DNA Methylation in Breast Tumor from High-risk Women in the Breast Cancer Family Registry

HUI-CHEN WU¹, MELISSA C. SOUTHEY², HANINA HIBSHOOSH³, REGINA M. SANTELLA¹,⁴ and MARY BETH TERRY¹,⁵

¹Department of Environmental Health Sciences, Mailman School of Public Health of Columbia University, New York, NY, U.S.A.; ²Department of Pathology, The University of Melbourne, Melbourne, Victoria, Australia; ³Department of Pathology and Cell Biology, College of Physicians & Surgeons of Columbia University, New York, NY, U.S.A.; ⁴Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center, New York, NY, U.S.A.; ⁵Department of Epidemiology, Mailman School of Public Health of Columbia University, New York, NY, U.S.A.

Abstract. To examine DNA methylation profiles in breast tumors of women with a strong breast cancer family history, we measured methylation by bisulfite sequencing in 40 genes in 40 breast tumor tissues from women in the Breast Cancer Family Registry. We selected candidate genes from analysis of the Cancer Genome Atlas project (TCGA) breast data. Compared to TCGA breast cancer, BCFR cases are younger and more likely to be ER-negative. Overall, we found that many of the methylation differences between BCFR tumor and normal adjacent tissues were smaller than that in TCGA samples. We found only 32% of tested genes were hypermethylated in BCFR; the largest difference was 36.1% for SEPW1, and the smallest difference was 10% for RYR2. These data suggest the importance of examining breast cancer cases including familial cases enriched with early-onset cancers to identify methylation markers that can be examined in blood as biomarkers for early detection.

Studies examining the associations between whole-genome DNA methylation and breast cancer classification found that there were distinct methylation patterns by hormone receptor status (1, 2) and by BRCA mutation state (3). In addition, DNA methylation profiles can also identify a cluster of breast cancers that are not classified by current expression subtypes (4). Only few studies have examined breast tumor tissues on DNA methylation in selected genes in familial breast cancer (5-7). Examining methylation in the promoter region of 10 selected genes in both inherited and non-inherited breast cancer, Esteller et al. (5) suggested that hereditary breast cancers have methylation levels similar to that of sporadic tumors in these genes. However, another study reported that the median cumulative methylation index calculated as the sum of the percentage methylation for 11 genes was significantly lower in BRCA1-related breast cancers than in sporadic breast cancers (6).

Recent data from The Cancer Genome Atlas (TCGA) has dramatically expanded the number of genes identified as aberrantly methylated in breast cancer (8). The aim of the current study was to examine the methylation values in selected genes that are highly methylated in TCGA breast tumor tissues in breast cancer tissues from high-risk women. By analyzing data of Illumina HumanMethylation450 array in breast cancer in TCGA Network, we first selected 40 genes that are hypermethylated in breast tumor than adjacent non-tumor tissues. We then compared DNA methylation levels of these 40 genes using high throughput targeted bisulfite sequencing in 40 breast tumor and non-tumor tissue pairs from high-risk women from the New York site of the Breast Cancer Family Registry (BCFR). We also examined the relationship of methylation in each loci with estrogen receptor (ER) status and age at cancer diagnosis.

Materials and Methods

Study participants. We selected breast cancer cases from families participating in the New York site of the BCFR (9, 10). We recruited high risk breast and/or ovarian cancer families from clinical and community settings within the Metropolitan New York area who
were eligible to participate if they met one of the following criteria: i) have a female relative with breast or ovarian cancer diagnosed before age 45 years, ii) have a female relative with both breast and ovarian cancer regardless of age at diagnosis, iii) have two or more relatives with breast or ovarian cancer diagnosed after age 45 years, iv) be a male with breast cancer diagnosed at any age, or v) have a family member with a known BRCA mutation. The study was approved by Columbia University’s Institutional Review Board; written informed consent was obtained from all subjects, and strict quality controls and safeguards were used to protect confidentiality. For DNA methylation analysis, we analyzed 40 formaldehyde-fixed paraffin embedded (FFPE) breast tumors and adjacent non-tumor tissues from breast cancer cases who have family history of breast cancer, including 19 ER-negative and 21 ER-positive tumors.

DNA extraction and bisulfite treatment. With the H&E-stained slide used as a guide, the pathologist (Dr. Hibshoosh) circled the region of breast carcinoma on the corresponding unstained slide for macrodissection. Non-tumor tissue from outside the circled area was also removed. Tissues were de-paraffinized before DNA extraction. We used two 10-μm sections per case to extracted DNA by a standard phenol-chloroform protocol. Aliquots of DNA (1 μg) were bisulfite-treated with Epitect kit (Qiagen) as per the manufacturer’s instructions. The DNA was resuspended in 20 μL of distilled water and stored at −20°C until use.

**TCGA data analysis for selection of genes.** To identify breast-specific DNA methylation markers, we downloaded Illumina Infinium HumanMethylation450 data (Level 3 data) for tumor tissues and adjacent non-tumor tissues and clinical annotation tables for 96 breast invasive carcinomas from TCGA data portal (https://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp) (5/20/2014). We used paired sample t-test with Bonferroni’s correction for multiple testing to identify CpG sites that were differentially methylated between tumor and adjacent tissues. A statistically significant difference was defined as sites with a Bonferroni-corrected p-value <0.0005; a total of 84,584 CpG sites were identified. We then selected candidate genes (Table 1) based on the criteria for selecting driver methylation changes (11) from TCGA breast cancer data; they are: 1) effect size, with >20% higher methylation in tumor tissues, 3) genomic location, with CpG sites located in the CpG Island, 4) excluding any CpG sites with higher methylation in tumor tissues, 3) genomic location, with CpG sites located in the CpG Island, 4) excluding any CpG sites with higher methylation in tumor tissues, 3) genomic location, with CpG sites located in the CpG Island, 4) excluding any CpG sites with higher methylation in tumor tissues, 3) genomic location, with CpG sites located in the CpG Island.
Bisulfite DNA sequencing for measuring DNA methylation in candidate genes. DNA methylation was measured using an Illumina MiSEQ instrument as per the manufacturer’s instructions. We designed 40 pairs of oligonucleotide primers to cover the target CpG site region using MethPrimer with default parameters (12). The number of CpG sites per amplicon ranges from 4 to 34. We trimmed the Fastq files generated by sequencing (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) for both adapters and for a quality cut off of 30 using trimmed galore. Sequencing alignment and methylation calls was done via Bismark (13) and bowtie2 (14) and the methylation status and methylation level of each analyzed CpG-sites were then determined. The methylation level for each gene was assigned by averaging the methylation level of all CpG sites in the gene for each sample. Primers are available on requesting from the authors.

Statistical methods. We used paired-t-test to determine the differences in methylation between breast tumor and adjacent nontumor tissues of BCFR samples. We used Wilcoxon Rank Sum test to examine methylation by ER status (positive vs. negative) and age at cancer diagnosis (<50 vs. ≥50 years of age) among tumor tissues. We used SAS software 9.3 (SAS Institute, Cary, NC, USA) for the analyses.

Results

Table I presents the distributions of selected characteristics in BCFR cases and TCGA breast cancer cases. BCFR breast cases were younger than TCGA cases. The mean age of breast cancer diagnosis was 46.4±11.4 for BCFR cases and 57.2±15.4 for TCGA cases. Sixty-five percent of BCFR breast cases were diagnosed with breast cancer before the age of 50, while only 33% of TCGA breast cases were early onset cases. Most breast cancer cases (71%) in TCGA were ER positive and 53% cases in the BCFR were ER positive.

The overall methylation levels for each locus in tumor and adjacent tissues in BCFR tissue samples and corresponding methylation difference in TCGA breast and adjacent non-tumor tissues are given in Table II. Overall, the methylation levels for most of genes were higher in tumor samples than in adjacent tissues in BCFR tissue samples. Seven genes (SEPW1, PCDHGA4, CCDC36, C1orf14, RPTOR, C1orf114, and ZNF454) had methylation levels in tumors at least 20% higher than in adjacent tissues. The largest difference was 36.1% for SEPW1. A total of 13 genes (SEPW1 (74.4 vs. 38.6%), PCDHGA4 (46.0 vs. 17.9 %), CCDC36 (42.7 vs. 16.2%), C1orf14 (55.7 vs. 29.3%), RPTOR (53.2 vs. 28.9%), C1orf114 (54.6 vs. 31.0%), ZNF454 (32.9 vs. 10.1%), USP44 (40.1 vs. 22.4%), CSMD3 (29.1 vs. 12.1%), PRKAR1B (61.1 vs. 48.6%), SLC7A14 (32.9 vs. 21.0%), SOX2OT (32.1 vs. 20.6%) and RYR2 (24.0 vs. 14.0%) showed significantly higher methylation in breast tumor compared to adjacent non-tumor tissues. Although the technologies for DNA methylation measurement are different between our and TCGA, the methylation levels in TCGA breast tumor were higher than that in BCFR tumor samples, while the methylation levels in TCGA adjacent samples were lower than that in BCFR adjacent samples (Figure 1).

We also examined the methylation of these genes by ER status and age at cancer diagnosis in BCFR samples and TCGA breast tumor data. In TCGA breast tumor, methylation
Table III. Distribution of DNA methylation levels by estrogen receptor (ER) status (A) and by age at breast cancer diagnosis (B) in breast tumors from the New York site of the BCFR by bisulfite sequencing and from The Cancer Genome Atlas Illumina 450k array.

<table>
<thead>
<tr>
<th>Gene</th>
<th>BCFR Mean (SD),%</th>
<th>TCGA Mean (SD),%</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ER Positive</td>
<td>ER Negative</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean (SD),%</td>
<td>Mean (SD),%</td>
<td>p-Value</td>
</tr>
<tr>
<td>C1orf14</td>
<td>58.5 (27.7)</td>
<td>52.0 (24.3)</td>
<td>0.54</td>
</tr>
<tr>
<td>ZNF454</td>
<td>37.8 (33.7)</td>
<td>23.4 (27.0)</td>
<td>0.28</td>
</tr>
<tr>
<td>GRM1</td>
<td>40.1 (30.6)</td>
<td>39.6 (26.8)</td>
<td>0.96</td>
</tr>
<tr>
<td>GPRC5B</td>
<td>13.9 (21.0)</td>
<td>7.3 (15.2)</td>
<td>0.31</td>
</tr>
<tr>
<td>VAX1</td>
<td>13.9 (14.6)</td>
<td>15.0 (15.8)</td>
<td>0.82</td>
</tr>
<tr>
<td>TULP1</td>
<td>33.6 (16.6)</td>
<td>29.3 (27.5)</td>
<td>0.65</td>
</tr>
<tr>
<td>LITD1</td>
<td>55.6 (28.0)</td>
<td>33.5 (27.2)</td>
<td>0.08</td>
</tr>
<tr>
<td>KLHDC7B</td>
<td>61.9 (33.6)</td>
<td>61.4 (30.1)</td>
<td>0.97</td>
</tr>
<tr>
<td>RPTOR</td>
<td>46.0 (28.1)</td>
<td>63.7 (27.5)</td>
<td>0.09</td>
</tr>
<tr>
<td>TBR1</td>
<td>37.8 (37.6)</td>
<td>14.0 (10.3)</td>
<td>0.41</td>
</tr>
<tr>
<td>TTBK1</td>
<td>21.7 (17.8)</td>
<td>21.0 (23.1)</td>
<td>0.92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>&lt;50 yrs of age</th>
<th>&gt;50 yrs of age</th>
<th>&lt;50 yrs of age</th>
<th>&gt;50 yrs of age</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPTOR</td>
<td>46.0 (28.1)</td>
<td>63.7 (27.5)</td>
<td>0.09</td>
<td>0.59 (0.24)</td>
<td>0.72 (0.15)</td>
</tr>
<tr>
<td>TBR1</td>
<td>37.8 (37.6)</td>
<td>14.0 (10.3)</td>
<td>0.41</td>
<td>0.47 (0.27)</td>
<td>0.62 (0.28)</td>
</tr>
<tr>
<td>TTBK1</td>
<td>21.7 (17.8)</td>
<td>21.0 (23.1)</td>
<td>0.92</td>
<td>0.40 (0.33)</td>
<td>0.64 (0.33)</td>
</tr>
</tbody>
</table>

In C1orf14, ZNF454, GPRC5B and KLHDC7B were higher, while methylation in GRM1, VAX1, TULP1, and LITD1 were lower in ER positive than ER negative (Table III). The methylation levels in 3 genes were higher in older cases of TCGA samples (Table III). We did not observe any different by ER status or age in BCFR tumor samples.

Discussion

In this study, we examined DNA methylation profiles in breast tumor tissues from women at high risk. Among 40 tested genes that are hypermethylated in TCGA breast tumor tissues, we found that only 13 of them were highly methylated in our sample of high-risk women. Consistent with the TCGA breast data, methylation levels in SEPW1, PCDHGA4, CCDC36, C1orf14, RPTOR, C1orf14, ZNF454, USP44, CSMD3, PRKAR1B, SLC7A14, SOX2OT and RYR2 were significantly higher in the breast tumor than in adjacent histologically normal tissues. The difference in methylation between tumor and adjacent tissues ranged from 10.0% for RYR2 to 36.1% for SEPW1. Compared to data from TCGA breast tissues that are from mainly sporadic breast cancer cases, the methylation levels were lower in our BCFR tumor samples, however, the levels were higher in our adjacent samples then in TCGA adjacent tissues.

Differences in the technologies for DNA methylation measurement between our study (bisulfite sequencing) and TCGA (Illumina arrays) cannot explain the observation of smaller differences in DNA methylation between breast tumor and adjacent non-tumor tissues in our study than in TCGA data, as we compared methylation difference between tumor and non-tumor tissues within the same platform. The TCGA was not enriched for women at very high risk (e.g., strong family history of breast cancer) and young cancer cases (33% in TCGA vs. 65% in BCFR samples). Clinical observations and molecular studies suggest biology of early-onset breast cancer is different from that of late-onset cases (15). In contrast to later-onset cancer, early-onset breast cancer is enriched with ER-negative tumor (16). We did not see any methylation difference by ER status, but we found that methylation status of 11 genes differed by either ER status or age of cancer diagnosis in TCGA samples. As we only examined a small fraction (<1%) of hypermethylated genes in TCGA breast data, to better characterize DNA methylation profiles in familial breast cancer, studies examining a large-scale of DNA methylation profiles in breast tumor tissues with detail family history of breast cancer are needed.

Examining epigenome-wide DNA methylation profiles in breast tumors, several studies have identified a number of alternately methylated genes including SEPW1, PCDHGA4, CCDC36, C1orf14, RPTOR, C1orf14, ZNF454, USP44, CSMD3, PRKAR1B, SLC7A14, SOX2OT, and RYR2 (8, 17). Comparing DNA methylation profiles between healthy breast tissue and ductal carcinoma in situ using HumanMethylation450 microarrays, Fleischer et al. (17) found a total 16,949 CpGs were differentially methylated, representing 5,659 genes including SEPW1, PCDHGA4, CCDC36, C1orf14, RPTOR, C1orf14, ZNF454, USP44, CSMD3, PRKAR1B, SLC7A14, SOX2OT, and RYR2 (8, 17). Comparing DNA methylation profiles between healthy breast tissue and ductal carcinoma in situ using HumanMethylation450 microarrays, Fleischer et al. (17) found a total 16,949 CpGs were differentially methylated, representing 5,659 genes including SEPW1, PCDHGA4, CCDC36, C1orf14, RPTOR, C1orf14, ZNF454, USP44, CSMD3, SLC7A14, RYR2, and RPTOR, suggesting hypermethylation of these selected genes occurs early in breast cancer development including familial breast cancer. Hypermethylation in the promoter regions of key genes might allow for clonal selection and growth of tumor cells (18).
Using whole-exome sequencing in breast tumor tissue, mutations in RYR2, C1orf14, SLC7A14, and RPTOR have been identified (8, 19-21). The observation of hypermethylation of these genes in breast cancer suggests that epigenetic silencing via methylation is another important mechanism for inactivating these genes in breast cancer tumorigenesis. Genome-wide DNA methylation studies have supported the correlation of CpG island DNA hypermethylation and gene expression (4, 22-24). Using MEXPRESS (25) (http://mexpress.be/) to visualize and examine DNA methylation and expression data from data in TCGA breast tumor, methylation was significantly associated with gene expression in SEPWII (person correlation coefficient $r=-0.11$), PCDHGA4 ($r=-0.21$), CCDC36 ($r=0.42$), RPTOR ($r=0.24$), ZNF454 ($r=-0.60$), USP44 ($r=0.63$), CSMD3 ($r=0.01$), PRKAR1B ($r=0.39$), SLC7A14 ($r=-0.32$), and RYR2 ($r=-0.19$). Further studies are needed to better understand the role of DNA methylation alterations in these genes involved in breast tumorigenesis.

In summary, we found that the genes we selected from TCGA breast data, because they were dysregulated in DNA methylation were not all dysregulated in breast tumor tissues from women with a strong breast cancer family history. As accumulating data suggest the importance of DNA methylation alternations for cancer risk, a more comprehensive set of breast cancer cases including familial cases enriched with early-onset cancers need to be studied in order to identify methylation changes in genes that can be examined in the peripheral blood as biomarkers for early detection.

Conflicts of Interest

There is no conflict for all authors.

Acknowledgements

The Authors thank the Epigenetics Shared Resource of the Herbert Irving Comprehensive Cancer Center for performing the targeted bisulfite sequencing. We would like to gratefully acknowledge all of the families that participate in the New York site of the BCFR. The Breast Cancer Family Registry (BCFR) was supported by grant UM1 CA164920 from the USA National Cancer Institute. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the BCFR, nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR.

References


Received December 23, 2016
Revised January 13, 2017
Accepted January 17, 2017