA demethylation with active histone modifications to be involved in transcriptional regulation

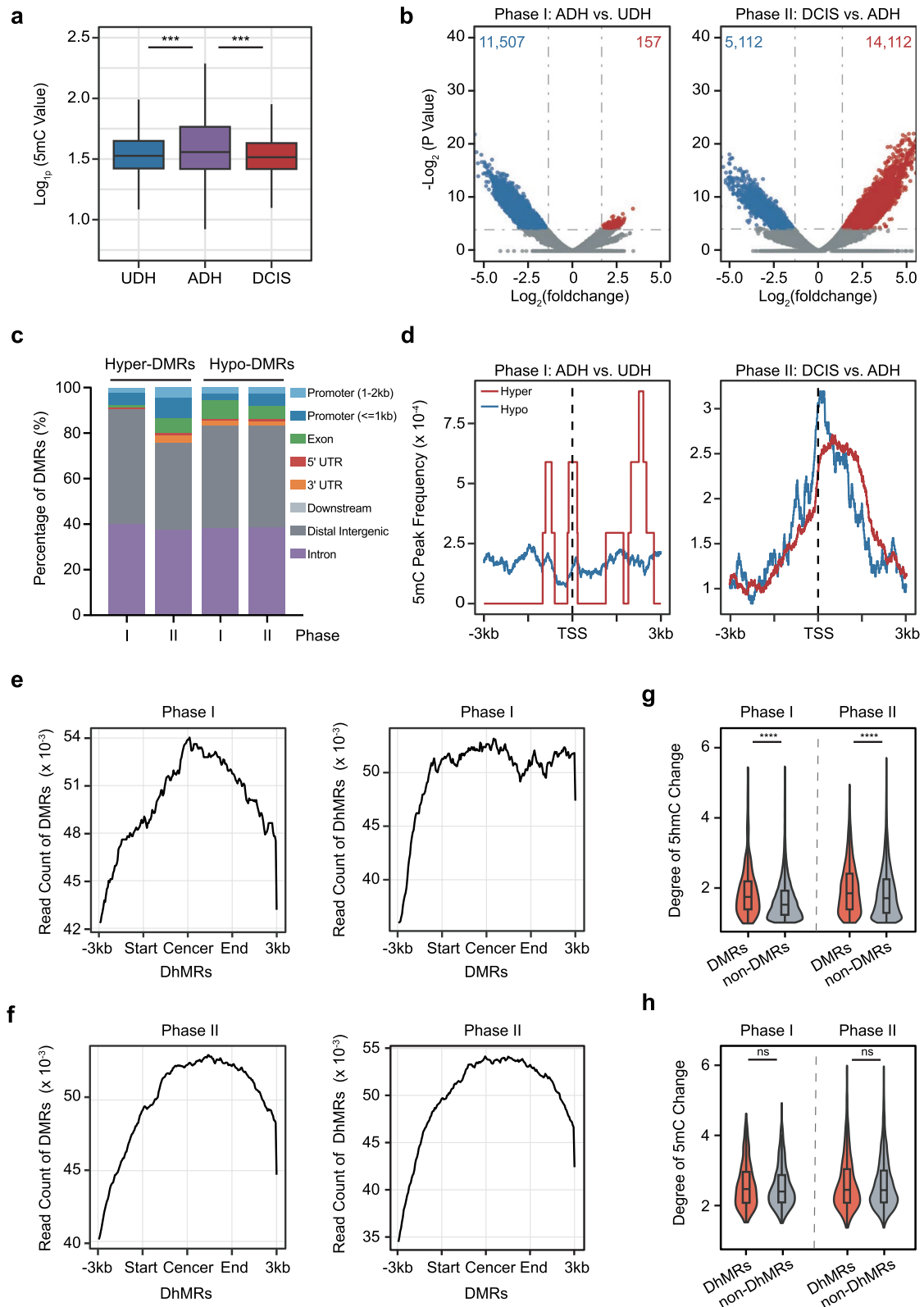
Both DNA methylation and hydroxymethylation are pivotal transcriptional regulators with profound effects on gene expression. To investigate the impact of dynamic changes of 5hmC and 5mC on RNA expression during breast tumorigenesis, a comprehensive set of publicly available RNA-seq data from breast tumors (GSE47462) [42] was re-analyzed, which represents various stages of breast tumorigenesis (Additional file 2: Fig. S3a). As shown in Fig. 4a, a substantial number of differentially expressed genes (DEGs) concurrent differential 5hmC or 5mC levels. These DhMRs/DMRs-associated DEGs are implicated in the development of cancer (Additional file 2: Fig. S3b–S3c).

To gain deeper insights into the effects of DNA 5hmC and 5mC on RNA expression within breast tumors, RNA-seq analyses were performed on the same samples previously subjected to hMeDIP-seq and MeDIP-seq (Tables 1, 2 and 3). The genes were categorized into three evenly sized groups based on the trisectional quantiles of their RNA expression levels, and the distribution of 5hmC and 5mC among these genes in each group was assessed. As a result, a positive correlation was observed between RNA expression levels and the frequencies of 5hmC in TSS and upstream regions (Fig. 4b and Additional file 2: Fig. S3d). Conversely, 5mC levels near TSS exhibited a negative regulatory effect on RNA expression (Fig. 4c and Additional file 2: Fig. S3e).

On the basis of the multifaceted effects of 5hmC and 5mC on transcriptional regulation and the crosstalk between DNA modifications and histone marks, we next investigated whether histone marks play a role in the regulatory effects of DNA 5hmC and 5mC. The distribution of common histone marks, including H3K27ac, H3K27me3, H3K36me3, H3K4me1, H3K4me3, and H3K9me3, around the corresponding DhMRs and DMRs was explored utilizing ChIP-seq data of MCF7 cells from the ENCODE database [53]. This analysis unveiled the correlations between DhMRs/DMRs and histone modifications occurring in

(See figure on next page.)

**Fig. 3** Dynamic changes of genomic 5mC along with 5hmC in the early phase of breast tumorigenesis. **a** Boxplots showing relative DNA 5mC levels of UDH, ADH, and DCIS samples. **b** Volcano plots displaying DMRs ( $P$  value  $< 0.05$ , and  $|\text{Log}_2(\text{foldchange})| > 1$ ) identified in the early stages of breast precancerous lesions. **c** Distribution of hyper-DMRs and hypo-DMRs across genomic regions in the early-stage breast precancerous lesions. **d** Distribution of DMRs in TSS-surrounding regions ( $\text{TSS} \pm 3$  kb), the red lines indicate hyper-DMRs in the later stage, and the blue lines indicate hypo-DMRs in the later stage. **e, f** Enrichment of DMRs around DhMRs (left), and the enrichment of DhMRs around DMRs (right) in phase I (**e**) and phase II (**f**). **g** Comparison of the degree of 5hmC changes between DMRs and non-DMRs. **h** Comparison of the degrees of 5mC changes between DhMRs and non-DhMRs. \*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0001$ , ns: no significant. DMRs: differentially methylated regions; DhMRs: differentially hydroxymethylated regions; TSS: transcription starting site



**Fig. 3** (See legend on previous page.)

**Table 3** Information of the samples used for genomic profiling analyses of DNA 5hmC, 5mC, and RNA-seq

Sample	Method	Raw pairs	Trim pairs	Unique pairs	Unique rate (%)	(h)MRs	(h)MGs
UDH1	Input	21,694,103	21,201,259	16,664,861	76.82		
	hMeDIP-seq	66,437,141	64,760,579	50,860,166	76.55	264,584	20,565
	MeDIP-seq	144,661,672	137,764,572	91,912,952	63.54	342,764	21,519
	RNA-seq	35,116,804	34,777,444				
UDH2	Input	89,503,093	87,925,630	69,568,313	77.73		
	hMeDIP-seq	33,220,332	31,390,467	25,538,825	76.88	88,399	15,941
	MeDIP-seq	41,524,771	39,741,983	29,509,201	71.06	75,024	15,311
	RNA-seq	55,411,299	55,162,069				
UDH3	Input	24,001,398	23,161,184	16,248,049	67.70		
	hMeDIP-seq	29,832,486	28,625,365	20,609,759	69.08	178,785	20,477
	MeDIP-seq	67,025,400	64,749,035	49,146,230	73.32	467,752	22,614
	RNA-seq	45,575,272	45,243,353				
ADH1	Input	33,389,588	32,532,107	24,891,123	74.55		
	hMeDIP-seq	25,811,182	24,875,325	19,623,977	76.03	67,914	13,916
	MeDIP-seq	25,585,307	24,725,851	16,557,385	64.71	220,253	21,159
	RNA-seq	55,411,299	55,162,069				
ADH2	Input	119,148,894	117,220,763	93,656,263	78.60		
	hMeDIP-seq	34,064,157	32,339,195	27,069,535	79.47	183,055	19,948
	MeDIP-seq	37,668,391	36,032,190	28,819,395	76.51	36,487	11,499
	RNA-seq	41,990,728	41,686,531				
ADH3	Input	93,731,982	91,765,402	72,890,970	77.77		
	hMeDIP-seq	56,840,643	54,773,750	43,923,707	77.28	389,717	22,402
	MeDIP-seq	59,689,825	57,174,175	44,050,258	73.80	313,726	21,715
	RNA-seq	37,657,850	37,377,912				
DCIS1	Input	57,929,335	55,560,631	43,656,365	75.36		
	hMeDIP-seq	35,728,170	33,599,009	26,755,783	74.89	101,558	17,806
	MeDIP-seq	40,580,168	39,062,373	29,861,708	73.59	161,844	19,548
	RNA-seq	56,082,948	55,605,593				
DCIS2	Input	94,202,415	92,401,771	73,323,285	77.84		
	hMeDIP-seq	92,841,382	90,145,171	72,515,792	78.11	846,147	23,885
	MeDIP-seq	58,205,897	56,357,412	44,152,287	75.86	386,749	22,488
	RNA-seq	35,781,614	35,456,120				
DCIS3	Input	58,623,645	56,434,336	42,773,299	72.96		
	hMeDIP-seq	52,950,098	51,323,358	36,063,719	68.11	389,751	20,112
	MeDIP-seq	76,335,219	74,010,486	57,364,929	75.15	331,041	19,732
	RNA-seq	47,365,157	46,980,799				

**Table 4** Numbers of DhMRs and DMRs in each phase of breast tumorigenesis

	Hyper-DhMRs	Hypo-DhMRs	Hyper-DMRs	Hypo-DMRs
Phase 1: ADH vs UDH	345	20,492	157	11,486
Phase 2: DCIS vs ADH	19,704	23,582	14,108	5107
Phase 3: IDC vs DCIS	29,781	15,191		

*P* value < 0.05 and |Log<sub>2</sub>(foldchange)| > 1

distinct phases of breast tumorigenesis. Specifically, hypo-DhMRs in phase I and hyper-DhMRs in phase III exhibited enrichment of the active histone marks H3K27ac and H3K4me3, which are frequently located at active promoters and associated with active transcription (Fig. 4d–4e). In parallel, hypo-DMRs in II appeared to show a slight enrichment of enhancer-specific histone modifications H3K27ac and H3K4me1 (Fig. 4f–4g). Overall, both DhMRs and DMRs exhibited specific associations with several active histone marks, rather than repressive ones, such as H3K27ac and H3K4me3. Moreover, these associations were phase-specific, suggesting a dynamic interplay between DNA modifications and histone marks that contributes to the progression of breast tumorigenesis. These results uncover the significance of 5hmC and 5mC enrichment around TSSs in influencing gene transcription, possibly through crosstalk with histone marks. Such regulatory mechanism is of great significance for promoting breast tumorigenesis.

#### Cooperation of TET2 with ER complex and FOS to function in genomic repatterning of 5-hydroxymethylation and 5-methylation

Conversion of 5mC to 5hmC relies on the DNA demethylation enzyme TETs. As previously reported, the expression of TET2 decreases with the progression of breast precancerous lesions [28]. To investigate the effect of TET2 on DNA demethylation during breast tumorigenesis, we conducted a comprehensive analysis by integrating public and in-house data from different stages of breast lesions.

We first obtained ChIP-seq data of TET2 in MCF7 cells from the GEO database (GSE153251) [46] to identify the binding regions of TET2. A significant proportion (approximately 56.4%) of TET2 binding regions were distributed across distal intergenic and promoter regions (Additional file 2: Fig. S4a). Meanwhile, it was observed that TET2 preferred to bind active and primed enhancers and promoters marked by H3K27ac, H3K4me1, and H3K4me3. Intriguingly, TET2-binding sites around H3K27ac and H3K4me1 displayed a bimodal enrichment pattern (Additional file 2: Fig. S4b). To explore the role of TET2 in transcriptional regulation, we performed integrated analyses on RNA-seq and hMeDIP-seq data

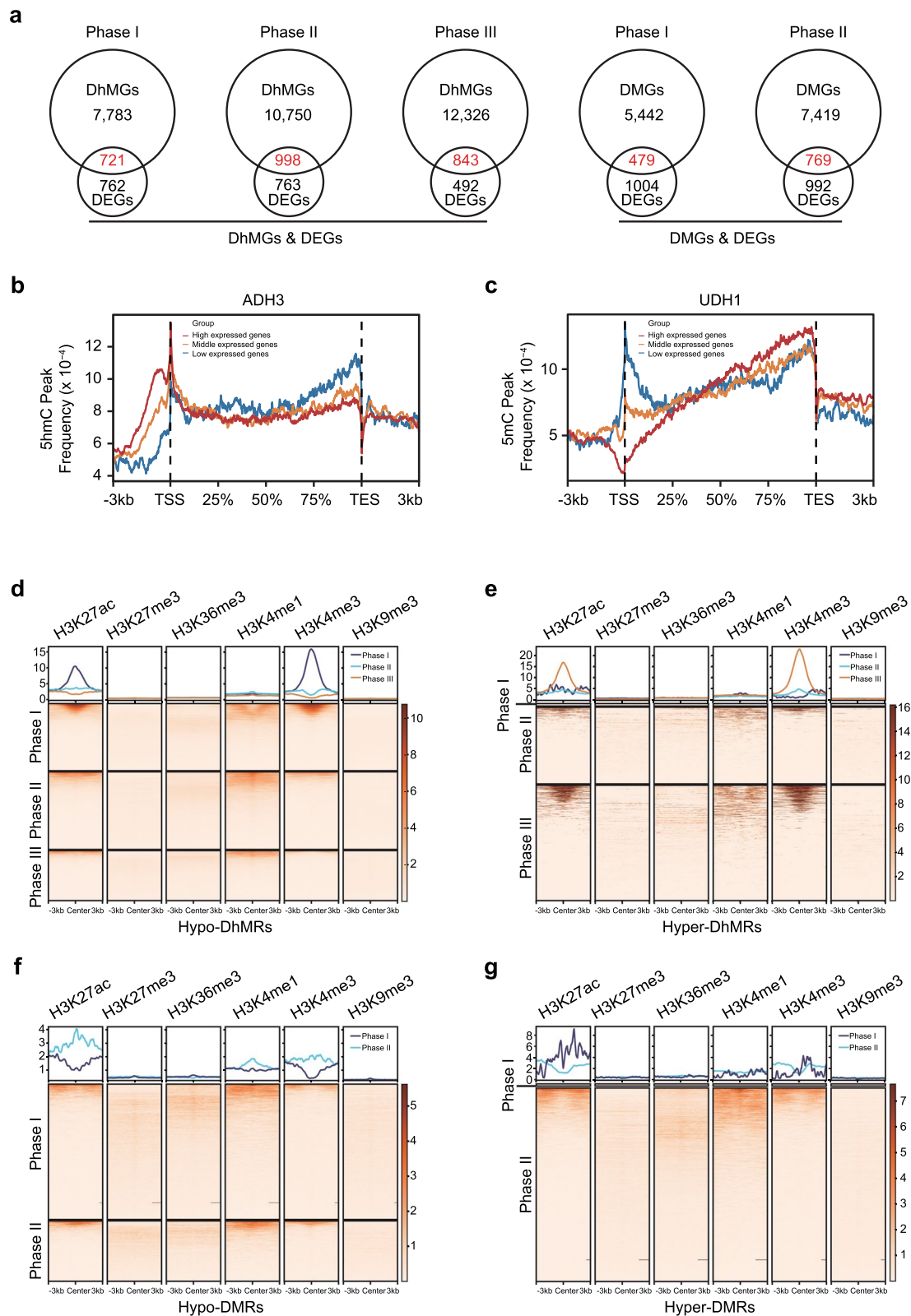
obtained from breast tumors. Our findings highlighted that both TET2-binding sites and 5hmC deposition exerted a positive effect on RNA expression, with the highest levels of RNA expression in genomic regions meeting both criteria (Fig. 5a and Additional file 2: Fig. S4c). Consequently, we proposed that TET2-related 5hmC modification significantly contributes to the transcriptional regulation in breast tumors.

To gain a deeper understanding of the role of TET2 in modulating the dynamic changes of 5hmC during breast tumorigenesis, we compared the levels of 5hmC between hMRs with and without TET2 enrichment at each stage of breast tumors (Fig. 5b). The changing pattern of 5hmC in TET2-binding regions across different stages was in agreement with the global 5hmC changes in breast tumorigenesis (Fig. 2b), and the 5hmC levels in TET2-binding regions across different stages also exhibited a continuous decline from UDH to DCIS, followed by an increasing trend. In terms of enrichment levels, 5hmC levels in TET2-binding regions were stably and significantly higher than those in non-TET2-binding regions in all stages of breast lesions (Fig. 5b). Additionally, thousands of DhMRs in all phases were recognized by TET2 (Additional file 2: Fig. S4d). The proportion of TET2-binding hyper-DhMRs in promoter, exon, and 3'UTR regions progressively increased from phase I to phase III, while hypo-DhMRs displayed an opposite trend (Fig. 5c). Based on our observations that DhMRs are preferentially located in cis-regulatory regions marked by active histone modifications (Fig. 4d–e), we further explored the enrichment of histone marks near TET2-binding DhMRs and observed a robust enrichment of active histone mark H3K27ac (Fig. 5d–e). Moreover, enhancer-specific histone mark H3K4me1 displayed a similar pattern to that of H3K27ac in TET2-binding hypo-DhMRs, albeit at a relatively low level of enrichment (Fig. 5d). Interestingly, both TET2-binding hypo-DhMRs in phase I and TET2-binding hyper-DhMRs in phase III exhibited associations with H3K4me3 (Fig. 5d–e).

Given the interplay between DhMRs and DMRs in genomic distribution and the crucial role of TET2 in DNA demethylation, we hypothesized that TET2 might also be involved in dynamic DMRs concurrently with TET2-targeted DhMRs. In our study, a less number of

(See figure on next page.)

**Fig. 4** DNA epigenetic modifications are associated with gene expression and specific histone modifications' enrichment. **a** Venn diagram showing the overlap between DEGs and DhMGs/DMGs in each phase of breast tumorigenesis. **b, c**. Distribution profiles of 5hmC (**b**) and 5mC (**c**) peak in genes expressed at high (red), medium (yellow) and low (blue) levels. **d, e**. Enrichment of H3K27ac, H3K27me3, H3K36me3, H3K4me1, H3K4me3, and H3K9me3 signals flanking hypo-DhMRs (**d**) and hyper-DhMRs (**e**) in each phase of breast tumorigenesis. **f, g**. Enrichment of H3K27ac, H3K27me3, H3K36me3, H3K4me1, H3K4me3, and H3K9me3 signals around hypo-DMRs (**f**) and hyper-DMRs (**g**) in early phases of breast tumorigenesis. DEG: differentially expressed genes; DhMG: differentially hydroxymethylated gene; DMG: differentially methylated gene



**Fig. 4** (See legend on previous page.)

TET2-binding DMRs were detected compared to TET2-binding DhMRs (Additional file 2: Fig. S4e). However, the TET2-binding hypo-DMRs exhibited a similar enrichment pattern of active histone marks to TET2-binding DhMRs, such as H3K27ac and H3K4me1 (Additional file 2: Fig. S4f). Thus, co-occupation of active histone marks and TET2-binding DhMRs/DMRs in promoter and enhancer regions suggests that TET2-related 5hmC and 5mC modifications play crucial roles in transcriptional regulation via manipulating the activity of regulatory elements during breast tumorigenesis.

As TET2 is capable of coactivating crucial transcriptional factors (TFs) such as ER $\alpha$  (ESR1) and GATA3 in breast cancer cells, we then explored the potential TFs that involved in TET2-related DhMRs/DMRs during breast tumorigenesis. Through an examination of enriched motifs in TET2-binding DhMRs, the motifs associated with several TFs were identified, including ESR1, FOXA1, FOS, FOSL2, FOXM1, and JUNB (Fig. 5f). Similar to TET2, these TFs are typically located within active chromatin regions marked by H3K27ac, H3K4me1, and H3K4me3 (Additional file 2: Fig. S4g). To confirm the involvement of aforementioned TFs and GATA3 [46], we compared their enrichment levels surrounding DhMRs/DMRs regions at each phase of breast tumorigenesis. Our analyses revealed that ESR1, GATA3, FOXA1, and FOS exhibited pronounced enrichment around TET2-binding DhMRs in at least one phase of breast tumorigenesis (Fig. 5g and Additional file 2: Fig. S4h). As shown in Fig. 4b and 4c, both 5hmC and 5mC DNA modifications associated with transcriptional regulation were predominantly distributed around TSS, which prompts the investigation of whether these four TFs are also located around DMRs during breast tumorigenesis. In Additional file 2: Fig. S4i, a strong enrichment of these TFs was observed in TET2-binding DMRs in phase I. In contrast, only TET2-binding hypo-DMRs, rather than hyper-DMRs, were enriched with the TFs in phase II. Conversely, no similar phenomenon was observed in non-TET2-binding DMRs (Additional file 2: Fig. S4j). Therefore, we deduced that TET2-binding DhMRs and hypo-DMRs located

in promoters and enhancers play a role in modulating gene expression. Additional transcriptional factors ESR1, GATA3, FOXA1 and FOS co-localize with TET2 and are likely to be involved in the dynamic changes of 5hmC and 5mC throughout breast tumorigenesis.

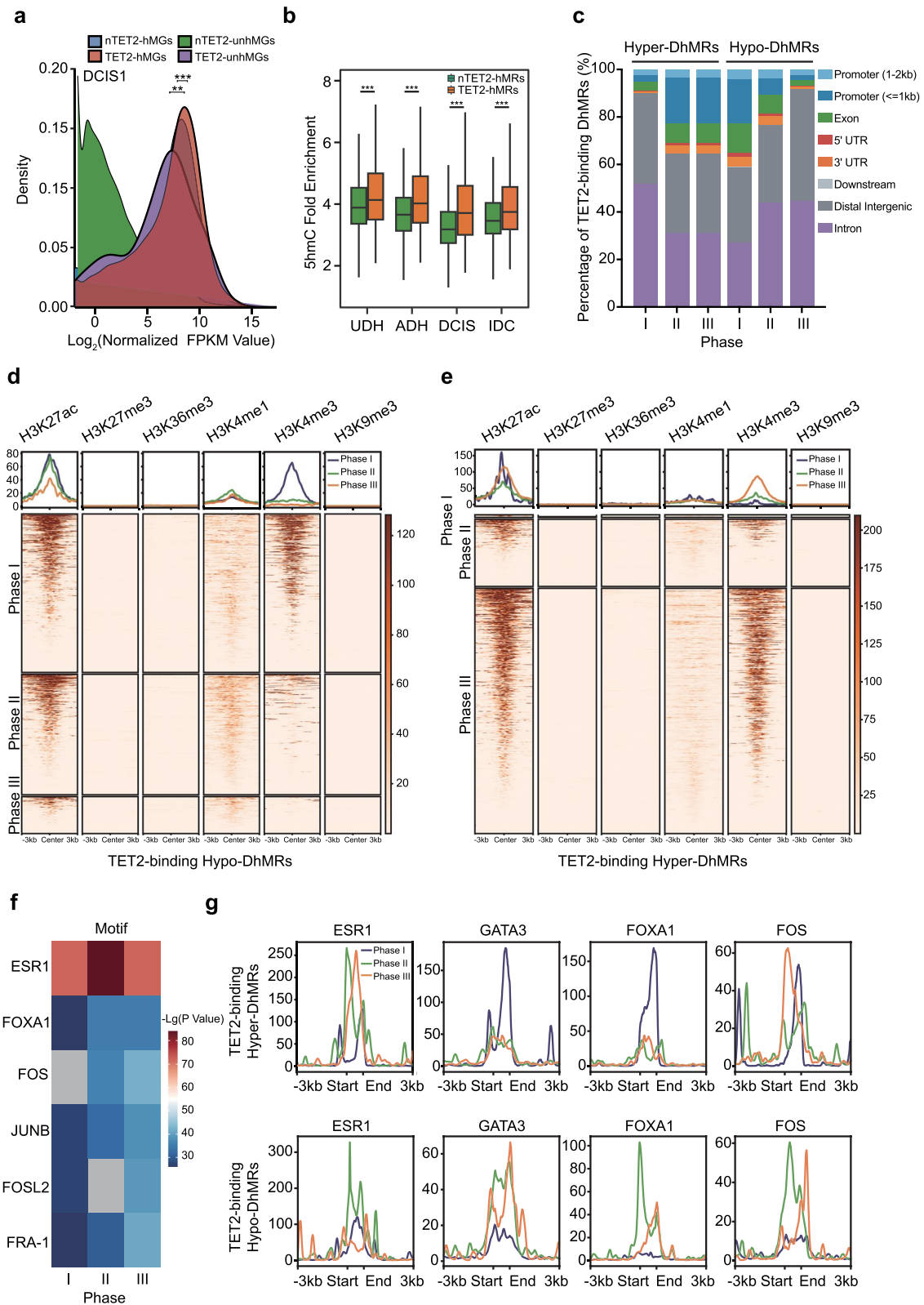
#### Identification of differentially hydroxymethylated regions as potential biomarkers for detecting early-stage breast cancer

In recent decades, liquid biopsy has attracted increasing attention as a non-invasive alternative to tissue biopsy for cancer screening and monitoring [54]. Although cfDNA 5hmC has proven its value as a biomarker for various cancers [15, 20, 55, 56], its potential application in breast cancer remains underexplored. Given the integral role of 5hmC in breast tumorigenesis, we utilized 5hmC sequencing data of cfDNA from a previous study [50] to identify common DhMRs shared between cfDNA and primary breast tumors. Such DhMRs hold promise as diagnostic markers for breast cancer screening.

Compared to healthy control samples, only a slight decrease in the global levels of 5hmC in cfDNA from patients with early-stage breast cancer was observed, while the changes in primary tumor samples were more pronounced (Additional file 2: Fig. 5a–S5b). Despite limited alterations in global 5hmC levels, 4881 hyper-DhMRs and 3570 hypo-DhMRs, annotated to 3408 and 2676 genes, respectively, were identified in cfDNA. In the meantime, 3718 hyper-DhMRs and 12,254 hypo-DhMRs, annotated to 2977 and 5468 genes, respectively, were found in breast cancers compared to benign tumors ( $P$  value < 0.05 and  $|\text{Log}_2(\text{foldchange})| > 1$ ) (Fig. 6a–6b). DhMRs in cfDNA and breast tissues were predominantly enriched in intronic and intergenic regions, followed by promoter regions (Additional file 2: Fig. S5c–S5d). Subsequent KEGG analyses revealed that DhMGs in cfDNA exhibited a strong enrichment in cancer-related pathways, such as PI3K-Akt, RAS-RAP1, MAPK, and cell adhesion (Fig. 6c). In addition, DhMGs in breast cancers were significantly enriched in cancer-related pathways (Fig. 6d). To further explore the 5hmC signals that can

(See figure on next page.)

**Fig. 5** TET2 co-localizes with TFs and participates in breast tumorigenesis through mediating DNA demethylation. **a** Density plot showing the distribution of expression level of RNAs encoded by TET2-binding genes or hydroxymethylated genes. **b** Boxplots showing the relative 5hmC levels of the TET2-binding hMRs (TET2-hMRs) and non-TET2-binding hMRs (nTET2-hMRs) in the UDH, ADH, DCIS, and IDC samples. **c** Distribution of TET2-binding hyper-DhMRs and hypo-DhMRs across genomic regions in each phase of breast tumorigenesis. **d, e** Enrichment of H3K27ac, H3K27me3, H3K36me3, H3K4me1, H3K4me3, and H3K9me3 around TET2-binding hypo-DhMRs (**d**) and TET2-binding hyper-DhMRs (**e**) in breast tumorigenesis. **f** Heatmap showing the enrichment of canonical motifs recognized by the several TFs in TET2-binding DhMRs in each phase of breast tumorigenesis. **g** Enrichment of the crucial transcriptional factors around TET2-binding hyper-DhMRs and hypo-DhMRs in breast tumorigenesis.  $**P < 0.01$ ,  $***P < 0.001$ . nTET2-hMGs: Hydroxymethylated genes without TET2-binding regions; nTET2-unhMGs: Genes without 5hmC modifications and TET2-binding regions; TET2-hMGs: Hydroxymethylated genes with TET2-binding regions; TET2-unhMGs: Genes with TET2 binding regions but without 5hmC modifications



**Fig. 5** (See legend on previous page.)

potentially be used for early detection of breast cancer, 146 hypo-DhMRs and 12 hyper-DhMRs were identified, which exhibited concurrent changes in cfDNA and breast cancer tissues (Fig. 6e, f). Among the DhMRs-annotated genes, *KLF15*, *PTPRG*, *PPARGC1B* [57], and *ZFH3* are closely related to estrogen signaling pathways and the development of mammary epithelial cells. Meanwhile, *UNC5A* [58], *PIK3AP1*, *IGF1R*, and *HIF1A* [59] have been proved to be crucial in the development and metastasis of breast cancer. These findings indicate that 5hmC signals in cfDNA may reflect the genomic characteristics of the primary breast cancers, and may become valuable candidates for early-stage breast cancer screening.

## Discussion

In this study, we comprehensively analyzed the genome-wide distribution of 5hmC and 5mC in different stages of breast tumors and the intricate crosstalk among 5hmC, 5mC, and histone modifications in transcriptional regulation. We also uncovered the pivotal role of TET2 in mediating the dynamic changes of 5hmC and 5mC throughout breast tumorigenesis, and identified key transcription factors (ESR1, GATA3, FOXA1, and FOS) that collaborate with TET2 in orchestrating transcriptional regulation in breast tumors. Furthermore, we identified synchronous DhMRs in cfDNA and primary breast tissues, which hold potential as liquid biopsy biomarkers for breast cancer screening.

Despite the considerable reduction in DNA 5hmC observed in various cancers, the exact dynamic changes of 5hmC throughout breast tumorigenesis remain elusive. Our results revealed a gradual decreasing trend in 5hmC levels from UDH to DCIS. However, contrary to previous study [28], we observed an upward trend from DCIS to IDC, as evidenced by IHC and LC-MS/MS analysis. Consistently, our hMeDIP-seq analysis showed that the number of hyper-DhMR/DhMG was twice as high as the number of hypo-DhMR/DhMG in phase III. Considering that those DhMGs are closely related to the adhesion and invasion of cancer cells, we deduce that the cause of this intricate change in 5hmC in phase III is to meet the need for tumor progression from DCIS to IDC. However, the determining factors orchestrating the dynamic changes of 5hmC remains to be investigated.

It has been reported that breast tumor is a highly heterogeneous disease, including both intra- and inter-tumor heterogeneity. Concretely, there are significant variations in 5hmC levels not only among epithelial cells, mesenchymal cells, and infiltrating lymphocytes (intra-tumor heterogeneity), but also among different tumors (inter-tumor heterogeneity). As breast precursor lesions primarily arise from the luminal epithelial compartment of TDLUs, here the focus was exclusively on the tumorigenesis of the luminal epithelial cells. Therefore, purely abnormal ducts and cancer cells were meticulously collected for genome-wide and transcriptome-wide analyses using macro-dissection exclusively. Moreover, to reduce the inter-tumor heterogeneity of breast cancers of different molecular subtypes, we only selected DCIS and IDC samples of luminal subtype, which account for two-thirds of primary breast cancers. Consistently, all external data (ChIP-seq of histone marks, TET2, and transcriptional factors) analyzed in this study were generated using MCF7 cell line, which also belongs to the luminal subtype. Therefore, besides our findings about 5hmC in the breast cancer of luminal subtype, the role of 5hmC in breast cancers of other molecular subtypes remains elusive.

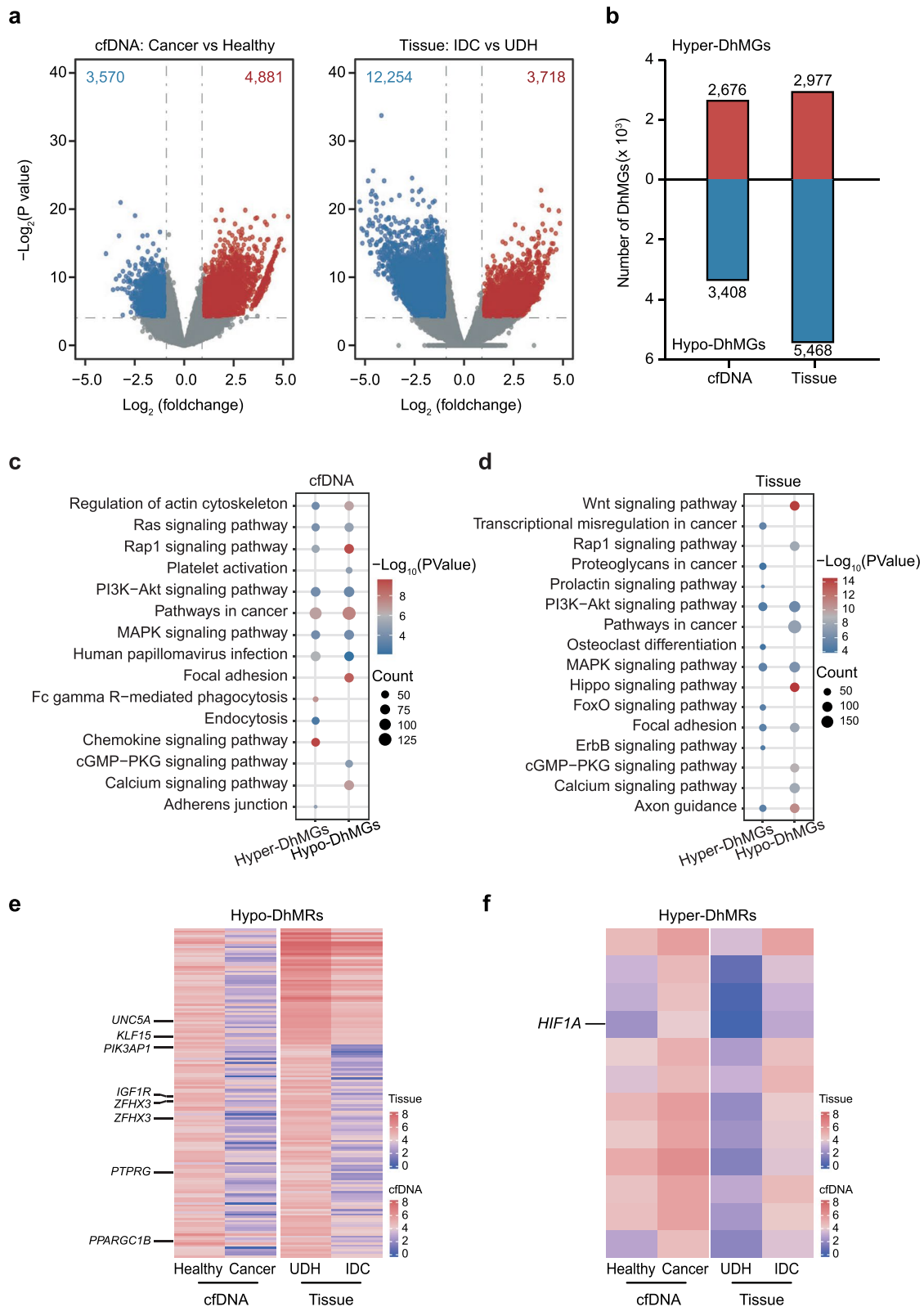
The genome-wide profiling revealed a broad reduction of both 5hmC and 5mC in the initial phases of breast tumorigenesis. Notably, DhMRs and DMRs were frequently situated around TSS and overlapped each other in phase I and II, underscoring the significance of active DNA demethylation in the early phases of breast tumorigenesis. In our study, changes in 5mC were more subdued than those in 5hmC throughout breast tumorigenesis, and the correlation between histone marks and DhMRs was notably stronger than that observed with DMRs, which lead to the speculation that 5hmC plays a more prominent role in the progression of breast cancer.

Intriguingly, a close association was observed between DhMRs and histone modifications at active enhancers and promoters other than gene bodies. These connections suggest that 5hmC located in these regulatory regions may influence the enrichment of the corresponding chromatin marks. Though the link between 5mC and transcriptional silencing is well established, no correlation between DMRs and repressive histone marks such as H3K27me3 and H3K9me3 was observed in breast lesions, as previously identified in other tissues [60, 61]. Similar to DhMRs,

(See figure on next page.)

**Fig. 6** Identification of potential 5-hydroxymethylcytosine signatures in cfDNA and early-stage primary breast cancers. **a** Volcano plots displaying hypo-DhMRs and hyper-DhMRs identified in cfDNA (left) and breast cancer tissues (right) compared to healthy controls. **b** The number of genes marked with hyper-DhMRs and hypo-DhMRs in cfDNA and breast cancer tissues. **c** KEGG enrichment analysis of hyper-DhMGs and hypo-DhMGs in cfDNA of breast cancer patients. **d** KEGG enrichment analysis of hyper-DhMGs and hypo-DhMGs in breast cancer tissues. **e, f** Heatmap showing the 5hmC levels of 146 regions exhibiting 5hmC loss (**e**) and 12 regions exhibiting 5hmC gain (**f**) in both cfDNA and breast tissues. Healthy: healthy controls; Cancer: patients with breast cancer





**Fig. 6** (See legend on previous page.)

DMRs in our study were also located in active chromatin regions, but with less pronounced enrichment. These observations hint that the repressive impact of 5mC on transcription may involve its collaboration with active histone marks or other repressive histone marks (except for H3K27me3 and H3K9me3) within breast tissue. In alignment with active histone marks, the abundance of 5hmC and 5mC around TSS showed a strong correlation with gene transcription levels, as described in previous studies [62, 63]. Collectively, our findings suggest that during breast tumorigenesis, both 5hmC and 5mC play a pivotal role in transcriptional regulation by coordinating with active histone marks in regulatory regions.

In the present study, we found that among all DhMRs/DMRs, active histone marks tended to be enriched in the TET2-binding DhMRs/DMRs, particularly in the promoter regions, indicative of co-localization of various TFs and TET2. Consistent with previous reports [46], ESR1 and GATA3 were observed to be involved in TET2-binding DhMRs. Furthermore, FOXA1, the pioneer factor for ESR1, was enriched in TET2-binding DhMRs. Although the interaction between FOXA1 and TET2 in prostate cancer has been reported, the role of FOXA1 in TET2-related DhMRs in breast cancer remains to be investigated. Given these observations, we speculated that the interaction of the ER complex with TET2 is vital in breast tissue. Additionally, in a study by Broome et al. [46], no global 5mC change was observed in TET2-knock down breast cancer cell lines. However, thousands of TET2-binding DMRs were identified through our comprehensive analysis of profiling data from clinical samples. Furthermore, the enrichment of ESR1, GATA3, and FOXA1 around TET2-binding DMRs mirrored the patterns observed in TET2-binding DhMRs, with ESR1 being notably more enriched in TET2-binding DMRs than in DhMRs. Therefore, we speculated that the TET2-ER complex may drive breast tumorigenesis by affecting both DNA methylation and demethylation. In addition to the ER complex, the involvement of the proto-oncogene FOS in TET2-related DhMRs/DMRs was first proposed. Moreover, FOXM1 was found to be enriched in TET2-binding DMRs in breast tumorigenesis. These findings indicate that FOS and FOXM1 may be additional interactors of TET2 in breast tumors, but further studies are needed to elucidate the nature of these interactions and the role of FOXM1 in TET2-binding DMRs.

Research on cfDNA-based 5hmC has made significant progress in cancer screening, primarily in the context of the digestive and hematologic systems [15, 20, 55, 56]. Recently, Curtis et al. proposed that metastatic seeding of breast cancer may occur 2–4 years before the diagnosis of the primary tumor [64]. Building on data from

Quake et al.'s study on cfDNA [50], thousands of DhMRs associated with breast cancer were identified between breast cancer patients and healthy controls. Compared to organs of the digestive system, changes of 5hmC in the cfDNA of breast cancer patients were relatively subtle, and these changes may be attributed to the limited blood flow to the breast. Therefore, the development of more sensitive 5hmC detection methods with low-input cfDNA remains imperative to gather more valuable insights for cancer screening. Our integrative analysis of cfDNA and breast tissue collectively suggests that 5hmC in cfDNA may serve as a valuable biomarker for early-stage invasive breast cancer screening. However, in order to develop reliable 5hmC-based biomarkers, a larger cohort and comprehensive pairwise comparisons are essential to identify common 5hmC features in cfDNA and primary breast tumors.

## Conclusions

Taken together, the dynamic changes of DNA 5hmC and 5mC in breast lesions and their effects on transcriptional regulation are crucial in propelling the malignant transformation of breast tumors. TET2-related DNA demethylation, histone marks, and TFs can be orchestrated in promoting breast tumorigenesis through transcriptional regulation. In addition, 5hmC-based biomarkers are valuable and remain to be investigated in the screening of breast precancerous lesions and liquid biopsy of breast cancer.

## Abbreviations

UDH	Usual ductal hyperplasia
ADH/ALH	Atypical ductal/lobular hyperplasia
DCIS/LCIS	Ductal/lobular carcinoma in situ
IDC	Invasive ductal breast cancer
5mC	5-Methylcytosine
5hmC	5-Hydroxymethylcytosine
cfDNA	Cell-free DNA
TF	Transcriptional factor
TDLU	Terminal duct lobular unit
LC-MS/MS	Liquid chromatography–mass spectrometry
hMeDIP-seq	Hydroxymethylated DNA immunoprecipitation sequencing
TSS	Transcription start site
DhMR	Differentially hydroxymethylated region
DMG	Differentially methylated gene
DMR	Differentially methylated region
DEG	Differentially expressed gene
MeDIP-seq	Methylated DNA immunoprecipitation sequencing
ChIP-seq	Chromatin immunoprecipitation followed by sequencing
IHC	Immunohistochemical staining
H & E	Hematoxylin and eosin staining
hMR	Hydroxymethylated methylated region
MR	Methylated region
hMG	Hydroxymethylated methylated gene
MG	Methylated gene
DhMG	Differentially 5-hydroxymethylated gene
GSEA	Gene set enrichment analysis
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
TES	Transcription ending site
unhMG	Genes without 5hmC modification

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13148-024-01726-7>.

Additional file 1. Representative movie displaying the process of target ducts/cells' macro-dissection.

Additional file 2: Figure S1. Genomic characteristics of 5hmC change in breast tumorigenesis. **a** Distribution of 5hmC peaks across each genomic region in UDH, ADH, DCIS, and IDC. **b** Distribution of normalized 5hmC read counts across the gene bodies. **c** Boxplots showing relative levels of 5hmC peak located at different genomic regions in UDH, ADH, DCIS, and IDC samples. **d** KEGG enrichment analysis of the differentially hydroxymethylated genes in each phase of breast tumorigenesis. **e** GSEA results of the DhMGs in each phase of breast tumorigenesis. NES > 1, Q value < 0.05. \*\*\*P < 0.001. Table S1. Detailed GSEA results of DhMGs in each phase of breast tumorigenesis. Figure S2. Genomic characteristics of 5mC changes along with 5hmC in breast tumorigenesis. **a** Distribution of 5mC peaks across each genomic region in UDH, ADH, and DCIS. **b** Distribution of normalized 5mC read counts across the gene bodies in early-stage breast tumors. **c** Boxplots showing relative 5mC levels of different genomic regions among UDH, ADH, and DCIS. **d** Numbers of significant differentially methylated genes identified in each phase of breast tumorigenesis. **e** KEGG enrichment analysis of the DMGs in the early stage of breast tumorigenesis. **f** GSEA results of the DMGs in each phase of breast tumorigenesis. NES > 1, Q value < 0.05. **g** Scatterplots displaying the regions co-occurring 5hmC and 5mC changes, blot spots indicate the regions with significant 5hmC and 5mC changes. \*\*\*P < 0.001. Table S2. Detailed GSEA results of DMGs in each phase of breast tumorigenesis (NES > 1, Q value < 0.05). Figure S3. Correlation between DNA epigenetic modifications and transcriptional regulation. **a** Number of differentially expressed genes in each phase of breast tumorigenesis through re-analyzing RNA-seq data in GSE47462. **b**, **c** KEGG enrichment analysis of the DEGs accompanied with 5hmC and 5mC change in each phase of breast tumorigenesis. **d**, **e** Distribution profiles of 5hmC and 5mC peak in genes expressed at high, medium and low levels. Figure S4. Transcriptional factors and histone modifications are involved in TET2-related DhMRs and DMRs' regulation in breast tumorigenesis. **a** Distribution of TET2-binding regions across each genomic region in MCF7 cells. **b** Enrichment of H3K27ac, H3K27me3, H3K36me3, H3K4me1, H3K4me3, and H3K9me3 around TET2-binding regions. **c** Effect of TET2-binding and DNA 5-hydroxymethylation on gene expression in UDH1, UDH2, ADH2, ADH3, DCIS2. **d**, **e** The number of DhMRs and DMRs overlapped with TET2-binding regions in each phase of breast tumorigenesis. **f** Enrichment of H3K27ac, H3K27me3, H3K36me3, H3K4me1, H3K4me3, and H3K9me3 around TET2-binding hypo-DMRs and hyper-DMRs in the early phases of breast tumorigenesis. **g** Enrichment of H3K27ac, H3K27me3, H3K36me3, H3K4me1, H3K4me3, and H3K9me3 around TF-binding sites including ESR1, GATA3, FOXA1, FOS, FOSL2, FOXM1, JUNB in MCF7 cells. **h** Enrichment of the crucial TFs identified by motif analyses around nonTET2-binding hypo-DhMRs and hyper-DhMRs. **i** Enrichment of the crucial TFs around TET2-binding hypo-DMRs and hyper-DMRs in breast tumorigenesis. **j** Enrichment of the crucial TFs around nonTET2-binding hypo-DMRs and hyper-DMRs. \* P < 0.05, \*\*P < 0.01, \*\*\*P < 0.001, ns: no significant. nTET2-hMGs: Hydroxymethylated genes without TET2-binding regions; nTET2-unhMGs: Genes without 5hmC modifications and TET2-binding regions; TET2-hMGs: Hydroxymethylated genes with TET2-binding regions; TET2-unhMGs: Genes with TET2 binding regions but without 5hmC modifications. Figure S5. Characteristics of 5hmC in cfDNA and primary breast tumors. **a** Boxplots showing global 5hmC level in cfDNA of healthy controls and breast cancer patients. **b** Boxplots showing global 5hmC level in breast tumors of UDH and IDC. **c**, **d** Distribution of differentially hydroxymethylated regions across genomic regions in cfDNA and breast cancer tissues. \*\*\*P < 0.001.

Additional file 3: Table S3. Top 500 hyper/hypo-DhMRs and annotated genes in different phases of breast tumorigenesis.

Additional file 4: Table S4. Top 500 hyper/hypo-DMRs and annotated genes in different phases of breast tumorigenesis.

## Acknowledgements

Not applicable.

## Author contributions

SLW performed clinical samples collection, and experiments. LY and CH performed bioinformatic analysis. FY and YZ performed LC-MS/MS detection. QL and CM provided experimental support. XW provided support in bioinformatic analysis. SLW and YL wrote the manuscript. YN supervised the project and wrote the manuscript. FJ and WMT conceived and initiated the project. All the authors have read and approved the final manuscript.

## Funding

This work was supported by the National Natural Science Foundation of China (grant numbers 82073282 to FJ, and 82303129 to SLW), the CAMS Innovation Fund for Medical Sciences (Grant Number 2021-I2M-1-020 to YN), and the Fundamental Research Funds for the Central Universities (Grant Number 3332022037 to CM).

## Availability of data and materials

The datasets have been deposited in the Genome Sequence Archive under the accession number GSA-Human: HRA005156 (bioProject accession: PRJCA018482).

## Declarations

### Ethics approval and consent to participate

All the human tissue samples were obtained under a protocol (AF-SOP-07-1.1-0.1) approved by the Medical Scientific Research Ethics Committee of the First Affiliated Hospital of China Medical University. All subjects provided written informed consent according to the Institutional Guidelines. This study is compliant with all relevant ethical regulations regarding research involving human resources.

### Consent for publication

All authors have given consent for publication.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Department of Surgical Oncology and Breast Surgery, The First Affiliated Hospital of China Medical University, Shenyang 110000, China. <sup>2</sup>Department of Pathology, Institute of Basic Medical Sciences Chinese Academy of Medical Sciences, School of Basic Medicine Peking Union Medical College, Beijing 100005, China. <sup>3</sup>State Key Laboratory of Medical Molecular Biology, Department of Biochemistry and Center for Bioinformatics, Institute of Basic Medical Sciences Chinese Academy of Medical Sciences, School of Basic Medicine Peking Union Medical College, Beijing 100005, China. <sup>4</sup>National Institute of Measurement and Testing Technology, Chengdu 610021, China. <sup>5</sup>Beijing National Laboratory for Molecular Sciences (BNLMS), MOE Key Laboratory of Bioorganic Chemistry and Molecular Engineering, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China. <sup>6</sup>State Key Laboratory of Common Mechanism Research for Major Diseases, Center for Bioinformatics, National Infrastructures for Translational Medicine, Institute of Clinical Medicine and Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100871, China. <sup>7</sup>Molecular Pathology Research Center, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China.

Received: 4 May 2024 Accepted: 9 August 2024

Published online: 11 September 2024

## References

1. Fan L, Strasser-Weippl K, Li JJ, St Louis J, Finkelstein DM, Yu KD, Chen WQ, Shao ZM, Goss PE. Breast cancer in China. *Lancet Oncol*. 2014;15(7):e279-289.

2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin*. 2015;65(2):87–108.
3. Dupont WD, Page DL. Risk factors for breast cancer in women with proliferative breast disease. *N Engl J Med*. 1985;312(3):146–51.
4. Wellings SR, Jensen HM, Marcum RG. An atlas of subgross pathology of the human breast with special reference to possible precancerous lesions. *J Natl Cancer Inst*. 1975;55(2):231–73.
5. Wellings SR, Jensen HM. On the origin and progression of ductal carcinoma in the human breast. *J Natl Cancer Inst*. 1973;50(5):1111–8.
6. Oyama T, Iijima K, Takei H, Horiguchi J, Iino Y, Nakajima T, Koerner F. Atypical cystic lobule of the breast: an early stage of low-grade ductal carcinoma in-situ. *Breast Cancer*. 2000;7(4):326–31.
7. Page DL, Dupont WD, Rogers LW, Rados MS. Atypical hyperplastic lesions of the female breast. A long-term follow-up study. *Cancer*. 1985;55(11):2698–708.
8. Kader T, Hill P, Zethoven M, Goode DL, Elder K, Thio N, Doyle M, Semple T, Sufyan W, Byrne DJ, et al. Atypical ductal hyperplasia is a multipotent precursor of breast carcinoma. *J Pathol*. 2019;248(3):326–38.
9. Thomas PS. Diagnosis and management of high-risk breast lesions. *J Natl Compr Canc Netw*. 2018;16(11):1391–6.
10. Pashayan N, Pharoah PDP. The challenge of early detection in cancer. *Science*. 2020;368(6491):589–90.
11. Song CX, He C. Balance of DNA methylation and demethylation in cancer development. *Genome Biol*. 2012;13(10):173.
12. Skvortsova K, Stirzaker C, Taberlay P. The DNA methylation landscape in cancer. *Essays Biochem*. 2019;63(6):797–811.
13. Lian CG, Xu Y, Ceol C, Wu F, Larson A, Dresser K, Xu W, Tan L, Hu Y, Zhan Q, et al. Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. *Cell*. 2012;150(6):1135–46.
14. Chen K, Zhang J, Guo Z, Ma Q, Xu Z, Zhou Y, Xu Z, Li Z, Liu Y, Ye X, et al. Loss of 5-hydroxymethylcytosine is linked to gene body hypermethylation in kidney cancer. *Cell Res*. 2016;26(1):103–18.
15. Cai J, Chen L, Zhang Z, Zhang X, Lu X, Liu W, Shi G, Ge Y, Gao P, Yang Y, et al. Genome-wide mapping of 5-hydroxymethylcytosines in circulating cell-free DNA as a non-invasive approach for early detection of hepatocellular carcinoma. *Gut*. 2019;68(12):2195–205.
16. Liu J, Jiang J, Mo J, Liu D, Cao D, Wang H, He Y, Wang H. Global DNA 5-hydroxymethylcytosine and 5-formylcytosine contents are decreased in the early stage of hepatocellular carcinoma. *Hepatology*. 2019;69(1):196–208.
17. Liu H, Xu T, Cheng Y, Jin MH, Chang MY, Shu Q, Allen EG, Jin P, Wang X. Altered 5-hydroxymethylcytosine landscape in primary gastric adenocarcinoma. *DNA Cell Biol*. 2019;38(12):1460–9.
18. Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. *Cell*. 2012;150(1):12–27.
19. Thomson JP, Meehan RR. The application of genome-wide 5-hydroxymethylcytosine studies in cancer research. *Epigenomics*. 2017;9(1):77–91.
20. Zeng C, Stroup EK, Zhang Z, Chiu BC, Zhang W. Towards precision medicine: advances in 5-hydroxymethylcytosine cancer biomarker discovery in liquid biopsy. *Cancer Commun*. 2019;39(1):12.
21. Tang Q, Cheng J, Cao X, Surowy H, Burwinkel B. Blood-based DNA methylation as biomarker for breast cancer: a systematic review. *Clin Epigenet*. 2016;8:115.
22. Lo PK, Sukumar S. Epigenomics and breast cancer. *Pharmacogenomics*. 2008;9(12):1879–902.
23. Wilkins OM, Johnson KC, Houseman EA, King JE, Marsit CJ, Christensen BC. Genome-wide characterization of cytosine-specific 5-hydroxymethylation in normal breast tissue. *Epigenetics*. 2020;15(4):398–418.
24. Ramasamy D, Rao A, Balaiah M, Vittal Rangan A, Sundersingh S, Veluswami S, Thangarajan R, Mani S. Locus-specific enrichment analysis of 5-hydroxymethylcytosine reveals novel genes associated with breast carcinogenesis. *Cells*. 2022;11(19):2939.
25. Tsai KW, Li GC, Chen CH, Yeh MH, Huang JS, Tseng HH, Fu TY, Liou HH, Pan HW, Huang SF, et al. Reduction of global 5-hydroxymethylcytosine is a poor prognostic factor in breast cancer patients, especially for an ER/PR-negative subtype. *Breast Cancer Res Treat*. 2015;153(1):219–34.
26. Yang H, Liu Y, Bai F, Zhang JY, Ma SH, Liu J, Xu ZD, Zhu HG, Ling ZQ, Ye D, et al. Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. *Oncogene*. 2013;32(5):663–9.
27. Yang Z, Jiang W, Liu F, Zhou Y, Yin H, Ai S. A novel electro-chemical immunosensor for the quantitative detection of 5-hydroxymethylcytosine in genomic DNA of breast cancer tissue. *Chem Commun*. 2015;51(78):14671–3.
28. Zhang Z, Jin Y, Zhang W, Chu C, Zhang K, Gao X, Zhou J, Zou L, Tang F, Wang H, et al. Values of 5mC, 5hmC, and TET2 for identifying the presence and progression of breast precancerous lesion. *J Clin Lab Anal*. 2019;34(5):e23162.
29. Wu SL, Zhang X, Chang M, Huang C, Qian J, Li Q, Yuan F, Sun L, Yu X, Cui X, et al. Genome-wide 5-hydroxymethylcytosine profiling analysis identifies MAP7D1 as a novel regulator of lymph node metastasis in breast cancer. *Genom Proteomics Bioinform*. 2021;19(1):64–79.
30. Salmeron-Barcenas EG, Zacapala-Gomez AE, Torres-Rojas FI, Antonio-Vejar V, Avila-Lopez PA, Banos-Hernandez CJ, Nunez-Martinez HN, Dircio-Maldonado R, Martinez-Carrillo DN, Ortiz-Ortiz J, et al. TET enzymes and 5hmC levels in carcinogenesis and progression of breast cancer: potential therapeutic targets. *Int J Mol Sci*. 2023;25(1):272.
31. Park SY, Kwon HJ, Lee HE, Ryu HS, Kim SW, Kim JH, Kim IA, Jung N, Cho NY, Kang GH. Promoter CpG island hypermethylation during breast cancer progression. *Virchows Arch*. 2011;458(1):73–84.
32. Hoque MO, Prencipe M, Poeta ML, Barabano R, Valori VM, Copetti M, Gallo AP, Brait M, Maiello E, Apicella A, et al. Changes in CpG islands promoter methylation patterns during ductal breast carcinoma progression. *Cancer Epidemiol Biomark Prev*. 2009;18(10):2694–700.
33. Danforth DN. Molecular profile of atypical hyperplasia of the breast. *Breast Cancer Res Treat*. 2018;167(1):9–29.
34. Fleischer T, Frigessi A, Johnson KC, Edvardsen H, Touleimat N, Klajic J, Riis ML, Haakensen VD, Warnberg F, Naume B, et al. Genome-wide DNA methylation profiles in progression to in situ and invasive carcinoma of the breast with impact on gene transcription and prognosis. *Genome Biol*. 2014;15(8):435.
35. Minning C, Mokhtar NM, Abdullah N, Muhammad R, Emran NA, Ali SA, Harun R, Jamal R. Exploring breast carcinogenesis through integrative genomics and epigenomics analyses. *Int J Oncol*. 2014;45(5):1959–68.
36. Teschendorff AE, Gao Y, Jones A, Ruebner M, Beckmann MW, Wächter DL, Fasching PA, Widschwendter M. DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. *Nat Commun*. 2016;7:10478.
37. Muse ME, Titus AJ, Salas LA, Wilkins OM, Mullen C, Gregory KJ, Schneider SS, Crisi GM, Jawale RM, Otis CN, et al. Enrichment of CpG island shore region hypermethylation in epigenetic breast field cancerization. *Epigenetics*. 2020;15(10):1093–106.
38. Yu Y, Pan HY, Zheng X, Yuan F, Zhou YL, Zhang XX. Ultrasensitive simultaneous detection of multiple rare modified nucleosides as promising biomarkers in low-pur breast cancer DNA samples for clinical multi-dimensional diagnosis. *Molecules*. 2022;27(20):7041.
39. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, Imamichi T, Chang W. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res*. 2022;50(W1):W216–21.
40. Ma C, Chang M, Lv H, Zhang ZW, Zhang W, He X, Wu G, Zhao S, Zhang Y, Wang D, et al. RNA m(6)A methylation participates in regulation of postnatal development of the mouse cerebellum. *Genome Biol*. 2018;19(1):68.
41. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPP. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
42. Brunner AL, Li J, Guo X, Sweeney RT, Varma S, Zhu SX, Li R, Tibshirani R, West RB. A shared transcriptional program in early breast neoplasias despite genetic and clinical distinctions. *Genome Biol*. 2014;15(5):R71.
43. Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, Brown GD, Gojis O, Ellis IO, Green AR, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*. 2012;481(7381):389–93.
44. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
45. Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*. 2015;31(14):2382–3.
46. Broome R, Chernukhin I, Jamieson S, Kishore K, Papachristou EK, Mao SQ, Tejado CG, Mahtey A, Theodorou V, Groen AJ, et al. TET2 is a component

- of the estrogen receptor complex and controls 5mC to 5hmC conversion at estrogen receptor cis-regulatory regions. *Cell Rep.* 2021;34(8):108776.
47. Luo Y, Hitz BC, Gabdank I, Hilton JA, Kagda MS, Lam B, Myers Z, Sud P, Jou J, Lin K, et al. New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res.* 2019;48(D1):D882–9.
  48. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell.* 2010;38(4):576–89.
  49. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 2016;44(W1):W160–5.
  50. Song C-X, Yin S, Ma L, Wheeler A, Chen Y, Zhang Y, Liu B, Xiong J, Zhang W, Hu J, et al. 5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages. *Cell Res.* 2017;27(10):1231–42.
  51. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. Software for computing and annotating genomic ranges. *PLoS Comput Biol.* 2013;9(8):e1003118.
  52. de Ruijter TC, van der Heide F, Smits KM, Aarts MJ, van Engeland M, Heijnen VCG. Prognostic DNA methylation markers for hormone receptor breast cancer: a systematic review. *Breast Cancer Res.* 2020;22(1):13.
  53. Luo Y, Hitz BC, Gabdank I, Hilton JA, Kagda MS, Lam B, Myers Z, Sud P, Jou J, Lin K, et al. New developments on the encyclopedia of DNA elements (ENCODE) data portal. *Nucleic Acids Res.* 2020;48(D1):D882–9.
  54. Ignatiadis M, Sledge GW, Jeffrey SS. Liquid biopsy enters the clinic—implementation issues and future challenges. *Nat Rev Clin Oncol.* 2021;18(5):297–312.
  55. Li W, Zhang X, Lu X, You L, Song Y, Luo Z, Zhang J, Nie J, Zheng W, Xu D, et al. 5-Hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic biomarkers for human cancers. *Cell Res.* 2017;27(10):1243–57.
  56. Guler GD, Ning Y, Ku CJ, Phillips T, McCarthy E, Ellison CK, Bergamaschi A, Collin F, Lloyd P, Scott A, et al. Detection of early stage pancreatic cancer using 5-hydroxymethylcytosine signatures in circulating cell free DNA. *Nat Commun.* 2020;11(1):5270.
  57. Li Y, Li Y, Wedren S, Li G, Charn TH, Desai KV, Bonnard C, Czene K, Humphreys K, Darabi H, et al. Genetic variation of ESR1 and its co-activator PPAR $\gamma$ C1B is synergistic in augmenting the risk of estrogen receptor-positive breast cancer. *Breast Cancer Res.* 2011;13(1):R10.
  58. Padua MB, Bhat-Nakshatri P, Anjanappa M, Prasad MS, Hao Y, Rao X, Liu S, Wan J, Liu Y, McElyea K, et al. Dependence receptor UNC5A restricts luminal to basal breast cancer plasticity and metastasis. *Breast Cancer Res.* 2018;20(1):35.
  59. de Heer EC, Jalving M, Harris AL. HIFs, angiogenesis, and metabolism: elusive enemies in breast cancer. *J Clin Invest.* 2020;130(10):5074–87.
  60. Rose NR, Klose RJ. Understanding the relationship between DNA methylation and histone lysine methylation. *Biochim Biophys Acta.* 2014;1839(12):1362–72.
  61. Agirre E, Oldfield AJ, Bellora N, Segelle A, Luco RF. Splicing-associated chromatin signatures: a combinatorial and position-dependent role for histone marks in splicing definition. *Nat Commun.* 2021;12(1):682.
  62. Bhattacharyya S, Pradhan K, Campbell N, Mazdo J, Vasantkumar A, Maqbool S, Bhagat TD, Gupta S, Suzuki M, Yu Y, et al. Altered hydroxy-methylation is seen at regulatory regions in pancreatic cancer and regulates oncogenic pathways. *Genome Res.* 2017;27(11):1830–42.
  63. He B, Zhang C, Zhang X, Fan Y, Zeng H, Liu J, Meng H, Bai D, Peng J, Zhang Q, et al. Tissue-specific 5-hydroxymethylcytosine landscape of the human genome. *Nat Commun.* 2021;12(1):4249.
  64. Hu Z, Li Z, Ma Z, Curtis C. Multi-cancer analysis of clonality and the timing of systemic spread in paired primary tumors and metastases. *Nat Genet.* 2020;52(7):701–8.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.