Hemimethylation Patterns in Breast Cancer Cell Lines

Cancer Informatics
Volume 18: 1–12
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1176935119872959

Shuying Sun[1], Yu Ri Lee[1] and Brittany Enfield[2]

[1]Department of Mathematics, Texas State University, San Marcos, TX, USA. [2]Global Engineering Systems, Cypress Semiconductor, Austin, TX, USA.

ABSTRACT: DNA methylation is an epigenetic event that involves adding a methyl group to the cytosine (C) site, especially the one that pairs with a guanine (G) site (ie, CG or CpG site), in a human genome. This event plays an important role in both cancerous and normal cell development. Previous studies often assume symmetric methylation on both DNA strands. However, asymmetric methylation, or hemimethylation (methylation that occurs only on 1 DNA strand), does exist and has been reported in several studies. Due to the limitation of previous DNA methylation sequencing technologies, researchers could only study hemimethylation on specific genes, but the overall genomic hemimethylation landscape remains relatively unexplored. With the development of advanced next-generation sequencing techniques, it is now possible to measure methylation levels on both forward and reverse strands at all CpG sites in an entire genome. Analyzing hemimethylation patterns may potentially reveal regions related to undergoing tumor growth. For our research, we first identify hemimethylated CpG sites in breast cancer cell lines using Wilcoxon signed rank tests. We then identify hemimethylation patterns by grouping consecutive hemimethylated CpG sites based on their methylation states, methylation "M" or unmethylation "U." These patterns include regular (or consecutive) hemimethylation clusters (eg, "MMM" on one strand and "UUU" on another strand) and polarity (or reverse) clusters (eg, "MU" on one strand and "UM" on another strand). Our results reveal that most hemimethylation clusters are the polarity type, and hemimethylation does occur across the entire genome with notably higher numbers in the breast cancer cell lines. The lengths or sizes of most hemimethylation clusters are very short, often less than 50 base pairs. After mapping hemimethylation clusters and sites to corresponding genes, we study the functions of these genes and find that several of the highly hemimethylated genes may influence tumor growth or suppression. These genes may also indicate a progressing transition to a new tumor stage.

KEYWORDS: Methylation, hemimethylation, bioinformatics, breast cancer

RECEIVED: August 5, 2019. **ACCEPTED:** August 5, 2019.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project was also supported by Dr. Sun's Texas State University Research Enhancement Program (REP) grant.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Shuying Sun, Department of Mathematics, Texas State University, 601 University Drive, San Marcos, TX 78666, USA. Email: ssun@txstate.edu

Introduction

Cancer is a leading cause of death. In 2019, there will be an estimated 1 762 450 new cancer cases diagnosed and 606 880 cancer deaths in the United States, according to the American Cancer Society Cancer Facts and Figures 2019. Among different types of cancer, about 62 930 new cases of breast carcinoma in situ are expected to be diagnosed in 2019.[1] Medical researchers have developed many methods to fight this disease, and early detection has become a key factor. A Surveillance, Epidemiology, and End Results (SEER) review of breast cancer cases in the United State has found that patients diagnosed in the earlier stages of the disease have a significantly higher chance of survival.[2] Common detection methods like self or clinical breast exams and mammograms are fairly successful means of detecting tumors once they have developed. However, by the time a tumor has grown large enough to be identified by these methods, the patient may already be in a late stage of the disease. Thus, methods of detecting breast cancer at earlier stages have become progressively important. Genetic and epigenetic biomarkers are especially important and may serve as early indicators of tumor growth in many cancers including breast cancer.

Epigenetics is a relatively new field of biology that examines how gene activity is affected by external modifications to DNA and not by changes to the DNA sequence itself.[3] One significant epigenetic mechanism is DNA methylation, a biological process in which a methyl group (CH_3) is added to the fifth carbon of a cytosine-guanine dinucleotide (CG or CpG site) on a DNA molecule in a mammalian cell. It is a natural and essential process associated with DNA replication and cell differentiation. DNA methylation is facilitated by a group of DNA methyltransferases (DNMTs) and comes in 2 forms: de novo methylation and maintenance methylation.[3] In de novo methylation, bare DNA is methylated in a tissue-specific pattern as shown in Figure 1A. During maintenance methylation, the new DNA strand formed during DNA replication becomes methylated at the same CpG sites as the first strand,[4] as shown in Figure 1B.

Previous researchers have used microarray technologies to study DNA methylation. Although these technologies allow researchers to study expression levels of numerous genes, all microarray technologies (except for the Illumina microarray) cannot generate methylation signals at the single CpG site level. In addition, microarray technologies lack the ability to separate strands of DNA. As a result of these weaknesses, many early studies assume symmetric methylation of CpG sites, or methylation of a CpG site on both the forward and reverse strands.[5,6] Within the last decade, next-generation sequencing (NGS) technologies have filled these gaps by combining bisulfite treatment with parallel DNA sequencing processes.[5-7] On a single strand of DNA, with bisulfite treatment, unmethylated cytosines are converted to uracils while methylated

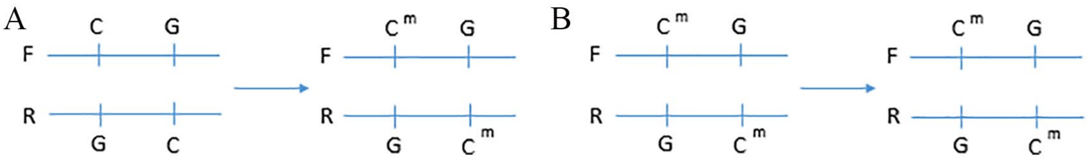

Figure 1. Two types of methylation: (A) Example of de novo methylation; (B) Example of maintenance methylation. "F" and "R" mean forward (or "+") and reverse (or "−") stands, respectively. "CG" on the F and R strands means a CG site is not methylated; "C^{m}G" on the F or R strand means a CG site is methylated.

cytosines remain unchanged. After amplification, the modified DNA is sequenced with one of several advanced sequencing machines available today.[8] We can then obtain the methylation signals on the forward and reverse strands allowing us to investigate asymmetric methylation, or hemimethylation (HM), which does exist and has been reported in several studies.[9-13] To clarify, we emphasize that HM in this article means that DNA methylation at a CpG site only occurs on one strand, not on the other strand. It is not allele-specific methylation that is common in imprinting, and it is not partial methylation either.

Shao et al[10] report the existence of hemimethylated CpG sites in both carcinomas and controls when studying CpG sites of the Sat2 gene. They show that HM exists in both singletons and clusters. Singletons are the hemimethylated CpG sites that are not formed a cluster with other consecutive CpG sites. Hemimethylation is significantly more likely to occur in clusters in ovarian cancerous cells than in control or normal cells. The HM clusters reported in their previous studies exist in 2 forms: regular (or consecutive) and polarity (or reverse) patterns.[9,10] Regular (or consecutive) HM clusters occur when successive CpG sites are hemimethylated on the same strand, either the forward or the reverse strand as shown in Figure 2A. For example, "MMM-UUU" is an HM cluster of 3 CpG sites with "MMM" on the forward/F strand and "UUU" on the reverse/R strand; "MM-UU" is an HM cluster of 2 CpG sites with "MM" on the forward/F strand and "UU" on the reverse/R strand. On the contrary, polarity (or reverse) clusters arise when consecutive CpG sites are hemimethylated on opposite strands as shown in Figure 2B. For example, "MU-UM" is a polarity HM cluster of 2 CpG sites with "MU" on the forward/F strand and "UM" on the reverse/R strand; "UM-MU" is an HM cluster of 2 CpG sites with "UM" on the forward/F strand and "MU" on the reverse/R strand. Both these 2 types of HM clusters have been found and reported in literature.[9,10] HM clusters may indicate different methylation events and may have a more substantial impact on the function of a particular gene. In addition to HM clusters, there are also individual or singleton HM CpG sites that are not in a cluster with other CpG sites.[10,12]

It is both useful and important to study HM. First, a recent study by Xu and Corces shows that hemimethylated CpG sites are inherited over several cell divisions.[11,13] This finding challenges the previous understanding and model of methylation and HM, in which HM was thought to be transient. Second, the identification of HM patterns is crucial for understanding different methylation events (eg, methylation maintenance and de novo methylation) and the establishment of different methylation patterns (eg, hypomethylation and hypermethylation).[12] For example, previous studies suggest that hemimethylated CpG sites are intermediates in active demethylation during carcinogenesis and not just due to a failure of maintenance methylation during replicative DNA synthesis. Hemimethylation can help researchers trace the footprints of DNA methylation in cancer.[9,10] Third, a recent publication shows that stably inherited HM regulates chromatin interaction and transcription.[13] Therefore, it is very likely that HM affects gene expression in cancer cells significantly. Before studying the function or role of HM patterns in cancers, we should first identify them in a whole genome. Therefore, the identification of hemimethylated sites is the focus of our current article.

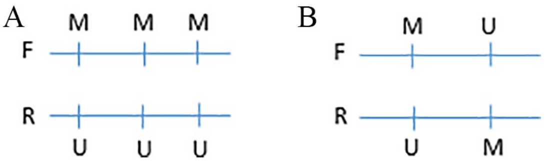

Figure 2. Two hemimethylation clusters (regular and polarity clusters). "F" and "R" mean forward (or "+") and reverse (or "−") stands, respectively. "M" means methylation and "U" means unmethylation.

The goal of our research is to use publicly available bisulfite-sequencing data to study HM in breast cancer cell lines across the entire genome. The HM sites or patterns we want to identify are singleton HM CpG sites and 2 types of HM clusters as shown in Figure 2. We obtain publicly available reduced representation bisulfite-sequencing (RRBS) data (GSE27003) of 7 breast cancer cell lines (BT20, BT474, MCF7, MDAMB231, MDAMB468, T47D, and ZR751).[14] We then identify hemimethylated CpG sites using Wilcoxon signed rank tests and study the genes and promoters that contain these CpG sites.

Methodology

Data preparation

We use the hg19 version of the human genome as a reference to align raw sequencing reads. All data sets in our project have been processed and analyzed using publicly available software packages: BRAT-bw,[15] Perl,[16,17] and R.[18] The preprocessed methylation sequencing data sets consist of all CpG base pairs found on the forward and reverse strand for each cell line. In total, there are 27 999 103 applicable CpG sites in the whole genome. For these 28 million CpG sites, we choose to further

Table 1. HM CpG sites and percentage of HM sites that form clusters.

MEAN DIFFERENCE	HM SITES	HM SITES IN CLUSTERS	PERCENTAGE
|Mean difference| ⩾ 0.4	19 736	3492	17.69%
|Mean difference| ⩾ 0.6	15 526	2526	16.27%
|Mean difference| ⩾ 0.8	10 136	1382	13.63%

Abbreviation: HM, hemimethylation.
The first column is the mean difference cutoff values. The second column shows the total number of identified HM CpG sites. The third column is the number of HM sites that form or belong to a cluster with at least 2 consecutive HM sites.

analyze those with at least 4 methylation signals among 7 cell lines on each strand (forward and reverse). There are 464 674 CpG sites (ie, 1.6% of all CpG sites in a human genome) passing these criteria.

Statistical and bioinformatic analysis

To identify HM CpG sites, we analyze the RRBS data of 7 breast cancer cell lines using the Wilcoxon signed rank test. This statistical test is used when the normality (normal distribution) assumption is violated and the sample size is small. It is often used to determine whether the centers of 2 sets of data are significantly different from each other. To ensure quality data, Wilcoxon signed rank tests are performed on CpG sites that have at most 3 missing observation values in the cancer forward and reverse data. At each CpG site, we use Wilcoxon signed rank tests to compare the methylation signals of the forward and reverse strands of 7 breast cancer cell lines. The Wilcoxon test is a rank-based test, and there are different methods of dealing with tied data and zero values including the "Wilcoxon" and "Pratt" methods.[19] The "Pratt" method first ranks absolute differences including zeros and then discards all the ranks corresponding to the zero-differences, whereas the "Wilcoxon" method first deals with the zero-differences and then ranks the remaining absolute differences. The "Pratt" and "Wilcoxon" methods produce similar results for all CpG sites and the same results for CpG sites with P value < .05 for our data. For our study, we choose to use "Wilcoxon" to deal with tied data. Once we get the output of the Wilcoxon signed rank tests and the P value for each CpG site, we calculate and report the absolute mean difference of the forward and the reverse strand methylation signals.

We use the mean difference and P value to select HM CpG sites. The mean difference at each CpG site is the difference between the mean/average methylation signals of the forward and reverse strands. It tells us whether the cancer cell lines show biologically significant HM signals at a CpG site. We look for CpG sites whose absolute mean differences are greater than a cutoff value in cancer cell lines. The P value will tell us whether a CpG site has statistically significant HM signals. As we are looking for CpG sites that are hemimethylated in cancer, we select CpG sites with P value < .05 and forward and reverse strand methylation mean difference (absolute value) $d \geqslant 0.4$, 0.6, and 0.8. Then, we identify HM patterns that meet both biological and statistical significance criteria by extracting consecutive HM CpG sites based on their methylation states and further study these HM patterns.

For bioinformatic analysis, we provide annotations for all HM CpG sites by finding which genes have HM sites in their gene body or promoter regions using the R code written by ourselves. We then use the GeneCards (a gene database)[20] and the Molecular Signatures Database (MSigDB)[21] to study these genes' functions. We use the ConsensusPathDB (CPDB)[22-25] to conduct pathway analyses. More detailed information about these databases and related results will be shown in the Results section.

Results

Hemimethylated CpG sites and clusters

Using Wilcoxon signed rank tests on cancer cell lines and applying stringent cutoff values, we have identified HM CpG sites as those with P value < .05 and an absolute mean difference $d \geqslant 0.4$, 0.6, and 0.8 (see Table 1). This table shows the summary for the number and percentage of HM CpG sites that form clusters. The HM CpG sites that are not in a cluster are called singletons. Table 1 shows that the number of HM CpG sites belonging to clusters decreases as the mean difference cutoff value increases. When the cutoff value increases, the number of HM clusters decreases too (see Table 2). However, for the 3 cutoff values, the percentages of HM CpG sites belonging to clusters are relatively consistent (about 13%-17% as shown in Table 1). This consistency implies that breast cancer samples may contain a certain number of HM clusters. A more detailed summary of 2 types of HM clusters (regular and polarity clusters) is shown in Table 2. Table 2 shows the number and percentage of regular HM clusters and polarity clusters.

Although we have applied different cutoff values, we will conduct a more detailed analysis for HM sites obtained based on absolute value of mean difference $d \geqslant 0.4$. Next, we zoom in to summarize the HM clusters with the forward and reverse strand methylation mean difference (absolute value) $d \geqslant 0.4$ (see Table 3). Among the total 1719 clusters in Table 3, 1558 are polarity clusters and 161 are regular clusters. Among the 1558 polarity clusters, 1534 are MU-UM and 24 are UM-MU

Table 2. Summary of hemimethylation clusters.

IMEAN DIFFERENCEI⩾0.4			IMEAN DIFFERENCEI⩾0.6			IMEAN DIFFERENCEI⩾0.8		
REGULAR HM CLUSTER			REGULAR HM CLUSTER			REGULAR HM CLUSTER		
MM-UU	UU-MM	LENGTH > 2	MM-UU	UU-MM	LENGTH > 2	MM-UU	UU-MM	LENGTH > 2
69	56	36	42	34	12	15	10	6
42.86%	34.78%	22.36%	47.73%	38.64%	13.64%	48.39%	32.26%	19.35%
IMEAN DIFFERENCEI⩾0.4			IMEAN DIFFERENCEI⩾0.6			IMEAN DIFFERENCEI⩾0.8		
POLARITY CLUSTER			POLARITY CLUSTER			POLARITY CLUSTER		
MU-UM	UM-MU		MU-UM	UM-MU		MU-UM	UM-MU	
1534	24		1160	7		652	4	
98.46%	1.54%		99.40%	0.60%		99.39%	0.61%	

Abbreviation: HM, hemimethylation.
The Table shows regular HM clusters with 2 or more CpG sites and polarity clusters with only 2 CpG sites. The number of clusters and percentage of each type are calculated based on each mean difference cutoff value.

Table 3. Frequency or count of HM clusters with $d \geqslant 0.4$.

HM CLUSTERS	FREQUENCY/COUNT	POLARITY OR NOT
MMMMM-UUUUU	1	Regular
MMMM-UUUU	3	Regular
MMM-UUU	12	Regular
MMMU-UUUM	1	Regular
MM-UU	69	Regular
MMU-UUM	2	Regular*
MUM-UMU	1	Regular*
MU-UM	1534	Polarity
MUU-UMM	1	Regular*
UM-MU	24	Polarity
UMU-MUM	1	Regular*
UU-MM	56	Regular
UUU-MMM	7	Regular
UUUU-MMMM	4	Regular
UUUUU-MMMMM	2	Regular
UUUUUUU-MMMMMMM	1	Regular
	Total: 1719	

Abbreviation: HM, hemimethylation.
The first and second columns of Table 3 are cluster patterns and counts. The third column indicates whether an HM pattern is a regular or polarity cluster. Four regular clusters are labeled as "Regular*." These clusters are classified as regular clusters with more than 2 CpG sites, but each of them has at least 1 pair of polarity CpG sites that are embedded in this regular cluster. For example, "MMU-UUM" is a regular cluster, but its last 2 CpG sites have the "MU-UM" polarity pattern. Because only 4 of the 1719 clusters are like this, we consider them as regular clusters and define that a polarity cluster consists of only 2 CpG sites to simplify our definition. We point these 4 clusters out using the label "Regular*" to show the complexity of hemimethylation clusters.

clusters (see Tables 2 and 3). Among the 161 regular clusters, most of them are short 2-CpG clusters, including 69 MM-UU and 56 UU-MM (see Table 2). The patterns observed in Table 3 are similar to the previous finding regarding the HM patterns of the breast cancer cell line MCF7 (see Table 4 of Sun and Li[12]). In addition, if we summarize the clusters for $d \geqslant 0.6$ and $d \geqslant 0.8$, we get similar patterns for the count/frequency of different clusters (data not shown).

We calculate the length of HM clusters using the physical distance between the first and the last CpG site in a cluster (see Figure 3). For regular HM clusters, 154 of 161 clusters (ie, 95.65%) are less than or equal to 40 base pairs; 101 of 161 (ie, 62.73%) are just about 10 base long. For polarity clusters, 1506 of 1558 (ie, 96.66%) are about 40 to 50 base long. The cluster size or length patterns may show that when CpG sites are very close to each other (eg, less than 50 bases), they are likely to hemimethylate together. In addition, the chromosomal locations of hemimethylated clusters span the entire genome (see Figure 4).

Gene annotation

We have mapped the 19 736 HM sites in the cancer cell lines to their corresponding genes and promoters and have found that 6831 genes and 1399 promoter regions contain HM sites. A more detailed summary of the HM site distribution for genes and promoter regions is shown in Table 4. This table shows that 2194 genes have only 1 HM CpG site in their gene bodies; 1875 genes have 2 HM CpG sites; 885 genes have 3 CpG sites, and so on. Approximately 93% of the identified genes have at most 9 HM CpG sites (see Table 4). Even though most of the identified genes have no more than 9 HM sites, there are 319 genes that have 10 or more hemimethylated CpG sites (see Table 4). In addition, almost all the identified promoter regions have less than 5 HM sites (see Table 4).

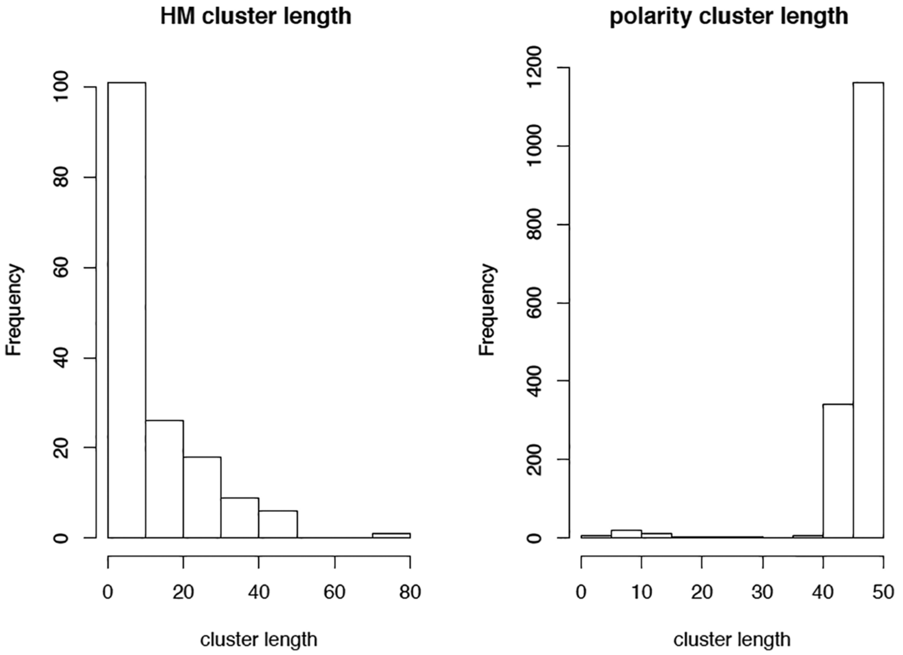

Figure 3. The length of HM clusters. HM indicates hemimethylation.
The left plot in Figure 3 is the histogram of the regular HM cluster length; the right plot is the histogram of the polarity cluster length.

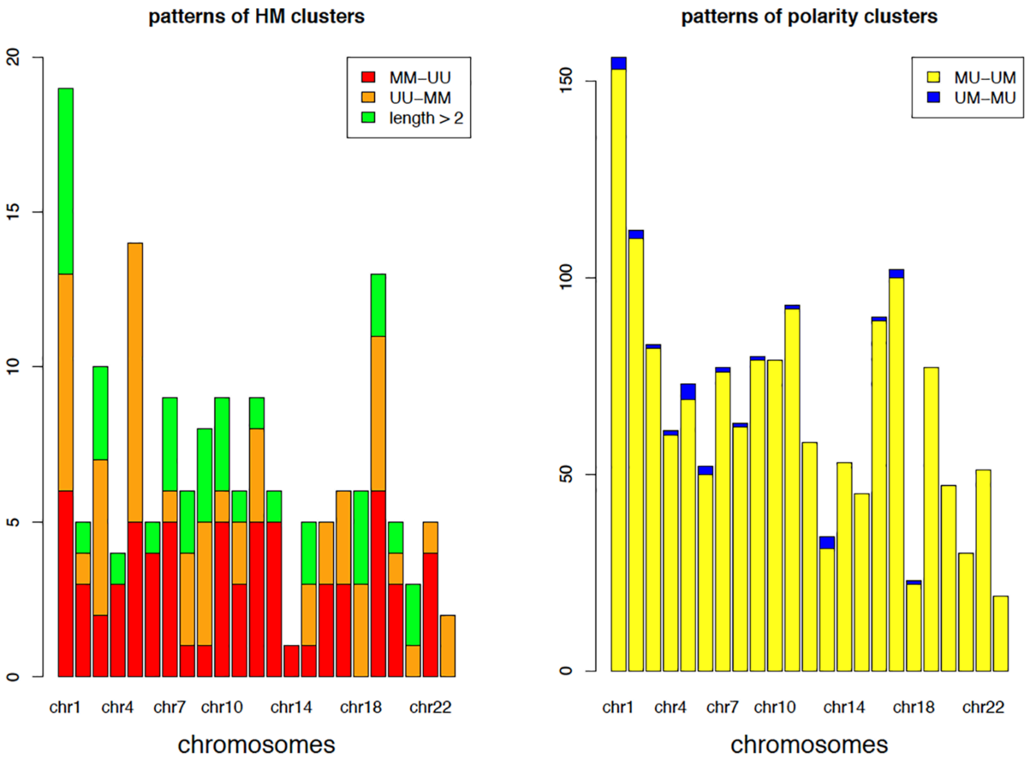

Figure 4. Hemimethylated CpG sites by chromosome. HM indicates hemimethylation.
The horizontal axis corresponds to the 23 pairs of chromosomes; 1 stacked bar represents each chromosome. The vertical axis corresponds to the number of HM clusters; 1 color corresponds to each type. For example, in the left plot, red is for "MM-UU," orange is for "UU-MM," and green is for clusters with more than 2 CpG sites (ie, length > 2).

Significant genes

The 7 breast cancer cell lines that we use to perform Wilcoxon signed rank tests are generated using the RRBS technique, which only sequences a small percentage of all CpG sites in a genome dependent on the insert size and alignment rate.[26] However, we do identify hemimethylated sites on each chromosome as shown in Figure 4. Therefore, we can say that HM in breast cancer spans the entire genome based on our results. After identifying genes with HM CpG sites, we have researched the functions of all genes with at least 25 HM sites and detailed them in Table 5.[27] The third column is some

Table 4. Summary of HM CpG site distribution.

SUMMARY OF HM SITES										
NO. OF HM CpG SITE PER GENE (*n*)	1	2	3	4	5	6	7	8	9	⩾10
No. of genes with *n* HM CpG sites	2194	1875	885	581	335	277	172	101	92	319
SUMMARY OF HM SITES ON PROMOTER REGIONS										
NO. OF HM CpG SITES PER PROMOTER REGION (*n*)	1	2	3	4	5					
No. of promoter regions with *n* HM CpG sites	849	373	105	50	22					

Abbreviation: HM, hemimethylation.

general description for each gene. In addition, we have also done some further research to find which of these 45 genes are involved in breast cancer (see Table 6). This further research is conducted using the GeneCards Batch Queries based on molecular function, phenotype, and genetic variants.[20] Our results show that 14 of these 45 genes are closely related to breast cancer, and in the second column of Table 6, we briefly summarize our findings based on the GeneCards Batch Queries.

Hemimethylation of hypermethylated breast cancer genes

Previous studies have identified tumor suppressor genes that exhibit hypermethylation in breast cancer cells.[4,28-31] Several of the hypermethylated tumor suppressor genes also have HM CpG sites as demonstrated in our analysis. Thirty-five genes hypermethylated in breast cancer are shown in Table 7. Most of the matched genes have at most 6 HM CpG sites except TP73 and TERT. It is very likely that there are more HM CpG sites located in these genes because the RRBS protocol can only sequence a small number of CpG sites in the genome. Different from the majority of the above genes, TP73 and TERT have 12 and 35 sites, respectively. Mutations of the TERT gene are associated with elongated telomere length commonly found in cancer cells such as breast cancer.[32] Furthermore, methylation of TP73 is associated with an increase in the malignancy and abnormality of breast cancer cells.[33] Hemimethylation of these genes may indicate important changes regarding the methylation events in breast cancer cells.

Table 7 shows that only 40% (14 genes out of 35) of known hypermethylated genes are hemimethylated (or have HM CpG sites in their gene body or promoter regions). The possible reason is that the RRBS protocol can only sequence a small number of CpG sites, about 3% to 6% of all CpG sites in a human genome, dependent on the insert size and alignment rate (see Table 1 of Doherty and Couldrey[26]). It is likely that the other CpG sites located in these hypermethylated genes (gene bodies or promoter regions) are hemimethylated, but the RRBS protocol does not generate data for these CpG sites, so we cannot determine the HM patterns for them. If possible, whole genome bisulfite sequencing (WGBS) may be a better option because it uses similar bisulfite-sequencing techniques but generates data for more than 99% of CpG sites. Hemimethylation analysis based on the WGBS data can give a better answer.

Hemimethylation of oncogenes and tumor suppressor genes

The Molecular Signatures Database (MSigDB) is created based on Gene Set Enrichment Analysis (GSEA).[21,34] It provides lists of genes that fit into certain gene sets such as oncogenes, tumor suppressors, and transcription factors. We compare our list of hemimethylated genes (each of which has one or more hemimethylated CpG sites) with the MSigDB list of oncogenes and tumor suppressors and find 157 oncogenes and 32 tumor suppressors have hemimethylated CpG sites. Because the MSigDB provides lists of oncogenes and tumor suppressor genes for all types of cancers, our finding is based on a large database, not on a database only for breast cancer. Table 8 lists the genes in these groups with the most HM sites. Among the genes listed in Table 8, 6 of them are closely related to breast cancer according to the GeneCards Batch Queries.[20] They are CBFA2T3, CRTC1, GNAS, SEPT9, APC, and FANCA. The HM of oncogenes (which are overexpressed in cancer cells) and tumor suppressors (which are silenced in cancer cells) may indicate complex methylation pattern changes and the instability of cancer DNA.

Gene pathways

ConsensusPathDB (or CPDB) is a database-type software package that integrates different types of functional interactions between physical entities like genes, RNA, proteins, protein complexes, and metabolites.[22-25] It provides different types of biological interaction analyses for a given set of genes. For the 45 genes that have at least 25 HM CpG sites, we use the CPDB to identify induced network modules. Even though the user provides a long list of genes as the input file, the CPDB produces a network with only the significantly enriched/represented genes. In our analysis, we focus on finding networks including genes with significant protein interaction and genetic interaction (see Figure 5). In this figure, 14 black-colored genes

Table 5. Forty-five genes with at least 25 hemimethylated CpG sites.

GENE	HEMIMETHYLATION SITES	DESCRIPTION
PTPRN2	90	Member of the protein tyrosine phosphatase (PTP) family that regulates a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation.
MAD1L1	87	Mitotic spindle-assembly checkpoint component that prevents the onset of anaphase until all chromosome are properly aligned at the metaphase plate. May play a role in cell cycle control and tumor suppression.
PRDM16	72	Zinc finger transcription factor. Translocation results in overexpression, which plays a role in myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML).
DIP2C	48	Encodes a member of the disco interacting protein homolog 2 family expressed in the nervous system.
CBFA2T3	44	Encodes a member of the myeloid translocation gene family, which interacts with DNA-bound transcription factors. Also known to be a putative breast tumor suppressor.
PRKAR1B	40	Encodes protein kinase A (PKA) enzyme that assists cell in regulation of metabolism, ion transport, and gene transcription.
KDM4B	39	Regulates gene expression by demethylating histone. Known to bind to ESR1 leading to tumorigenesis of various cancers.
KCNT1	37	Member of potassium sodium-activated channel subfamily T member 1.
SORCS2	36	Containing receptor for sortilin-related VPS10 domain.
KCNQ1	35	Encodes a voltage-gated potassium channel required for repolarization phase of the cardiac action potential.
MACROD1	35	Estrogen and androgen-responsive gene; known to have higher expression in hormone-dependent cancer cells such as MCF7.
TERT	35	Ribonucleoprotein polymerase that maintains telomere ends by addition of telomere repeats. Deregulation of telomerase expression in somatic cells may be involved in oncogenesis.
EXD3	35	Protein required for 3'-end trimming of AGO1-bound miRNAs.
NCOR2	34	Mediates transcriptional silencing of certain target genes. Aberrant expression of this gene is associated with certain cancers.
RPTOR	34	Component of a signaling pathway that regulates cell growth in response to nutrient and insulin levels.
C7orf50	34	Chromosome 7 open reading frame 50, poly(A) RNA binding is related to GO annotations.
TRAPPC9	34	Encodes a protein that likely plays a role in NF-kappa-B signaling.
CACNA1 H	33	Encodes a protein in the voltage-dependent calcium channel complex.
HDAC4	33	Histone deacetylase; plays a critical role in transcriptional regulation, cell cycle progression, and developmental events. Affects transcription factor access to DNA.
ADAMTS2	33	Responsible for the degradation of a major proteoglycan of cartilage, leading to arthritic disease.
SEPT9	32	Involved in cytokinesis and cell cycle control. Candidate for ovarian tumor suppressor gene.
NOTCH1	30	GO annotations include transcription factor activity and sequence-specific DNA binding.
VAV2	30	Second member of the VAV guanine nucleotide exchange factor family of oncogenes.
MOB2	29	MOB kinase activator 2.
FBRSL1	29	Fibrosin-line 1 protein.
GNAS	29	GNAS complex locus protein.
FAM20C	29	Encodes a protein that binds calcium and phosphorylates proteins involved in bone mineralization. Mutations in this gene are associated with Raine syndrome.
COL18A1	28	Collagen type XVIII alpha 1 chain protein.
CUX1	28	Member of the homeodomain family of DNA binding proteins. May regulate gene expression, morphogenesis, differentiation, and cell cycle progression.

(Continued)

Table 5. (Continued)

GENE	HEMIMETHYLATION SITES	DESCRIPTION
MEGF6	27	Multiple EGF like domains 6 protein.
OBSCN	27	Obscurin, cytoskeletal calmodulin and titin-interacting RhoGEF protein.
SHANK2	27	Encodes synaptic proteins that may function as molecular scaffolds in the postsynaptic density of excitatory synapses. Alterations of the protein may be associated with susceptibility to autism spectrum disorder.
RASA3	27	Member of the GAP1 family of GTPase-activating proteins.
SOLH	27	Calpain 15 protein.
BAIAP2	27	BAI1 associated protein 2 protein.
AGAP1	27	Member of an ADP-ribosylation factor involved in membrane trafficking and cytoskeleton dynamics.
CDH4	27	Cadherin 4 protein.
COL5A1	27	Collagen type V alpha 1 chain protein.
PRKCZ	26	Protein kinase C Zeta protein.
ASPSCR1	26	UBX domain containing tether for SLC2A4 protein, related pathways are transcriptional misregulation in cancer.
KCNQ2	26	Potassium voltage-gated channel subfamily Q member 2 protein.
ZC3H3	26	Zinc finger CCCH type containing 3.
LMF1	25	Lipase maturation factor 1 protein.
RBFOX3	25	RNA binding protein.
IQSEC1	25	Promotes binding of GTP and is particularly important in regulating cell adhesion. Highly expressed in the prefrontal cortex.

are from our provided seed gene list (ie, about one third of the 45 genes in Table 5). Purple-colored genes are the intermediate nodes that are not from our provided seed gene list. However, these intermediate genes associate 2 or more seed genes with each other and overall have significantly many connections within the induced network module. These black and purple genes are significantly enriched/represented ones, and the significance is determined based on the Genes2Networks approach.[35] In particular, intermediate genes are ranked and selected according to the significance of association with the seed gene list. The association is quantified by a z-score calculated for each intermediate node based on the binomial proportion test. The default z-score in the CPDB is used for our analysis.

In Figure 5, TERT is a hub gene that has many genetic interactions with other genes; see the blue lines/connections with other purple-colored genes. Among these purple genes, estrogen receptor alpha (ERα or ESR1) is a typical breast cancer gene. ERα displays gene regulatory interaction with TERT, which is abnormally active in most cancer cells for its trait of dividing uncontrollably. Over the years, estrogens have been recognized as an important factor associated with breast cancers; more than half of all breast cancers overexpress ERα, and 70% of them respond to estrogen hormone therapy.[36] HDAC4 has protein interactions with a few genes; see the yellow lines connecting HDAC4 with related genes (the left side of Figure 5). This gene is said to be involved in the MTA1-mediated epigenetics regulation of ESR1 expression in breast cancer.[37] MTA1 is a transcriptional coregulator that can perform as a transcriptional corepressor, and with the combination of other components of NuRD, MTA1 acts as a transcriptional corepressor of BRCA1 and ESR1.[20] BRCA1 is a tumor suppressor gene that helps prevent cells from growing rapidly, and its expression is impeded by MTA1. The biological connections among genes with at least 25 hemimethylated CpG sites imply that the genes associated with breast cancer may also be hemimethylated.

Discussion

In this article, we analyze the HM patterns in 7 breast cancer cell lines. The novel contribution of this article lies in that this is the first-ever thorough analysis of breast cancer HM. Both HM clusters and singleton sites are identified. On the contrary, our article has certain limitations. First, there is no statistical analysis done on a number of control or normal breast samples to compare with breast cancer cell lines. This is because so far we could not find suitable normal breast sample sequencing

Table 6. Fourteen genes involved in breast cancer.

GENE	INVOLVEMENT IN BREAST OR BREAST CANCER
PTPRN2	Related to increased cell death in breast cancer cell line MDA-MB-435
MAD1L1	1. Related to increased cell death in breast cancer cell line MDA-MB-435 2. Genetic variants (VAR_019714,VAR_019718) of this gene are found in a breast cancer cell line.
DIP2C	Genetic variants (VAR_035905, VAR_035907) are found in this gene (found in a breast cancer sample)
CBFA2T3	This gene is a putative breast tumor suppressor. Alternative splicing results in transcript variants.
MACROD1	Overexpressed by estrogens in breast cancer MCF-7 cells, probably via an activation of nuclear receptors for steroids (ESR1 but not ESR2)
TERT	Related to an anomaly of the structure of the breast and the abnormal growth of breast tissue
CACNA1H	Related to ductal breast carcinoma, and a single nucleotide polymorphism (SNP) (rs761025927) found in this gene (uncertain-significance)
HDAC4	1. Involved in the MTA1-mediated epigenetic regulation of ESR1 expression in breast cancer. 2. Related to increased cell death in breast cancer cell line MDA-MB-435 3. A genetic variant (VAR_036042) of this gene was reported in a breast cancer sample 4. Related to the anomaly of the structure of the breast.
NOTCH1	NOTCH1 is 1 of 4 known genes encoding the NOTCH family of proteins, a group of receptors involved in the Notch signaling pathway. Activation of Notch has been shown to be correlative with mammary tumorigenesis in mice and increased expression of Notch receptors has been observed in a variety of cancer types including cervical, colon, head and neck, lung, renal, pancreatic, leukemia, and breast cancer. A number of treatment modalities have been explored related to Notch inhibition especially in breast cancer with mixed results.
GNAS	1. Related to the abnormal growth of breast tissue and neoplasm of breast. 2. Two SNPs (rs11554273 and rs121913495) are found in this gene. 3. Related to the presence of abnormally increased levels of prolactin in the blood (prolactin is a peptide hormone produced by the anterior pituitary gland that plays a role in breast development and lactation during pregnancy).
FAM20C	Related to an anomaly of the sternum, also known as the breastbone.
CUX1	A genetic variant (VAR_036285) is found in this gene (found in a breast cancer sample)
OBSCN	Genetic variants (VAR_035534, VAR_035537, VAR_035538) of this gene are found in a breast cancer sample
COL5A1	Related to an anomaly of the sternum, also known as the breastbone.

data, except a WGBS data set discussed below. Second, we mainly focus on using the statistical and bioinformatic data analysis to identify HM sites. It would be more meaningful to further investigate the genes with a large number of hemimethylated sites, for example, the methylation events in their gene bodies, promoters, and enhancer regions. This type of research would require the wet lab work, which is beyond our capacity. Third, it is also meaningful to study hemimethylated genes' expression levels and how these levels are related to the methylation and HM patterns of these genes. This research is beyond the scope of this article as new gene expression data of these genes should be generated for further studies. Fourth, even though RRBS data analysis shows the existence of HM in various genes, it does not paint the entire picture for us. This is because RRBS mainly captures CpG rich sections of DNA, leaving out regions with scattered CpG sites. In fact, for the RRBS cancer cell lines we analyze, there are less than 2% of the CpG sites with at least 3× coverage in each of the breast cancer cell lines. With regards to our project, if we had the WGBS data for cancer cell lines, we would see a better picture of the HM patterns in breast cancer data.

Our focus of this article is to find significant HM CpG sites between breast cancer forward and reverse strands. In addition, we have compared the 7 breast cancer cell lines (RRBS data) with a WGBS data set generated from the human mammary epithelial cell (HMEC), which is considered as a normal breast sample (GSE29127).[38] As we have only 1 normal breast sample to compare, we do not apply the Wilcoxon signed rank test to analyze the HMEC. Instead, we filter out the data by selecting both forward and reverse strands with at least 3× coverage. In addition, we let the absolute mean difference between the forward and reverse strand methylation signals be greater than 0.4, 0.6, and 0.8 to identify hemimethylated CpG sites. To find the most significant HM sites, absolute mean difference greater than 0.8 is used on both 7 breast cancer cell lines and the single normal sample (HMEC). Only 2 hemimethylated CpG sites are identified both in cancer and normal sample as shown in Figure 6. However, 9477 CpG sites are hemimethylated in cancer cell lines. We choose to show Figure 6 and related results in the Discussion section rather than the Results section as this comparison is only based on 1 normal sample.

Table 7. Genes known to be hypermethylated in breast cancer.

APAF1	**CDKN1*(1)**	EPM2AIP1	GPC3	**MYOD1*(1)**	**TERT*(35)**	WDR79
IKIP	CDKN2A	MLH1	**GSTP1*(2)**	PGR	**TGFBR1*(2)**	**TP73*(12)**
CCND2*(2)	CDKN2B	**ESR1*(6)**	**HOXA5*(1)**	**SOCS1*(1)**	THBS1	TWIST1
CDH1	**CST6*(1)**	FHIT	HOXA6	STAT1	**TIMP3*(2)**	**WT1*(5)**
CDH13*(5)	DAPK2	GJB2	HSD17B4	SYK	TP53	WIT1

The genes with asterisks and highlighted in bold are found to have HM CpG sites as noted in the parentheses. For example, "TERT*(35)" means the gene TERT is a hypermethylated gene, and it covers 35 HM CpG sites.

Table 8. Oncogenes and tumor suppressors with HM sites.

ONCOGENES		TUMOR SUPPRESSORS	
GENE	NO. OF HM SITES	GENE	NO. OF HM SITES
ASPSCR1	26	APC	3
BCL11B	13	CBLC	3
BCR	17	FANCA	6
CARD11	13	FANCC	3
CBFA2T3	44	PTCH1	4
CRTC1	11	SMARCA4	7
GNAS	29	STK11	5
MN1	10	TSC1	3
NOTCH1	30	TSC2	14
NTRK1	11	WT1	5
PDE4DIP	10		
PRDM16	72		
SEPT9	32		

Abbreviation: HM, hemimethylation.

This simple comparison result may not be generalized or comparable with the comparison results based on multiple normal breast samples.

Our method uses the Wilcoxon signed rank test with a significance level of .05 on each base position listed. The more tests we perform with this significance level, the higher the chance we have of performing a type I error, also known as a false positive. For our study, this means the possibility that we have identified an HM CpG site incorrectly is higher if we keep the significance level the same for each test. Theoretically, we should do a multiple testing correction. However, this is a challenging task for this type of large genome data with complex features. To address this issue, we have used both the *P* value and the mean difference value to select the HM sites. This selection will help us to avoid a high false positive rate and also make sure the selected sites are both statistically and biologically significant.

Although DNA methylation plays an important role in gene regulation by affecting gene expression, the relationship of HM patterns and gene expression in cancer are not well studied yet. To the best of our knowledge, so far only Xu and Corces report some related research work.[11,13] They find that hemimethylated sites are inherited over several cell divisions; stably inherited HM may regulate chromatin interaction. They also show that gene body HM is associated with increased transcription. According to these findings, it is important to study the relationship between HM patterns and gene expression in cancers.

In the past, many studies are conducted to address important questions related to methylation. These studies include different topics of methylation pattern analysis and identification, such as integrative data analysis for methylation and other data (eg, gene expression),[39-41] methylation patterns for noncoding RNA,[42] and pan cancer data analysis.[43,44] Many of the previous methylation studies are conducted by assuming symmetric methylation (ie, not considering HM). Because of the existence of HM and its impact on transcription, it is favorable if a research article states which DNA strand is analyzed after bisulfite conversion. In the future, it would be optimal if a methylation analysis is conducted on 2 DNA strands separately as suggested by Naue and Lee.[45]

The 7 breast cancer cell lines we analyzed have their own unique characteristics or belong to different subtypes.[46] For example, 4 (BT474, ZR751, MCF7 and T47D) are ER+, and 3 (BT20, MDAMB231, and MDAMB468) are ER−; BT474 is luminal B; MCF7, T47D, and ZR751 are luminal A. Therefore, it is likely that they have different methylation and HM profiles. Our analysis can identify the hemimethylated sites that are common among these 7 cell lines, but it does not identify the HM profile for each single cell line or each breast cancer subtype. With technical replicates of the sequencing data for each cell line, identifying the HM profile for each subtype can be done in the future.

Conclusion

In this article, we have conducted the first-ever research work on identifying HM in breast cancer cell lines. Our statistical analysis of RRBS data has shown the existence of genome-wide

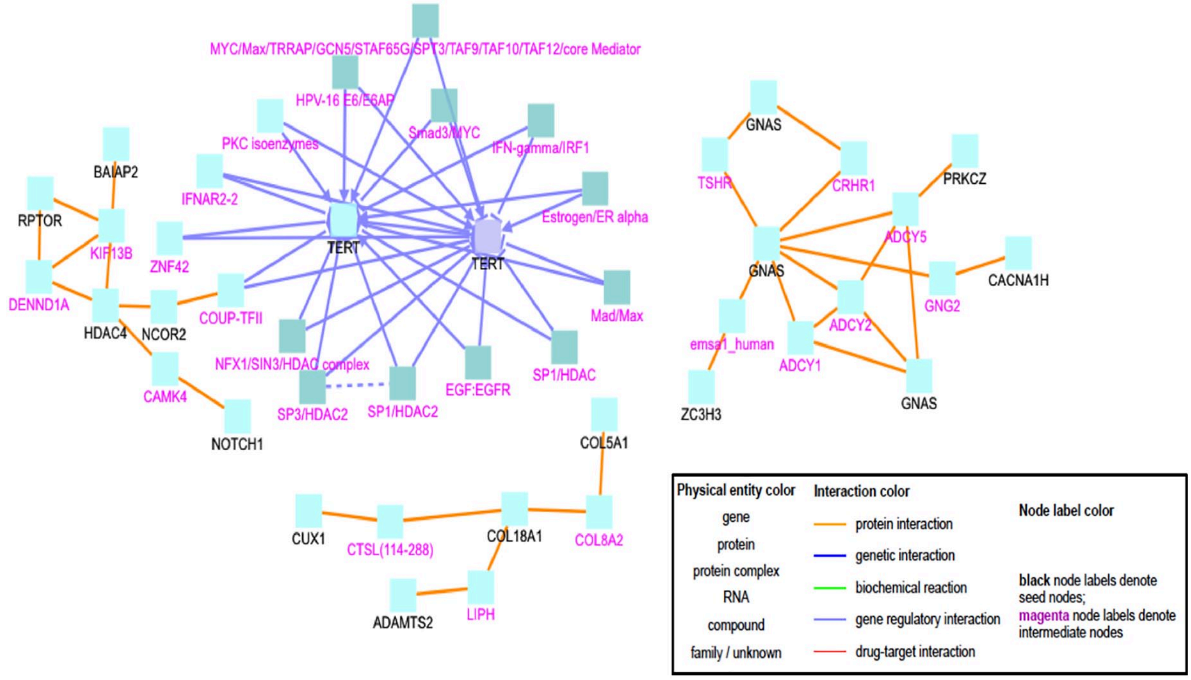

Figure 5. Relationships of 45 genes with at least 25 hemimethylated CpG sites.

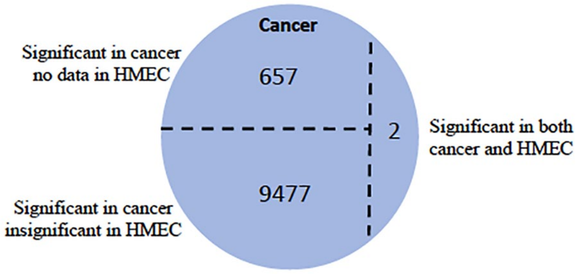

Figure 6. Comparing HM CpG sites in the HMEC and 7 breast cancer cell lines. HM indicates hemimethylation; HMEC, human mammary epithelial cell.

In total, 10 136 HM CpG sites are hemimethylated in breast cancer cell lines (ie, with the Wilcoxon test *P* value < .05 and absolute mean difference at least 0.8 as shown in Table 1). About 657 of these 10 136 CpG sites are hemimethylated in breast cancer cell lines, but there are no data in the HMEC sample. “No data in HMEC” means there are no sequencing reads or not enough sequencing reads covering those CpG sites (ie, <3× coverage). 9477 of these 10 136 CpG sites are hemimethylated in breast cancer, but not in HMEC. Only 2 of these 10 136 CpG sites are hemimethylated in both breast cancer cell lines and HMEC.

HM in breast cancer cell lines. Some of the genes that contain hemimethylated CpG sites may play a role in tumor growth or suppression. In addition, some of the hemimethylated genes associated with breast cancer are connected through biological pathways. Several of the hemimethylated genes are also known to be hypermethylated in breast cancer. In conclusion, these results suggest that certain genes in breast cancer cells undergo active methylation or demethylation, which results in genome-wide HM and may indicate a transition between different stages of breast cancer. This transition may occur before tumors develop. Thus, further study of HM may serve as a method to identify breast cancer in earlier stages and increase the chances of patient survival.

Acknowledgements

This project was completed with the use of Texas State University facilities and resources. The authors are grateful for the 3 reviewers, whose questions and comments help us improve this article greatly.

Author Contributions

S.S. initiated the project, suggested all key original ideas, and oversaw the whole process. Y.L. and B.E. conducted the main data analysis. All 3 authors contributed to the interpretation of data analysis and the writing of this article. S.S. gave suggestions over the course of the project and extensively reviewed and revised the final article. All authors contributed expertise and edits. All authors have read and approved the final article.

Availability of data and material

Data sets used in this article are publicly available. R code files are available upon request.

ORCID iD

Shuying Sun https://orcid.org/0000-0003-3974-6996

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin.* 2019;69:7-34.
2. Howlader N, Noone AM, Krapcho M, et al. *SEER Cancer Statistics Review, 1975-2016*. Bethesda, MD: National Cancer Institute. https://seer.cancer.gov/csr/1975_2016/. Updated April 2019.
3. Moore LD, Le T, Fan G. DNA methylation and its basic function. *Neuropsychopharmacology*. 2013;38:23-38.
4. Yang X, Yan L, Davidson NE. DNA methylation in breast cancer. *Endocr Relat Cancer.* 2001;8:115-127.

5. Beck S, Rakyan VK. The methylome: approaches for global DNA methylation profiling. *Trends Genet*. 2008;24:231-237.
6. Hurd PJ, Nelson CJ. Advantages of next-generation sequencing versus the microarray in epigenetic research. *Brief Funct Genomic Proteomic*. 2009;8:174-183.
7. Li Y, Tollefsbol TO. DNA methylation detection: bisulfite genomic sequencing analysis. *Methods Mol Biol*. 2011;791:11-21.
8. Li N, Ye M, Li Y, et al. Whole genome DNA methylation analysis based on high throughput sequencing technology. *Methods*. 2010;52:203-212.
9. Ehrlich M, Lacey M. DNA hypomethylation and hemimethylation in cancer. *Adv Exp Med Biol*. 2013;754:31-56.
10. Shao C, Lacey M, Dubeau L, Ehrlich M. Hemimethylation footprints of DNA demethylation in cancer. *Epigenetics*. 2009;4:165-175.
11. Sharif J, Koseki H. Hemimethylation: DNA's lasting odd couple. *Science*. 2018;359:1102-1103.
12. Sun S, Li P. HMPL: a pipeline for identifying hemimethylation patterns by comparing two samples. *Cancer Inform*. 2015;14:235-245.
13. Xu C, Corces VG. Nascent DNA methylome mapping reveals inheritance of hemimethylation at CTCF/cohesin sites. *Science*. 2018;359:1166-1170.
14. Sun Z, Asmann YW, Kalari KR, et al. Integrated analysis of gene expression, CpG island methylation, and gene copy number in breast cancer cells by deep sequencing. *PLoS ONE*. 2011;6:e17490.
15. Harris EY, Ponts N, Le Roch KG, Lonardi S. BRAT-BW: efficient and accurate mapping of bisulfite-treated reads. *Bioinformatics*. 2012;28:1795-1796.
16. Hall JN, Schwartz RL. *Effective Perl Programming: Writing Better Programs With Perl*. Reading, MA: Addison Wesley; 1998.
17. Cozens S. *Advanced Perl Programming* (Safari Books Online (Firm)). 2nd ed. Sebastopol, CA; Farnham, UK: O'Reilly Media; 2005. http://proquest.safaribooksonline.com/0596004567.
18. Team TRC. *R: A Language and Environment for Statistical Computing*. Version 3.0.1. Vienna: R Foundation for Statistical Computing; 2013.
19. Conover WJ. On methods of handling ties in the Wilcoxon signed-rank test. *J Am Stat Assoc*. 1973;68:985-988.
20. GeneCards—Gene Database. www.genecards.org.
21. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102:15545-15550.
22. Herwig R, Hardt C, Lienhard M, Kamburov A. Analyzing and interpreting genome data at the network level with ConsensusPathDB. *Nat Protoc*. 2016;11:1889-1907.
23. Kamburov A, Pentchev K, Galicka H, Wierling C, Lehrach H, Herwig R. ConsensusPathDB: toward a more complete picture of cell biology. *Nucleic Acids Res*. 2011;39:D712-D717.
24. Kamburov A, Stelzl U, Lehrach H, Herwig R. The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res*. 2013;41:D793-D800.
25. Kamburov A, Wierling C, Lehrach H, Herwig R. ConsensusPathDB —a database for integrating human functional interaction networks. *Nucleic Acids Res*. 2009;37:D623-D628.
26. Doherty R, Couldrey C. Exploring genome wide bisulfite sequencing for DNA methylation analysis in livestock: a technical assessment. *Front Genet*. 2014;5:126.
27. Brown GR, Hem V, Katz KS, et al. Gene: a gene-centered information resource at NCBI. *Nucleic Acids Res*. 2015;43:D36-D42.
28. Fiegl H, Millinger S, Goebel G, et al. Breast cancer DNA methylation profiles in cancer cells and tumor stroma: association with HER-2/neu status in primary breast cancer. *Cancer Res*. 2006;66:29-33.
29. Tan LW, Bianco T, Dobrovic A. Variable promoter region CpG island methylation of the putative tumor suppressor gene Connexin 26 in breast cancer. *Carcinogenesis*. 2002;23:231-236.
30. Widschwendter M, Jones PA. DNA methylation and breast carcinogenesis. *Oncogene*. 2002;21:5462-5482.
31. Sun S, Chen Z, Yan PS, Huang YW, Huang TH, Lin S. Identifying hypermethylated CpG islands using a quantile regression model. *BMC Bioinformatics*. 2011;12:54.
32. Bojesen SE, Pooley KA, Johnatty SE, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat Genet*. 2013;45:371-384, 384e1-2.
33. Marzese DM, Hoon DS, Chong KK, et al. DNA methylation index and methylation profile of invasive ductal breast tumors. *J Mol Diagn*. 2012;14:613-622.
34. Mootha VK, Lindgren CM, Eriksson KF, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*. 2003;34:267-273.
35. Berger SI, Posner JM, Ma'ayan A. Genes2Networks: connecting lists of gene symbols using mammalian protein interactions databases. *BMC Bioinformatics*. 2007;8:372.
36. Ali S, Coombes RC. Estrogen receptor alpha in human breast cancer: occurrence and significance. *J Mammary Gland Biol Neoplasia*. 2000;5:271-281.
37. Seo JH, Park JH, Lee EJ, et al. ARD1-mediated Hsp70 acetylation balances stress-induced protein refolding and degradation. *Nat Commun*. 2016;7:12882.
38. Hon GC, Hawkins RD, Caballero OL, et al. Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res*. 2012;22:246-258.
39. Li C, Lee J, Ding J, Sun S. Integrative analysis of gene expression and methylation data for breast cancer cell lines. *BioData Min*. 2018;11:13.
40. Ma X, Liu Z, Zhang Z, Huang X, Tang W. Multiple network algorithm for epigenetic modules via the integration of genome-wide DNA methylation and gene expression data. *BMC Bioinformatics*. 2017;18:72.
41. Ma X, Sun PG, Zhang ZY. An integrative framework for protein interaction network and methylation data to discover epigenetic modules [published online ahead of print April 30, 2018]. *IEEE/ACM Trans Comput Biol Bioinform*. doi:10.1109/TCBB.2018.2831666.
42. Ma X, Yu L, Wang P, Yang X. Discovering DNA methylation patterns for long non-coding RNAs associated with cancer subtypes. *Comput Biol Chem*. 2017;69:164-170.
43. Yang X, Gao L, Zhang S. Comparative pan-cancer DNA methylation analysis reveals cancer common and specific patterns. *Brief Bioinform*. 2017;18:761-773.
44. Zhang J, Huang K. Pan-cancer analysis of frequent DNA co-methylation patterns reveals consistent epigenetic landscape changes in multiple cancers. *BMC Genomics*. 2017;18:1045.
45. Naue J, Lee HY. Considerations for the need of recommendations for the research and publication of DNA methylation results. *Forensic Sci Int Genet*. 2018;37:e12-e14.
46. Dai X, Cheng H, Bai Z, Li J. Breast cancer cell line classification and its relevance with breast tumor subtyping. *J Cancer*. 2017;8:3131-3141.