Breast cancer family history and allele-specific DNA methylation in the legacy girls study

Hui-Chen Wu, Catherine Do, Irene L. Andrulis, Esther M. John, Mary B. Daly, Saundra S Buys, Wendy K. Chung, Julia A. Knight, Angela R. Bradbury, Theresa H. M. Keegan, Lisa Schwartz, Izabela Krupska, Rachel L. Miller, Regina M. Santella, Benjamin Tycko & Mary Beth Terry

To cite this article: Hui-Chen Wu, Catherine Do, Irene L. Andrulis, Esther M. John, Mary B. Daly, Saundra S Buys, Wendy K. Chung, Julia A. Knight, Angela R. Bradbury, Theresa H. M. Keegan, Lisa Schwartz, Izabela Krupska, Rachel L. Miller, Regina M. Santella, Benjamin Tycko & Mary Beth Terry (2018) Breast cancer family history and allele-specific DNA methylation in the legacy girls study, Epigenetics, 13:3, 240-250, DOI: 10.1080/15592294.2018.1435243

To link to this article: https://doi.org/10.1080/15592294.2018.1435243

View supplementary material

Accepted author version posted online: 13 Feb 2018.
Published online: 02 Apr 2018.

Submit your article to this journal

Article views: 202

View Crossmark data
Breast cancer family history and allele-specific DNA methylation in the legacy girls study

Hui-Chen Wu, Catherine Do, Irene L. Andrulis, Esther M. John, Mary B. Daly, Saundra S Buys, Wendy K. Chung, Julia A. Knight, Angela R. Bradbury, Theresa H. M. Keegan, Lisa Schwartz, Izabela Krupska, Rachel L. Miller, Regina M. Santella, Benjamin Tycko and Mary Beth Terry

ABSTRACT

Family history, a well-established risk factor for breast cancer, can have both genetic and environmental contributions. Shared environment in families as well as epigenetic changes that also may be influenced by shared genetics and environment may also explain familial clustering of cancers. Epigenetic regulation, such as DNA methylation, can change the activity of a DNA segment without a change in the sequence; environmental exposures experienced across the life course can induce such changes. However, genetic-epigenetic interactions, detected as methylation quantitative trait loci (mQTLs; a.k.a. meQTLs) and haplotype-dependent allele-specific methylation (hap-ASM), can also contribute to inter-individual differences in DNA methylation patterns. To identify differentially methylated regions (DMRs) associated with breast cancer susceptibility, we examined differences in white blood cell DNA methylation in 29 candidate genes in 426 girls (ages 6-13 years) from the LEGACY Girls Study, 239 with and 187 without a breast cancer family history (BCFH). We measured methylation by targeted massively parallel bisulfite sequencing (bis-seq) and observed BCFH DMRs in two genes: ESR1 (ΔΔ4.9%, \( P = 0.003 \)) and SEC16B (ΔΔ3.6%, \( P = 0.026 \)), each of which has been previously implicated in breast cancer susceptibility and pubertal development. These DMRs showed high inter-individual variability in methylation, suggesting the presence of mQTLs/hap-ASM. Using single nucleotide polymorphisms data in the bis-seq amplicon, we found strong hap-ASM in SEC16B (with allele-specific differences ranging from 42% to 74%). These findings suggest that differential methylation in genes relevant to breast cancer susceptibility may be present early in life, and that inherited genetic factors underlie some of these epigenetic differences.

Abbreviations: BCA: breast cancer; BCFH: breast cancer family history; bis-seq: bisulfite sequencing; BMI: body mass index; DMR: differentially methylated regions; EWAS: epigenome-wide association study; GWAS: genome-wide association study; hap-ASM: haplotype-dependent allele-specific methylation; LD: linkage disequilibrium; LEGACY: Lessons in Epidemiology and Genetics of Adult Cancer from Youth; mQTL: methylation quantitative trait loci; SNPs: single nucleotide polymorphisms; WBC: white blood cells

Introduction

Family history is a well-established risk factor for breast cancer (BCA) [1], increasing risk 2- to 4-fold, depending on the number of affected 1st- and 2nd-degree relatives and their ages at diagnosis [1]. Three classes of BCA susceptibility genes—high, intermediate, and low—exhibit different levels of risk and prevalence in the general population (reviewed in [2]); pathogenic variants in high-risk genes, including BRCA1, BRCA2, PTEN, and TP53, are relatively rare and explain about 20% of the proportion of the genetic variance in risk, whereas pathogenic

CONTACT Mary Beth Terry mt146@columbia.edu  Department of Epidemiology, Mailman School of Public Health of Columbia University, 722 West 168th St., 1611, New York, NY 10032, USA

*The first two authors contributed equally to this work.

Supplemental data for this article can be accessed at https://doi.org/10.1080/15592294.2018.1435243.

© 2018 Informa UK Limited, trading as Taylor & Francis Group
variants in intermediate-risk genes, including ATM, BRIP1, CHEK2, and PALB2, explain a further 3% of the heritability of the disease. More recently, ~100 common, low-risk single nucleotide polymorphisms (SNPs) have been associated with BCA risk through genome-wide association studies (GWAS) and explain an estimated 16–20% of heritability [3]. The remaining familial clustering is likely driven by shared environment as well as epigenetic changes that may be influenced by both genetics and shared environment.

Epigenetic regulation, such as DNA methylation of BCA-associated genes [4–6], can change the activity of a DNA segment without changing the underlying DNA sequence and can be altered throughout life (reviewed in [7,8]). DNA methylation, occurs mostly at cytosine residues in CpG dinucleotides and has an important role in regulating gene expression and tumorigenesis [6]. DNA methylation alterations in BRCA1 have been observed in BCA tissues [4,5], and modest case-control differences in BRCA1 methylation of white blood cells (WBCs) have also been reported [9,10]. In addition, several genome-wide association studies (EWAS) using WBCs have found that low average epigenome-wide methylation was associated with higher risk of BCA [11,12]. In addition, Xu et al. have reported that a specific CpG methylation signature using WBC DNA improves risk prediction for BCA [13].

In addition, epigenetic factors modifying DNA methylation, a subset of CpG sites have strong associations with SNPs in cis [14–16], suggesting that DNA methylation patterns can be driven by genetic differences. Haplotype-dependent allele-specific DNA methylation (hap-ASM) is the physical counterpart of such mQTLs [14–18]. Thus, to accurately identify environmental effects on epigenetic alterations, it is critical to identify the extent that mQTLs and hap-ASM drive epigenetic differences across individuals. In addition, disease associations involving mQTLs and/or hap-ASM can provide useful insights to genetic risk [17].

Distinct from genomic imprinting, where the methylation of an allele is determined by its parent-of-origin, for loci with hap-ASM the local sequence context (haplotype) acts in cis to dictate the methylation status of local CpGs. Loci that show hap-ASM can be identified by bisulfitsequencing (bis-seq), in heterozygotes, if SNPs are present in the bis-seq amplicons. The closely related term mQTL (strictly, cis-mQTL) refers to loci in which the CpG methylation levels, which are typically scored in case series by microarray-based methods such as Illumina Methylation BeadChips, correlate strongly with genotypes at nearby SNPs, as assessed by SNP genotyping in the same series of individuals. Hap-ASM loci and mQTLs in linkage disequilibrium (LD) with disease-associated SNPs from GWAS can translate into modest but important methylation differences between low- and high-risk populations. From a cautionary standpoint, reviews of EWAS design have discussed the potential pitfall of falsely concluding that epigenetic changes are due to environmental pathways or disease progression, rather than genetic differences between cases and controls [19]. At the same time, the finding of hap-ASM in a haplotype block that also contains a relevant GWAS peak can be useful for providing biological support for that statistical peak, and for localizing regulatory DNA sequences within the haplotype block, which reveal themselves by conferring the physical asymmetry in methylation between the two alleles [17,18].

Measuring methylation in repetitive elements in WBC DNA, we previously found lower levels in young girls with a breast cancer family history (BCFH) compared to girls without a BCFH [20]. In order to identify specific differentially methylated regions (DMRs) associated with BCA susceptibility, the current study examined differences in WBC DNA methylation and hap-ASM between girls with and without a BCFH at 29 candidate loci using massively parallel targeted bis-seq.

**Results**

Table 1 shows descriptive statistics of LEGACY girls by BCFH and compares characteristics for those with and without blood specimens. Girls who provided blood were older (P < 0.0001), had higher body mass index (BMI; P < 0.0001) and were more likely to have a BCFH (P = 0.01) than those who did not. There were no differences in age and BMI by BCFH in girls who gave blood; 64.7% of the BCFH+ vs. 55.1% of the BCFH- girls were non-Hispanic White (P = 0.004).

**Identification of BCFH-DMRs in ESR1 and SEC16B**

The distribution of methylation in the 29 genes by BCFH is summarized in Table 2. Methylation in ESR1 was, on average, 4.9% (P = 0.003, FDR Q P = 0.08), higher in BCFH+ girls than BCFH- girls. The median values of average DNA methylation in ESR1 were 55.0% [interquartile range (IQR) = 21.6%] for BCFH+ girls and 49.2% (IQR = 25.2%) for BCFH- girls. (Table 2 and Figure 1). The ESR1 BCFH-DMR is in a downstream alternative promoter region, which shows dynamic chromatin states with histone marks characteristic of poised chromatin in H1-ESC, HepG2, and NHEK cells but repressed marks in GM12878 lymphoblastoid cells (Figure 2). The bottom panel of Figure 2 illustrates the variability in methylation across individual CpG sites for 10 individuals (5 with and 5 without a BCFH) with a shift in the distribution of methylation toward high net methylation in the BCFH+ group compared to the girls without a BCFH.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Baseline Blood Available</th>
<th>No Baseline Blood</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCFH+ girls</td>
<td>(n = 239)</td>
<td>(n = 187)</td>
</tr>
<tr>
<td>BCFH- girls</td>
<td>(n = 295)</td>
<td>(n = 319)</td>
</tr>
<tr>
<td>Age (y) – mean (SD)</td>
<td>10.9 (2.4)</td>
<td>10.5 (2.3)</td>
</tr>
<tr>
<td>Body Mass Index (kg/m²) – mean (SD)</td>
<td>18.9 (4.4)</td>
<td>18.7 (4.9)</td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>155 (64.7)</td>
<td>103 (55.1)</td>
</tr>
<tr>
<td>Hispanic White</td>
<td>57 (24.0)</td>
<td>38 (20.3)</td>
</tr>
<tr>
<td>Asian American</td>
<td>12 (5.0)</td>
<td>16 (8.6)</td>
</tr>
<tr>
<td>Pacific Islander</td>
<td>2 (0.8)</td>
<td>10 (5.3)</td>
</tr>
<tr>
<td>Mixed race/ethnicity</td>
<td>0 (0.0)</td>
<td>187 (100.0)</td>
</tr>
<tr>
<td>Breast Cancer Family History</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>0 (0.0)</td>
<td>187 (100.0)</td>
</tr>
<tr>
<td>First-degree</td>
<td>105 (43.9)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Only Second-degree</td>
<td>134 (56.1)</td>
<td>180 (60.0)</td>
</tr>
</tbody>
</table>
The mean of methylation in the DNA regions that we queried in \( SEC16B \) was 3.6% \((P = 0.026, \text{FDRQ} = 0.38)\) higher in girls with a BCFH than girls without a BCFH. (Table 2 and Figure 1). The median methylation levels in \( SEC16B \) were 75.1% for BCFH- girls and 72.8% for BCFH- girls. The statistically significant associations with \( ESR1 \) and \( SEC16B \) remained after adjusting for age, race/ethnicity, and BMI, and further adjustment by pubertal breast development stage and age at menarche. Although our main comparisons were between girls with and without a BCFH, we also calculated a continuous risk score. The Spearman correlation coefficients between BADINGA-CEA risk score, which includes both family history and any known BCA genes, and methylation were 0.11 \((P = 0.04)\) for \( ESR1 \) and 0.11 \((P = 0.06)\) for \( SEC16B \), respectively, supporting the positive overall association with the binary construct of family history.

The above two BCFH-DMRs showed high inter-individual variability in methylation levels, while methylation levels in other genes such as DLGAP2 and CCNL1 showed little variation, both overall and when compared by BCFH (Figure 1). Although there was high inter-individual variability of the methylation of \( CYP1A1 \), the levels of methylation did not differ by BCFH (Figure 1).

**BCFH-DMR and hap-ASM in \( SEC16B \)** suggests a candidate genetic variant associated with BCA susceptibility

Methylation of the BCFH-DMRs showed high inter-individual variability (Figure 1). Since our quality control metrics (Supplementary Methods) supported a biological, rather than technical, explanation for this inter-individual variability in methylation, we examined whether it might be explained by hap-ASM. The BCFH-DMR at \( SEC16B \) is in the upstream promoter/proximal enhancer region. The region shows dynamic chromatin marks associated with active enhancer in some cell types and quiescent chromatin in other cell types (Figure 3). The amplicon of \( SEC16B \) also covered the common SNP, rs6682862 (Figure 3), and thus could be assessed for hap-ASM (see Materials and methods). We found strong hap-ASM in \( SEC16B \) in 25 out of 32 heterozygous blood samples with significant absolute differences between the reference and alternate allele ranging from 20% to 75% (with bootstrapped Wilcoxon \( P \) value from 0.02 to \( 6 \times 10^{-5} \)). In addition, in all 25 samples with hap-ASM, the reference allele was methylated while the alternate allele was unmethylated, indicating hap-ASM and not genomic imprinting (for which a random, genotype-independent, pattern of ASM is expected) as the mechanism (Figure 3). This conclusion was confirmed by binomial test (binomial test \( P \) value = \( 3 \times 10^{-5} \)). To determine if our findings in WBC are relevant for breast tissue, we performed bis-seq in 10 normal breast tissue samples and 10 peripheral blood T lymphocyte samples (relevant to the host immune response to cancers), from different individuals, to test for hap-ASM in \( SEC16B \).

Hap-ASM was present in 3 out of 4 heterozygous breast samples and 2 out 3 heterozygous T cell samples. Although the difference between alleles was slightly smaller in breast and T cells than in whole blood WBCs, ranging from 21.4% to 23.7%, the direction of this difference was consistent across all the tissues and cell types (Figure 3 and Supplementary Figure 1). In addition, the net methylation was similar in breast and blood \([70.1\% \text{ (SD = 8)} \) compared to 70% \text{ (SD = 13.9)}\]. These findings indicate that the polymorphism in \( SEC16B \) has similar effects in whole blood, T cells, and whole breast tissues.

Methylation of \( SEC16B \) was negatively associated with age (Spearman’s correlation coefficient of -0.13, \( P = 0.02 \)). In multi-variable models adjusting for BCFH, age, and rs6682862 genotype, only genotype was independently associated with DNA methylation \((P \text{ value} = 3.5 \times 10^{-30}, R^2 = 0.40)\), suggesting that inter-individual variations in \( SEC16B \) methylation are mostly attributable to \( cis \)-acting effects of nearby SNPs (Figure 3).

In addition to \( SEC16B \), 9 additional loci with no evidence of BCFH-DMRs were potentially informative for hap-ASM analysis, namely \( BRCA1, CCDC85A, CYP1A1, DLGAP2, RAD51L1, SLC39A14, \) and \( ZNF483 \), and the intergenic regions queried on chromosome 12 and 16 (referred as Chr12 and Chr16 loci) (Supplementary Table 2). Bis-seq revealed hap-ASM in \( BRCA1 \) for 19 of 136 heterozygous samples and in \( CCDC85C \) for 13 of 97 heterozygous samples (Supplementary Table 2). The finding of hap-ASM in only a subset of heterozygous samples is similar to previous observations for other loci \([17,21]\) and can be explained by extended haplotype effects, in which a particular combination of more than one SNP is required in \( cis \) to confer significant ASM.

**Discussion**

Large epidemiologic studies have reported case-control difference in DNA methylation of WBC DNAs in BCA susceptibility gene such as \( BRCA1 \) in WBC DNAs \([9,10]\). In this study, we
compared the methylation status between girls with and without a BCFH to identify specific DMRs associated with BCA susceptibility. We observed two BCFH DMRs: one region located in a BCA susceptibility gene, encoding estrogen signaling receptor 1 (ESR1), and one region located in SEC16B, encoding a transport protein involved in growth and development. While both of these DMRs showed small but statistically significant differences in net methylation between girls with and without a BCFH, these differences were not statistically significant after a stringent test for false positives. However, because we selected our list of candidate loci based on a priori criteria, correcting for false positives may be unnecessarily conservative. In any event, our findings need to be further tested in replication studies.

Multiple CpG sites in the amplicon of ESR1 showed consistently higher mean methylation in girls with a BCFH compared to girls without a BCFH. ESRs are involved in BCA and studies on BCA cell lines have observed that methylation in a promoter or enhancer of ESR1 is associated with decreased expression of ERα mRNA and protein [23]. To assess the functional relevance of methylation in these regions, we first looked at the correlation between methylation and expression using public TCGA data. We used MEXPRESS [24] (http://mexpress.be/) to visualize and examine DNA methylation and expression data from breast cancers. Methylation measured by Illumina 450K BeadChips was negatively correlated to gene expression in ESR1; the Pearson’s correlation coefficients were -0.72 for cg00601836, -0.34 for cg15626350, and -0.37 for cg04063345. Thus, our

Figure 1. Methylation differences between BCFH+ and BCFH- girls in ESR1, SEC16B, CYP1A1, DLGAP2, and CCNL1. Net CpG methylation values (percent) in BCFH- girls and BCFH+ girls in 5 gene regions examined in WBC DNA. The red line represents the median methylation. ESR1, SEC16B, and CYP1A1 showed high inter-individual variability of the methylation level, while the distribution of methylation levels of DLGAP2 and CCNL1 did not. For the CYP1A1, DLGAP2 and CCNL1, there is no difference between BCFH+ and BCFH- girls, while for ESR1 and SEC16B, there are small but statistically significant differences. Importantly, the range of methylation values for these loci is wide, suggesting the presence of hap-ASM, in which one local haplotype acts in cis and dictates a low methylation level while another haplotype dictates high methylation.
observation of higher methylation in BCFH+ girls should indicate less expression of ESR1. This observation is consistent with decreased estrogen signaling associated with increased ESR1 methylation [25], recruiting the ESR1 protein to AhR and away from estrogen receptor (ER) target genes, increased ER degradation, and synthesis of other inhibitory factors [25].

Several SNPs associated with BCA that map to ESR1 have been reported in GWAS [22,26], and at least 5 SNPs, including rs9918437, have been found to be related to imbalances in ER expression [22]. The ESR1 amplicon we studied here is located within 58 kb of rs9918437. In addition, rs10913469, a GWAS SNP related to weight and BMI, is located within a 150 kb window [27,28]. Although there are no informative SNPs in the ESR1 amplicon and LD data are not available in the 1000 genomes project for the closest common SNPs [29], hap-ASM might be present, with the hap-ASM index SNP being in LD with these GWAS SNPs. Using an array-based indirect approach, Day et al. identified mQTLs in blood lymphocytes at cg04063345 and cg15626350, which are located in our DMR, suggesting the presence of hap-ASM in this ESR1 region [30]. In addition, mRNA levels differ between two ESR1 alleles marked by rs2077647, which is about 1.2 kb 5’ of our amplicon [31]. Environmental exposures might also contribute to inter-individual variation in methylation of ESR1 [32,33]. For

Figure 2. BCFH-DMR in an alternate promoter of ESR1. Map and bis-seq showing the BCFH-DMR in an alternate promoter of ESR1. Zoomed-in view of ESR1, showing relevant ENCODE tracks and the amplicons utilized for targeted bis-seq. The BCFH-DMR is located in a downstream ESR1 alternative promoter region and marked by poised chromatin marks. Graphical representation of bis-seq data for 10 examples of BCFH- girls (left) and BCFH+ girls (right) are shown. Each column represents a consecutive CpG covered by the amplicon, and each line a unique DNA fragment. White circles are unmethylated CpGs and black circles are methylated CpGs.
Figure 3. Hap-ASM BCFH-DMR and mQTLs in the promoter of SEC16B. Map and results of bis-seq showing the BCFH-DMR and hap-ASM in the promoter of SEC16B. (A) Map of SEC16B, showing relevant ENCODE tracks and the amplicons utilized for targeted bis-seq. (B) Bis-seq data showing hap-ASM in the SEC16B promoter region. Graphical representation of 5 representative whole-blood samples and 1 human normal breast tissue. Additional representations in breast tissues and T cells from peripheral blood are shown in the Supplemental figure 1. This region overlaps with the common SNP, rs6682862. Allele A and B are analyzed and represented separately. The
example, an animal study reported that, in utero, bisphenol A (BPA) exposure disrupted epigenetic programming of ESR1 in the brain [32]. DNA methylation alterations in multiple CpG sites of ESR1 were also reported to be correlated with maternal smoking during pregnancy and offspring's methylation patterns in ESR1 [33].

The second BCFH-DM we observed is in SEC16B, a gene associated with obesity and age at menarche [34,35]. A GWAS peak SNP associated with menarche and obesity, rs633715, and one for BMI, rs10913469, are located 85 kb and 25 kb upstream of the hap-ASM region, respectively [28], although these SNPs appear to lie just outside of the haplotype block that contains the hap-ASM index SNP, rs6682862. Interestingly, allele-specific expression (assessed by eQTL) of SEC16B has been described in peripheral blood monocytes [36]. The BCFH-DMR at SEC16B is located in the gene promoter. In the TCGA data, methylation at cg00647232, which is covered by our bisseq ampiclon, showed a significant positive correlation to gene expression (Pearson’s coefficient = 0.09). Our findings of hap-ASM in this locus in blood and breast tissues, together with multiple eQTLs associated with SEC16B expression [37], suggest that the small but significant DNA methylation difference that we have observed in this gene between the BCFH+ and BCFH- girls could manifest functionally as differences in genetic susceptibility. Although no SNP in SEC16B has yet been reported as linked to BCA susceptibility, the presence of eQTLs and hap-ASM in this gene nominates it as a candidate for further scrutiny, which could be done by genetic fine-mapping and testing for associations with BCA after taking age at menarche and BMI into account. Importantly, our breast tissue samples are from whole excisions and contain not only epithelial cells but also fibroblasts, endothelial cells, and adipocytes. Future work on isolated breast epithelial cells, and on purified adipocytes and fibroblasts, will be necessary to gain insight on the underlying mechanisms of SEC16B associated susceptibility. The analysis of hap-ASM in adipocytes could be particularly important since this gene has been associated with obesity by GWAS and, indeed, the effect on breast cancer risk may be biologically linked to adiposity [38]. In fact, we directly examined methylation at SEC16B locus in adipose tissue using ENCODE whole-genome bisulfite sequencing generated by the ENCODE consortium (https://www.encodeproject.org/). Although only 3 samples were sequenced, the high inter-individual variability in the methylation level and pattern is suggestive of hap-ASM. Regarding the functional relevance of hap-ASM in T cells, results from recent clinical trials using immune checkpoint inhibitors in multiple cancers, including BCA [39], support an active role of immune infiltration in cancers.

We downloaded TCGA public data from the cancer browser (https://genome-cancer.ucsc.edu/). We compared methylation between the 22 BCA bearing mutation in BRCA1 or BRCA2 and 630 cancers without any of these mutations. For ESR1, we did not observe a significant effect of BRCA1/2 mutations on methylation (difference between mutated vs. non-mutated cancers = -2.1%, P = 0.2). In addition, this non-significant trend was in the opposite direction to the one observed in BCFH+ vs. BCFH- comparison, suggesting that our findings for ESR1 is unlikely due to a higher prevalence of BRCA1/2 mutation in BCFH+ compared to BCFH- girls. For SEC16B, we observed a non-significant relative hypermethylation in the BRCA1/2 mutated vs. non mutated cancers (7%, P = 0.12). Although this hypermethylation was not significant, we cannot exclude that part of the difference that we observed in BCFH+ vs. BCFH-comparison is driven by the unbalanced distribution of BRCA1/2 mutations in our groups. However, although we expect a higher prevalence of these mutations in BCFH+ girls, the low prevalence of BRCA1 and BRCA2 in the general population (1.56% and 1.87% in TCGA data, respectively) suggests, if present, a modest confounding effect. We acknowledge that future work should assess the effect of BRCA1/2 mutations on net methylation difference as well as on the presence and strength of hap-ASM in SEC16B.

We observed possible hap-ASM in BRCA1 and CCDC85C; however, since the magnitude of the hap-ASM was variable and the proportion of samples showing it was low for these genes, further exploration of these loci is required. Other loci that showed a wide range of methylation but lacked informative SNPs in the ampiclons (e.g., ESR1) will also need to be examined for possible hap-ASM using additional ampiclons. We believe that a non-negligible part of differentially methylated loci identified by large sample size case/control methylation studies (EWAS) are in fact hap-ASM dictated by disease-associated SNPs, the difference between case and control reflecting the unbalanced distribution in cases vs. controls of these disease associated SNPs [21]. Therefore, mapping hap-ASM can help prioritizing disease-associated SNPs and unravel biological pathways disrupted by genetic variants.

A key strength of our study is that it is enriched with girls with a BCFH, permitting identification of genes or loci for which dysregulated methylation might partially explain family clustering of cancer. Using bis-seq, we assessed ASM and examined the potential genetic effect on the BCFH-DM. Advantages of the bis-seq method that we used here are the ability to examine methylation across multiple CpGs and to separate the sequencing reads by allele. An important limitation with using WBC DNA is that both net methylation and hap-ASM can be tissue- and cell type-specific [17]. However, an advantage of using
WBC DNA is that it can be collected repeatedly, particularly across development. For example, using a repeated sample design of 51 children before and after puberty, Almstrup et al. [40] identified methylation changes associated with puberty in a total of 457 CpG sites, including 5 CpG sites located in the promoter of thyroid receptor-interacting protein (TRIP6), a gene involved in cell signaling, cell invasiveness and transcriptional activation. We did not examine this gene here but plan on examining repeated measures of DNA methylation pre- and post-puberty in the LEGACY girls. Other types of regulatory sequences will also need to be examined; for example, higher methylation variability in dysregulation in DNA methylation of BCA genes, such as ATM in WBC DNA was observed in intragenic repetitive elements rather than in promoter regions [41]. Although we selected the list of candidate genes based on a priori considerations including pilot data using genome-wide DNA methylation platform, if we were to conservatively divide the P value by the number of tests we conducted, our results for SEC16B may be due to chance. Thus, although our findings suggest a potential role for mQTLs in genes important to BCA, they need to be replicated in independent series. Here, we examined an a priori list of targeted genes based on criteria described in the methods. Our results suggest that larger scale of screening of mQTLs/hap-ASM may be fruitful for characterizing differences that may be evident early in life, and may be partly genetically determined, in individuals with a cancer family history.

We recognize that the abundance of specific cell subtypes, which may have different levels of methylation in the genes of interest, likely impacts our results. However, we evaluated differences in methylation related to BCFH and, even if differences are partially the result of large changes in methylation of a small subset of cells, we have no reason to expect that these subsets differ by BCFH. We examined whether the ESR1 region where we found differences in methylation could be explained by cell type differences by examining publicly available methylation data (GEO accession: GSE35069) and did not find any differences in methylation by cell type, including WBC, peripheral blood mononuclear cells, monocytes, and neutrophils [42]. Our findings, if replicated, suggest that methylation in BCA related genes may differ between individuals with and without a BCFH, in part due to genetic-epigenetic interactions. To understand whether genetic or environmental factors early in life may be associated with changes in DNA methylation, prospective studies are needed that measure within-individual changes in exposures and DNA methylation over time and that carefully consider the role of genetic variation in explaining the inter-individual epigenetic variation.

Materials and methods

Study participants

Participants were girls ages 6–13 years participating in the LEGACY Girls Study, which is a multicenter prospective study following 1,040 girls and their participating parent or guardian, with data collection every six months on clinical and epidemiological factors (for details see www.legacygirlsstudy.org and [43,44]). Girls were invited to provide biospecimens, including blood at baseline and at follow-up visits. Almost half of the girls have a BCFH, defined as at least one first- or second-degree relative diagnosed with BCA. Forty-one percent of the girls (n = 426) provided blood at baseline. We also collected demographic and epidemiological information, including family history of breast cancer, by interviews, and anthropometric measurements (height and weight) using standardized protocols. Because we have already reported differences between blood and saliva DNA methylation [45], we did not include girls who only provided saliva at baseline (n = 509). All participating institutions obtained Institutional Review Board approval to conduct the study.

Bis-seq for DNA methylation analysis

The selection criteria we used to develop a panel of DNA methylation levels in 29 candidate loci was based on published genetic or epigenetic association, with at least one or more of the following categories: (i) genes that are associated with GWAS SNPs or mutations related to BCA risk [46,47]; (ii) genes related to age at menarche from GWAS [34]; (iii) genes related to growth and development from GWAS [48] or EWAS [49]; (iv) genes involved in DNA recombination and repair (http://sciencepark.mdanderson.org/labs/wood/dna_repair_genes.html); and (v) the top 8 candidate CpG sites that showed evidence of methylation changes between girls with and without a BCFH in our exploratory genome-wide DNA methylation profiling from pilot work in 48 girls, but outside the above selection criteria (Supplementary Table 1). For each locus, we selected the bis-seq primer locations based on the chromatin states defined by Ernst et al. [50] and available in the UCSC human genome browser [51]. We focused on active promoters and enhancers, insulators, and poised chromatin, since these regions are implicated in gene regulation in cancers and are often enriched in disease-associated DMRs and mQTLs/hap-ASM [17,52]. In addition, we covered a few loci with repressed chromatin or transcription associated chromatin states. When indicated, annotation of GWAS peaks was performed using the NHGRI-EBI catalog [27] and literature searches for smaller-scale genetic association studies.

We examined DNA methylation for an amplicon (approximately 200–300 bp) for these 29 selected loci using massive parallel targeted bis-seq on genomic DNA from the WBC samples, as described [17]. Unlike Methylation BeadChips, bis-seq provides a qualitative and quantitative approach to identify 5-methylcytosine at single base-pair resolution over multiple contiguous CpG sites, as well as the potential to assess methylation level after separation by alleles [17,53]. We also measured DNA from breast normal tissues in 10 women: 8 from adjacent non tumor tissues, 1 from breast reduction, and 1 from benign tumor and 10 T cells from adults without BCA [17]. Details of the DNA methylation measurements are in the Supplementary Methods. We calculated methylation percentages by averaging across all CpGs in the amplicon.

Statistical methods

Identification of differentially methylated regions (DMRs) associated with BCFH

To assess BCFH-DMRs, we compared average methylation difference for each amplicon between girls with and without a
BCFH using the Wilcoxon rank-sum test. We calculated false discovery rate (FDR) q value using PROC MULTITEST procedure. For each DMR, we carried out a multivariable analysis using linear regression models with DNA methylation percentage as the outcome, and categorical variables for BCFH, race/ethnicity, and pubertal development, and continuous variables for age and body mass index [BMI; weight (kg)/height squared (m)] as covariates. Using the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) risk model [54], we previously had estimated the absolute risk based on family pedigree information for each girl [55]. We also examined the correlation between BCFH-DMRs and the BOADICEA risk model. In addition, for loci with available genotyping data, SNP genotype was also included in the model (see genotyping method below). We set the significance level 2-sided P value<0.05.

Identification of hap-ASM
To identify hap-ASM we used an approach similar to that in Do et al. [17] Genotype calling for informative SNPs in the bis-sequencing reads. Overall, 10 amplicons covered at least one informative SNP and were included for hap-ASM analysis (Supplementary Table 2). Details of assessing hap-ASM are provided in the Supplementary Methods. All analyses were performed with SAS software 9.4 (SAS Institute, Cary, NC) and R 3.10.

Acknowledgments
The authors thank the LEGACY girls and their family members for their continuing contributions to the study and our colleagues at the participating family genetics and oncology clinics. The authors thank the Epigenetics Shared Resource and the Molecular Pathology Share Resource of the Herbert Irving Comprehensive Cancer Center for performing the targeted bisulfite sequencing and the Columbia University Tissue Bank for normal breast tissue. We also thank the staff at each LEGACY study site, including Alex Blacker, Connie Cady, Jocelyn Koo, Lisa Moy, Meera Sangaramoorthy, Enid Satariano (California site); Ann Johnson, Jasmine MacDonald, Melissa White (New York site); Gordon Glendon, Danielle Hanna, Teresa Selandor (Ontario site); Colleen Sands (Philadelphia site); and Briar Doi, Karen O'Toole (Utah site). The authors also thank the members of the Scientific Advisory Board (Drs. Tim A. Ashles, Christopher K. Daugherty, Michele Forman, Marcia E. Herman-Giddens, Robert Hiatt, Lawrence H. Kushi, Mary S. Wolff, Susan L. Teitelbaum, Martin J. Yaffe) and the Event Monitoring Committee (Drs. Jeffrey Botkin, Christopher K. Daugherty, Branylin DeRosa, Malcolm Pike, and Ms. Cheryl Schuman). The authors thank the contributing clinical centers for their efforts in recruiting our LEGACY participants (Ontario site: Clinical Genetics at Trillium Health Partners – Credit Valley Hospital, The Cancer Risk Assessment Centre at the Juravinski Cancer Centre, The Princess Margaret Hospital Familial Breast and Ovarian Cancer Clinic, The Mount Sinai Familial Breast Cancer Clinic and The Granovsky Gluskin Family Medicine Centre of Mount Sinai Hospital; Philadelphia site: the Clinical and Translational Research Center (CTR) [UL1TR000003 (NCATS) and UL1RR024134] and the Pediatric Research Consortium (PeRC) at The Children’s Hospital of Philadelphia.

Disclosure of Potential Conflicts of Interest
The authors declare that they have no competing interests.

Funding
This work was supported by awards from the National Cancer Institute [R01CA138822 and P30 CA013666] and the National Institute of Environmental Health Sciences [P30 ES009089] and the Breast Cancer Research Foundation. The LEGACY Girls Study is funded by the National Cancer Institute at the National Institutes of Health (Grants CA138638 to E. M. John, CA138819 to M. B. Daly and A. R. Bradbury, CA138822 to M. B. Terry, and CA138844 to I. L. Andruski) and the Canadian Breast Cancer Foundation (I. L. Andruski). The work at the New York site was also supported by the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant Number UL1 TR000040, formerly the National Center for Research Resources, Grant Number UL1 RR024156. Support was also provided by a grant from the Avon Foundation for Women, to M.B. Terry and B. Tycko. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References


