Published in final edited form as:

Epigenomics. 2013 April ; 5(2): 155–165. doi:10.2217/epi.13.6.

Real-time methylomic aberrations during initiation and progression of induced human mammary epithelial cell tumorigenesis

Natalie E Mitchell¹, MacKenzie L Wilson¹, Molly S Bray², David K Crossman³, and Trygve O Tollefsbol^{*,1,4,5,6,7}

¹Department of Biology, University of Alabama at Birmingham, 1300 University Boulevard, Birmingham, AL 35294, USA

²Department of Epidemiology, University of Alabama at Birmingham, 1665 University Boulevard, Birmingham, AL 35294, USA

³Department of Genetics, University of Alabama at Birmingham, 720 20th Street South, Birmingham, AL 35294, USA

⁴Center for Aging, University of Alabama at Birmingham, 1530 3rd Avenue South, Birmingham, AL 35294, USA

⁵Comprehensive Cancer Center, University of Alabama at Birmingham, 1802 6th Avenue South, Birmingham, AL 35294, USA

⁶Nutrition Obesity Research Center, University of Alabama at Birmingham, 1675 University Boulevard, Birmingham, AL 35294, USA

⁷Comprehensive Diabetes Center, University of Alabama at Birmingham, 1825 University Boulevard, Birmingham, AL 35294, USA

Abstract

Aim—Neoplastic transformation provides one of the few existing opportunities to analyze molecular changes in real time during the initiation and progression of breast cancer.

Materials & methods—Human mammary epithelial cells underwent neoplastic reprogramming, generating one line of semitransformed, premalignant cells and two separate, temporal lines of fully transformed human mammary epithelial cells (THMECs). An Illumina Infinium HumanMethylation27 BeadChip was used to analyze DNA methylation alterations in 27,578 CpG loci at three consecutive time points over an 80-day (d) transformation period.

Results—The mean β value for semitransformed human mammary epithelial cells CpG loci (0.245) was much greater than for either THMEC-40d (0.055) or THMEC-80d (0.066), indicating a large loss of methylation after neoplastic induction. In addition, 54% of CpG loci were

© 2013 Future Medicine Ltd

*Author for correspondence: Tel.: +1 205 934 4573, Fax: +1 205 975 6097, trygve@uab.edu.

Financial & competing interests disclosure

The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

hypermethylated during the THMEC-40d to THMEC-80d transition. We observed that the CpG loci exhibiting DNA methylation changes during early oncogenesis were enriched for biological functions like cellular movement; this was distinctly different than in the later, more progressive stages of the transformation process enriched for processes involving differentiation.

Conclusion—The timing of major methylomic changes may be important in directing the cell toward a more cancerous phenotype. In addition, gene-specific hypermethylation appears to silence developmentally related genes, leading to dedifferentiation.

Keywords

breast cancer; DNA methylation; epigenetics; Illumina Infinium; HumanMethylation27 BeadChip; neoplastic reprogramming

Breast cancer is the second leading cause of cancer deaths among women; however, if detected early enough, the survival rate for the disease is relatively high [1,101]. Although our knowledge of cancer has increased dramatically in recent years, the process of breast oncogenesis is still poorly understood. This lack of comprehension stems partly from the limited ability to analyze the disease in its very beginning stages during the transition from a normal cell to its neoplastic derivative. Isolating the genetic requirements necessary for the initial acquisition of a cancerous phenotype can provide novel insight into the disease and highlight potential ways to diagnose and treat breast cancer during its early development. While it is not possible to study tumorigenesis *in vivo*, the process of neoplastic reprogramming may be employed to examine tumorigenesis in real time.

Weinberg and colleagues accomplished oncogenic transformation through the viral-mediated serial gene transfer of three defined elements – *SV40 ER*, *hTERT* and *hRAS-V12* – to generate cancer cells from normal cells [2]. Although the procedure has been completed in epithelial and fibroblast cells [2,3], it has also been extended to generate several other cancer types, including breast cancer, through the use of human mammary epithelial cells (HMECs) [4,5]. While these individual genetic mutations are not universally found in all types of breast cancer, the model can closely approximate the progression of the disease by mimicking the genetic and cellular requirements for the development of breast cancer, including disruption of the p53 and pRb pathways and upregulation of Ras-associated pathways. In addition, this established model of breast tumorigenesis enables us to analyze the estrogen receptor (ER)-negative breast cancer subtype, the most lethal form of breast cancer, as HMECs lack expression of the ER [5].

We chose three time points in the neoplastic progression of breast cancer to examine alterations to the genome and to identify molecular changes that may play a significant role in breast carcinogenesis. Using this model system, we have previously identified both genes and proteins associated with an increase in metastatic potential that displayed differential expression throughout the course of an 80-day (d) transformation period [6]. Since changes to the epi-genome contribute to the cancerous phenotype by altering the regulation of gene and protein expression [7,8], we elected to also investigate the epigenetic regulatory mechanisms occurring during the initiation and progression of human mammary cell neoplastic induction.

DNA methylation has been shown to play a crucial role in defining the cell type and maintaining both the normal and diseased state [9–11]. However, compared with their normal cell counterparts, cancer cells display abnormal patterns of DNA methylation, generally characterized by widespread hypomethylation and site-specific hypermethylation [12]. Knowledge of specific DNA methylation aberrations during breast tumorigenesis may

provide potential targets for developing therapeutic strategies to treat the disease in its earliest stages.

Few studies have focused on analyzing global alterations in DNA methylation during the onset of breast oncogenesis, mainly due to the limited availability of models that mimic the disease in real time. The goal of this study was to complete a novel genome-wide investigation of changes in DNA methylation patterns using the established method of neoplastic reprogramming as a model for ER-negative breast cancer, and to examine the timing and nature of methylomic changes during the transition from a normal to cancerous breast cell.

Materials & methods

Generation of the neoplastic reprogramming model & cell culture

Semitransformed HMECs (SHMECS; HMECs transformed with *SV40 ER* and *hTERT*) and transformed HMECs (THMECs; SHMECs transformed with *hRAS-V12*) were generated and maintained as previously described [5,6]. SHMECs are premalignant breast cells, while THMECs are fully transformed and malignant breast cancer cells, capable of producing tumors *in vivo* [6]. For each cell line, four biological replicates were generated.

DNA extraction

DNA was harvested with the DNeasy Blood and Tissue Kit as per the manufacturer's protocol (Qiagen; CA, USA). DNA was suspended in distilled water and stored at -80°C .

Genome-wide methylation analysis

Methylomic analysis was performed using the HumanMethylation27 BeadChip and iScan system from Illumina, Inc. (CA, USA). Briefly, following bisulfite treatment (Zymo, Inc.; WA, USA), genomic DNA was amplified, fragmented and then hybridized to the BeadChip. A single-base pair extension reaction was performed, followed by staining with a detectable fluorescent dye label. The resulting products were imaged with the iScan Reader (Illumina, Inc). Quality standards for bisulfite conversion, staining, single-base extension, hybridization, stringency and nonspecific binding were verified. The level of methylation was determined at each CpG locus by the intensity of two possible fluorescent signals, one from the C (methylated) allele and one from the T (unmethylated) allele, using GenomeStudio Software (Illumina, Inc.). The relative methylation value for each CpG locus was expressed as a β value. The β value was calculated by dividing the average signal for the methylated probe by the total signal for intensity [13]. Data were quantile normalized and β values were represented on a scale from <0 (completely unmethylated) to >1 (completely methylated). All negative β values were set to 0 [14]. Please note that the HumanMethylation27 BeadChip is not extended to CpG shores.

Statistical analysis

Figure 1 displays a schematic overview of the experimental design, including data stringency filters. Principal component analysis (PCA) was completed to assess the clustering and similarity of sample types. Mean β values were calculated for experimental groups (SHMEC, THMEC-40d and THMEC-80d) by averaging the β values for four replicates at each time point. Comparison of group β value means was performed using an analysis of variance with Benjamini–Hochberg multiple-testing correction (false discovery rate corrected $p < 0.01$), generating a list of CpG loci ($n = 1996$) with statistically different β values [15]. *Post hoc* pairwise comparisons between cell lines were performed using a Tukey's honestly significant difference procedure. Three separate lists of differentially

methyated CpG loci were of interest: SHMEC versus THMEC-40d; SHMEC versus THMEC-80d; and THMEC-40d versus THMEC-80d – each representing the time-point transition during which the differential methylation was observed. A biological filter of an absolute methylation level difference ($|\Delta\beta|$) of at least 0.20 ($|\Delta\beta| > 0.20$), corresponding to at least a 20% difference in methylation for a CpG site between a given time-point transition, was then applied to each of the gene lists generated from the pairwise comparison [13]. Venn diagrams were created to assess the similarity of differentially methylated CpG sites between the gene lists from each pairwise comparison. All statistical analyses, including PCA and Venn diagrams, were carried out using GeneSpringGX Version 11.5.1 (Agilent Technology; CA, USA).

Ingenuity Pathway Analysis

Core bioinformatics analyses were performed using Ingenuity Pathway Analysis (IPA; Ingenuity Systems [102]). IPA contains a curated database of networks and biological relationships based on original peer-reviewed articles. Data sets containing Human Gene Organization gene identifiers for CpG loci differentially methylated only between SHMEC to THMEC-40d ($n = 805$) or THMEC-40d to THMEC-80d ($n = 196$) were uploaded and analyzed separately using IPA software. The functional analysis identified biological functions that were most significant ($p < 0.05$) to each data set.

Validation of DNA methylation status with Pyrosequencing® analysis

Primers were designed for randomly selected CpG sites (Supplementary Material 1; see www.future-medicine.com/doi/suppl/10.2217/epi.13.6) that passed all previously described filters and were identified as differentially methylated either from SHMEC to 40d or 40d to 80d, including *ZNF577* (cg16731240) and *HMG3* (cg13176979), respectively, using PSQ Assay Design software (Biotage; Uppsala, Sweden) and validated using bisulfite Pyrosequencing. Successful validation is based on the strength of the correlation coefficient, r [16,17]. The Target ID for each CpG loci is indicated in parentheses. The Pyrosequencing reaction and quantitation of DNA methylation was performed using the PyroMarkHS96 sequencer and PyroMark MD software (Qiagen; CA, USA).

Results

PCA

Normalized DNA methylation data were assessed for four biological replicates for every time point through PCA to evaluate the epigenetic profile of each cell type. As shown in Figure 2, SHMEC, THMEC-40d and THMEC-80d appear to have distinct DNA methylation profiles as PCA separated each cell type onto a different axis. However, the THMEC-80d samples tend to cluster closer toward the THMEC-40d axis, indicating that the two fully transformed cell lines have more similar methylation profiles than when individually compared with SHMEC. This would suggest that the quantity of DNA methylation changes is more abundant during the SHMEC to THMEC-40d transition than the THMEC-40d to THMEC-80d transition. In addition, the slight difference between THMEC-80d sample clustering is likely due to the heterogeneous nature of the disease progression.

Identification of statistically different β values

An analysis of variance with Benjamini–Hochberg multiple testing correction (false discovery rate corrected $p < 0.01$) was used to identify CpG loci that had statistically different β values over the course of neoplastic induction [15]. The total average β value for these CpG loci ($n = 1996$) was calculated for each cell line (Figure 3). SHMEC had the highest total average β value approximately 0.245, with THMEC-40d and THMEC-80d

having significantly lower mean β values of 0.055 and 0.066, respectively. Using a paired t-test, the mean β value for SHMEC CpG loci was much greater than for either THMEC-40d ($p < 0.0001$) or THMEC-80d ($p < 0.0001$). The difference in average β value between THMEC-40d and THMEC-80d was minimal compared with the difference between SHMEC and THMEC-40d, suggesting the loss of promoter region DNA methylation as an early event during oncogenesis.

Venn diagram analysis of differentially methylated CpG loci

The list of 1996 CpG loci was subjected to *post hoc* pairwise comparisons between cell lines using a Tukey's honestly significant difference *post hoc* test procedure. A biological filter was then applied, eliminating CpG loci without at least a 20% difference in $|\Delta\beta|$ value between a given time-point transition. Figure 1 displays the number of CpG sites identified as differentially methylated for each pairwise comparison. Venn diagrams assessing the similarity of differentially methylated sites between data sets are shown in Figure 4. Of the 1033 CpG loci differentially methylated from SHMEC to 80d, 804 were in common with the SHMEC versus THMEC-40d loci list and 195 were shared with the THMEC-40d versus THMEC-80d list. This extreme overlap (97%) indicated that the timing of functional methylation changes can be narrowed to occur either during the transition from SHMEC to 40d or 40d to 80d as opposed to throughout the entire transformation period from SHMEC to 80d. Thus, subsequent analyses were focused on the comparison of SHMEC versus THMEC-40d and THMEC-40d versus THMEC-80d loci lists.

Assessment of hypo- & hyper-methylation & CpG site location

A total of 1010 CpG sites were differentially methylated from SHMEC to THMEC-40d (Figure 1). Of these, 844 CpG sites (84%) were hypomethylated and 166 sites (16%) were hypermethylated. In addition, 358 CpG sites were differentially methylated from THMEC-40d to THMEC-80d. Of these, 160 (46%) were hypomethylated and 198 (54%) were hypermethylated. Because the microarray contains both canonical CpG islands and additional CpG dinucleotide loci within gene promoters, we assessed a possible distribution pattern of CpG site location and stage of breast oncogenesis by comparing the methylation status of CpG sites located within and outside of CpG islands. Table 1 displays the percentage of differentially methylated CpG sites found within and outside of CpG islands for hypomethylated and hypermethylated sites, respectively. During the SHMEC to THMEC-40d transition, 65% of hypomethylation occurred within CpG islands and 35% outside CpG islands, while hypermethylation displayed a near-even distribution of sites within CpG islands (53%) and outside (47%). From 40d to 80d, hypomethylation was detected inside CpG islands at a value of 53%, compared with a value of 47% for sites outside CpG islands. By contrast, hypermethylation was mainly found within CpG islands (82%).

Investigation of methylation trends

We further examined the DNA methylation patterns across the time points from baseline to 80d following induction of tumorigenesis (Figure 5). For those loci that were differentially methylated between SHMEC and THMEC-40d ($n = 1010$), more than 80% ($n = 805$) were hypomethylated from SHMEC to THMEC-40d and exhibited no methylation change ($|\Delta\beta| < 0.10$) between THMEC-40d and THMEC-80d (Figure 5A). Only six loci demonstrated continual hypomethylation at all time points. The remaining loci that were differentially methylated between SHMEC and THMEC-40d displayed contrasting patterns of methylation across the 80d experiment, with 166 loci (16%) initially hypermethylated from SHMEC to 40d, followed by hypomethylation from 40d to 80d, and 33 loci demonstrating

the opposite pattern of methylation. Interestingly, no loci followed a pattern of hypermethylation from SHMEC to 40d and no methylation change between 40d to 80d.

When examining the methylation patterns of those genes differentially methylated between THMEC-40d and THMEC-80d ($n = 396$, Figure 5B), 196 loci (60%) exhibited no methylation change from SHMEC to 40d and were hypermethylated from 40d to 80d. The remaining loci differentially methylated between 40d and 80d demonstrated hypermethylation from SHMEC to 40d, followed by either hypomethylation ($n = 160$) or continued hypermethylation from 40d to 80d ($n = 2$). Zero loci demonstrated no methylation change between SHMEC to 40d and hypomethylation from 40d to 80d. These results suggest that a specific timing mechanism may exist for major hypo- or hyper-methylation changes in the genome during the onset and progression of breast cancer. We elected to focus the remainder of our study on CpG loci exhibiting differential methylation during only one time-point transition over the course of the 80d transformation period.

Ingenuity Pathway Analysis

We used IPA software to evaluate the relevance of CpG loci displaying a single DNA methylation change with breast cancer. SHMEC versus THMEC-40d ($n = 805$ CpG loci, Supplementary material 2) and THMEC-40d versus THMEC-80d ($n = 196$ CpG loci, Supplementary material 3) gene lists, containing Human Genome Organization gene identifiers, were uploaded and analyzed separately in IPA. Figure 6 displays the top five molecular and cellular function categories representative of each gene list. The analysis indicates early differential methylation occurs in genes involved in such cellular processes as cellular movement and cellular growth and proliferation. Genes exhibiting later methylation changes from THMEC-40d to THMEC-80d were enriched for biological functions like tissue and cellular development.

Validation of DNA methylation β values by Pyrosequencing

Because prior studies have indicated a good correlation exists between the β values generated by the Illumina Infinium HumanMethylation27 BeadChip and percent methylation data from Pyrosequencing, we randomly selected two genes exhibiting differential methylation for validation of our microarray data [17,18]. *HGMN3* ($r = 0.952$) and *ZNF577* ($r = 0.825$) both displayed strong correlations ($r > 0.75$) between array data and percent methylation (Figure 7). This is not an exhaustive Pyrosequencing analysis but does validate DNA methylation data generated from the Illumina array. It is also important to note that Pyrosequencing can under-represent DNA methylation values, which may account for the difference in percent Pyrosequencing methylation levels as compared with those generated by the Illumina array [16].

Discussion

Although a number of genetic mutations have been linked to breast cancer, epigenetic modifications have also been shown to significantly influence the development of tumors [12,19]. In particular, aberrant changes in DNA methylation often contribute to alterations in the normal regulation of gene expression, permitting the development of a cancerous phenotype [12,20–22]. Few studies have been completed that assess epigenetic changes during the initiation of oncogenesis as it is impossible to determine when a cell will begin transforming from a normal to neoplastic state *in vivo*. However, the method of neoplastic reprogramming, which mimics the activation and inactivation of pathways commonly disrupted during breast cancer, may be used as a cell model of the disease [5]. We have chosen three time points in the neo-plastic progression to breast cancer to identify and analyze changes in DNA methylation patterns. Although we are aware that the use of three

time points is not reflective of the entire carcinoma process, because both the sequence and timing of genome-wide epigenetic modifications are crucial in understanding the progression of breast cancer, we feel the use of neoplastic transformation in a time-specific manner does provide the best existing opportunity to analyze the disease in real time *in vitro*.

Our results reveal distinct DNA methylation changes seen following the addition of *hRAS-V12* over the 80d transformation period. PCA separated each cell type onto a different dimensional axis, demonstrating the existence of distinct epigenetic profiles at various stages of tumorigenesis. THMEC-80d samples tend to cluster toward the THMEC-40d axis, which indicates an increased similarity between these DNA methylation profiles than when compared with the SHMEC line. In addition, this suggests that functional epigenetic alterations that occur during the early stages of breast cancer may be maintained during the progression of the disease.

We identified 1996 CpG loci displaying statistically different β values over the course of neoplastic induction. SHMEC had the highest total average β value at approximately 0.245, with THMEC-40d and THMEC-80d having significantly lower mean β values, approximately 0.055 and 0.066, respectively. Here, we show a large loss of DNA methylation between SHMEC to THMEC-40d, implicating the global loss of methylation during the initial transition to an oncogenic state and persistence throughout the progression of the disease. Among varying solid tumor types, global DNA hypomethylation occurs most frequently in breast cancer [23–25]. Markedly reduced DNA methylation levels have been observed in several established breast cancer cell lines, including MCF-7 and MDA-MB-231, compared with normal HMECs [26]. It is important to note that others have also reported similar trends in primary invasive breast carcinoma tissue samples, indicating a higher frequency of hypomethylation in ER-negative breast tumors as compared with ER-positive tumors, suggesting that the reduced methylation may account for the increased aggressiveness of this cancer subtype [27].

We then applied a stringent $|\Delta\beta|$ biological filter and determined that 1010 CpG loci were differentially methylated from SHMEC to THMEC-40d compared with 358 differentially methylated sites during the transition from THMEC-40d to THMEC-80d. Categorizing whether the CpG sites were hypo- or hyper-methylated over the course of transformation, as well as the loci locations, revealed distinct trends. Significantly, 84% of sites were hypomethylated from SHMEC to THMEC-40d, suggesting that the loss of promoter methylation is a widespread occurrence during the initiation of breast cancer. Others report similar loss of DNA methylation in CpG islands during induced epithelial–mesenchymal transition in MCF-7 breast cells [26]. In addition, 54% of sites were hypermethylated from THMEC-40d to THMEC-80d. Over 80% of this hypermethylation occurred within CpG islands, which is consistent with previous reports using both breast cell lines and tissue samples [26,28,29]. The major methylation trend observed during the SHMEC to THMEC-40d transition was hypomethylation, while hypermethylation of CpG loci occurred from THMEC-40d to THMEC-80d. Thus, it remains possible that genome-wide DNA methylation patterns may be an important step in directing the cell toward a more cancerous state. It is important to note that we also observed the occurrence of hypo- and hyper-methylation changes outside CpG islands. Although these methylation changes were less frequent, this suggests that significant DNA methylation aberrations may also be prevalent in regions outside of traditionally defined CpG islands, including CpG island shores. Thus, epigenetic regulation during breast oncogenesis is not restricted to traditionally defined CpG islands, but instead may occur in regions distant from the transcriptional start site; recent studies suggest this is also true of both colon and esophageal cancer [10,30].

To further understand the timing of DNA methylation changes, we examined CpG loci exhibiting significant differential methylation from SHMEC versus THMEC-40d for change in methylation trends following the 40d time point. Hundreds of CpG loci (805) were hypomethylated from SHMEC to THMEC-40d and exhibited no methylation change between THMEC-40d and THMEC-80d. Methylation trends prior to the 40d time point were also assessed for CpG loci identified as differentially methylated between THMEC-40d and THMEC-80d. We found that 196 loci exhibited no methylation change from SHMEC to 40d and were hypermethylated from 40d to 80d. We used IPA to assess the biological enrichment and significance of genes exhibiting these differentially methylated loci. Because the Illumina Human Methylation27 BeadChip platform contains well-defined genes, it is important to note the possibility for potential skew in the pathway analysis. In addition, not every gene displaying differential methylation will impact transcription levels.

Genes with early differential methylation from SHMEC to THMEC- 40d were most closely associated with the processes of cellular movement and growth, and proliferation. This is consistent with our prior proteomics analysis completed on the same neoplastic reprogramming model, which identified nine differentially expressed proteins over the course of oncogenesis. All were associated with maintaining an early metastatic state through motility and migration [6].

Biological functions of genes associated with CpG loci differentially methylated later in carcinogenesis from THMEC-40d to THMEC-80d involved several types of development, including both organismal and embryonic. It has previously been observed that a large scope of human cancers silence developmental genes that play decisive roles in differentiation of adult cells. In particular, groups of hypermethylated homeobox genes have been detected in both lung and breast cancer [31,32]. Helman *et al.* recently determined that DNA hypermethylation of lung cancer primarily functions as a repressive mechanism of genes required to maintain a differentiated state [33]. Because the top five biological functions associated with this gene list involve various developmental processes, it is plausible that the DNA hypermethylation functions to silence vital developmental genes during this stage of breast oncogenesis, leading to cellular dedifferentiation. In addition, several observations suggest a functional link exists between developmental genes, Polycomb group (PcG) protein binding and CpG island hypermethylation. The reported percentage of PcG-binding sites within CpG islands falls between 50 and 88% [34]. Interestingly, developmental regulator genes that are PcG-marked in adult stem cells have the highest probability to be permanently methylated and consequently silenced in cancers, permitting tumorigenesis [35]. Also, direct interactions between PcG proteins and DNMT enzymes have been described, highlighting a possible mechanism for DNMT recruitment [36,37]. Approximately 30% of genes associated with hypermethylated sites from THMEC-40d to THMEC-80d are potential targets of a component of the Polycomb repressive complex 2, which consists of three essential core subunits: EZH2, EED and SUZ12, and functions to mainly trimethylate H3K27 (data not shown) [38,39]. Additional studies will be necessary to fully explore the timing and significance of Polycomb repressive complex 2 binding and developmental gene repression through CpG island DNA hypermethylation in breast oncogenesis.

Conclusion

In summary, this study illustrates several interesting methylomic trends during the initiation and beginning progressive stages of breast oncogenesis, prompting numerous further questions about the disease. As noted, others have reported similar DNA methylation trend findings in primary breast cell lines, but have not yet indicated the timing of these epigenetic modifications. Here, we have identified potential lists of CpG loci for further analysis based

on the timing of differential methylation. Future studies on this subset of genes will be aimed toward assessing the importance of the timing of these multiple DNA methylation aberrations on gene transcription.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors appreciate the help of T Hardy in proofread-ing the manuscript. The authors thank JT DeAngelis for providing the cell lines.

This work was funded by grants to TO Tollefsbol from the American Institute for Cancer Research and the National Institutes of Health (CA 129415). The authors acknowledge the support of the UAB Comprehensive Cancer Center (P30-CA013148) Heflin Center for Genomic Sciences Genomics Core and its personnel for assisting with all portions of the Illumina Infinium HumanMethylation27 BeadChip experiments.

References

Papers of special note have been highlighted as:

■of interest

■●of considerable interest

- DeSantis C, Siegel R, Bandi P, Jemal A. Breast cancer statistics, 2011. *CA Cancer J. Clin.* 2011; 61(6):409–418. [PubMed: 21969133]
- Hahn WC, Counter CM, Lundberg AS, Beijersbergen RL, Brooks MW, Weinberg RA. Creation of human tumour cells with defined genetic elements. *Nature.* 1999; 400(6743):464–468. [PubMed: 10440377] ■Original neoplastic reprogramming paper.
- Casillas MA, Brotherton SL, Andrews LG, Ruppert JM, Tollefsbol TO. Induction of endogenous telomerase (hTERT) by c-Myc in WI-38 fibroblasts transformed with specific genetic elements. *Gene.* 2003; 316:57–65. [PubMed: 14563552]
- Liu J, Yang G, Thompson-Lanza JA, et al. A genetically defined model for human ovarian cancer. *Cancer Res.* 2004; 64(5):1655–1663. [PubMed: 14996724]
- Elenbaas B, Spirio L, Koerner F, et al. Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. *Genes Dev.* 2001; 15(1):50–65. [PubMed: 11156605] ■Neoplastic transformation applied to human mammary epithelial cells.
- DeAngelis JT, Li Y, Mitchell N, Wilson L, Kim H, Tollefsbol TO. 2D difference gel electrophoresis analysis of different time points during the course of neoplastic transformation of human mammary epithelial cells. *J. Proteome Res.* 2011; 10(2):447–458. [PubMed: 21105747] ■Proteomics analysis of a real-time human mammary epithelial cell neoplastic reprogramming model.
- Katto J, Mahlknecht U. Epigenetic regulation of cellular adhesion in cancer. *Carcinogenesis.* 2011; 32(10):1414–1418. [PubMed: 21705481]
- Baylin SB, Jones PA. A decade of exploring the cancer epigenome – biological and translational implications. *Nat. Rev. Cancer.* 2011; 11(10):726–734. [PubMed: 21941284] ■Excellent review of recent cancer epigenomics efforts.
- Doi A, Park IH, Wen B, et al. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat. Genet.* 2009; 41(12):1350–1353. [PubMed: 19881528]
- Irizarry RA, Ladd-Acosta C, Wen B, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* 2009; 41(2):178–186. [PubMed: 19151715] ■Provides evidence for important DNA methylation changes in regions outside of CpG islands.

11. Anderson RM, Bosch JA, Goll MG, et al. Loss of Dnmt1 catalytic activity reveals multiple roles for DNA methylation during pancreas development and regeneration. *Dev. Biol.* 2009; 334(1): 213–223. [PubMed: 19631206]
12. Herman JG, Baylin SB. Gene silencing in cancer in association with promoter hypermethylation. *N. Engl. J. Med.* 2003; 349(21):2042–2054. [PubMed: 14627790]
13. Bibikova M, Le J, Barnes B, et al. Genome-wide DNA methylation profiling using Infinium® assay. *Epigenomics.* 2009; 1(1):177–200. [PubMed: 22122642]
14. Sun Z, Chai HS, Wu Y, et al. Batch effect correction for genome-wide methylation data with Illumina Infinium platform. *BMC Med. Genomics.* 2011; 4:84. [PubMed: 22171553]
15. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA.* 2003; 100(16):9440–9445. [PubMed: 12883005]
16. Rajendram R, Ferreira JC, Grafodatskaya D, et al. Assessment of methylation level prediction accuracy in methyl-DNA immunoprecipitation and sodium bisulfite based microarray platforms. *Epigenetics.* 2011; 6(4):410–415. [PubMed: 21343703]
17. Hanna CW, Bloom MS, Robinson WP, et al. DNA methylation changes in whole blood is associated with exposure to the environmental contaminants, mercury, lead, cadmium and bisphenol A, in women undergoing ovarian stimulation for IVF. *Hum. Reprod.* 2012; 27(5):1401–1410. [PubMed: 22381621]
18. Shen J, Wang S, Zhang YJ, et al. Genome-wide DNA methylation profiles in hepatocellular carcinoma. *Hepatology.* 2012; 55(6):1799–1808. [PubMed: 22234943]
19. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat. Rev. Genet.* 2007; 8(4):286–298. [PubMed: 17339880]
20. Holliday R. Epigenetics: a historical overview. *Epigenetics.* 2006; 1(2):76–80. [PubMed: 17998809]
21. Herman JG, Merlo A, Mao L, et al. Inactivation of the *CDKN2/p16/MTS1* gene is frequently associated with aberrant DNA methylation in all common human cancers. *Cancer Res.* 1995; 55(20):4525–4530. [PubMed: 7553621]
22. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.* 2002; 3(6):415–428. [PubMed: 12042769]
23. Bernardino J, Roux C, Almeida A, et al. DNA hypomethylation in breast cancer: an independent parameter of tumor progression? *Cancer Genet. Cytogenet.* 1997; 97(2):83–89. [PubMed: 9283586]
24. Soares J, Pinto AE, Cunha CV, et al. Global DNA hypomethylation in breast carcinoma: correlation with prognostic factors and tumor progression. *Cancer.* 1999; 85(1):112–118. [PubMed: 9921982]
25. Veeck J, Esteller M. Breast cancer epigenetics: from DNA methylation to microRNAs. *J. Mammary Gland Biol. Neoplasia.* 2010; 15(1):5–17. [PubMed: 20101446]
26. Ruike Y, Imanaka Y, Sato F, et al. Genome-wide analysis of aberrant methylation in human breast cancer cells using methyl-DNA immunoprecipitation combined with high-throughput sequencing. *BMC Genomics.* 2010; 11:137. [PubMed: 20181289]
27. Fackler MJ, Umbricht CB, Williams D, et al. Genome-wide methylation analysis identifies genes specific to breast cancer hormone receptor status and risk of recurrence. *Cancer Res.* 2011; 71(19): 6195–6207. [PubMed: 21825015] ■■Assesses DNA methylation aberration based on estrogen receptor status of breast cancer subtype.
28. Tan J, Gu Y, Zhang X, et al. Hypermethylation of CpG islands is more prevalent than hypomethylation across the entire genome in breast carcinogenesis. *Clin. Exp. Med.* 2012; 13(1): 1–9. [PubMed: 22212995]
29. Park SY, Kwon HJ, Lee HE, et al. Promoter CpG island hypermethylation during breast cancer progression. *Virchows Arch.* 2011; 458(1):73–84. [PubMed: 21120523]
30. Alvarez H, Opalinska J, Zhou L, et al. Widespread hypomethylation occurs early and synergizes with gene amplification during esophageal carcinogenesis. *PLoS Genet.* 2011; 7(3):e1001356. [PubMed: 21483804]

31. Rauch T, Wang Z, Zhang X, et al. Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay. *Proc. Natl Acad. Sci. USA*. 2007; 104(13):5527–5532. [PubMed: 17369352]
32. Novak P, Jensen T, Oshiro MM, et al. Epigenetic inactivation of the *HOXA* gene cluster in breast cancer. *Cancer Res*. 2006; 66(22):10664–10670. [PubMed: 17090521]
33. Helman E, Naxerova K, Kohane IS. DNA hypermethylation in lung cancer is targeted at differentiation-associated genes. *Oncogene*. 2012; 31(9):1181–1188. [PubMed: 21804601]
34. van Vlodrop IJ, Niessen HE, Derks S, et al. Analysis of promoter CpG island hypermethylation in cancer: location, location, location! *Clin. Cancer Res*. 2011; 17(13):4225–4231. [PubMed: 21558408]
35. Easwaran H, Johnstone S, Vanneste L, et al. A DNA hypermethylation module for the stem/progenitor cell signature of cancer. *Genome Res*. 2012; 22(5):837–849. [PubMed: 22391556]
36. Mohammad HP, Cai Y, McGarvey KM, et al. Polycomb CBX7 promotes initiation of heritable repression of genes frequently silenced with cancer-specific DNA hypermethylation. *Cancer Res*. 2009; 69(15):6322–6330. [PubMed: 19602592]
37. Viré E, Brenner C, Deplus R, et al. The Polycomb group protein EZH2 directly controls DNA methylation. *Nature*. 2006; 439(7078):871–874. [PubMed: 16357870]
38. Ben-Porath I, Thomson MW, Carey VJ, et al. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat. Genet*. 2008; 40(5):499–507. [PubMed: 18443585]
39. Lee TI, Jenner RG, Boyer LA, et al. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell*. 2006; 125(2):301–313. [PubMed: 16630818]

Websites

101. SEER Cancer Statistics Review 1975–2009. http://seer.cancer.gov/csr/1975_2009_pops09/
45. Pathway analysis, mRNA, RNA-seq, NGS and microarray. www.ingenuity.com.

Executive summary

- The neoplastic transformation of human mammary epithelial cells (HMECs) is a well-established model that can be used to analyze real-time DNA methylation changes over the course of breast carcinogenesis.
- A large loss of DNA methylation was observed during initial transition to an oncogenic state, from semitransformed HMECs to transformed HMEC (THMEC)-40 days (d).
- DNA hypermethylation during the THMEC-40d to THMEC-80d transition occurred mainly in CpG islands.
- Genes displaying early differential methylation from semitransformed HMEC to THMEC-40d were enriched for the processes of cellular movement and cellular growth and proliferation, which is consistent with observations from a prior proteomics analysis completed on the same neoplastic reprogramming model.
- CpG loci differentially methylated later in breast oncogenesis from THMEC-40d to THMEC-80d were associated with several types of development, including both organismal and embryonic. This potentially provides a mechanism for cellular dedifferentiation.
- Our results suggest that the timing of differential DNA methylation may play a role in driving the cell toward an increasingly greater cancerous state.

Figure 1. Experimental design filtering methods

The statistical and biological filters applied to data appear on the left side of the diagram, whereas the types of analyses completed are listed on the right. Boxes contain the number of CpG loci represented in each data set. All cell lines are represented by orange boxes, SHMEC versus THMEC-40d comparisons are outlined in purple, SHMEC versus THMEC-80d comparisons are in blue and THMEC-40d versus THMEC-80d comparisons are indicated in green.

$|\Delta\beta|$: Absolute methylation level difference; ANOVA: Analysis of variance; d: Day; DMA: Differential methylation analysis; HSD: Honestly significant difference; PCA: Principal component analysis; SHMEC: Semitransformed human mammary epithelial cells; THMEC: Transformed human mammary epithelial cells.

Figure 2. Principal component analysis of DNA methylation data

The methylation data for all 27,578 CpG loci from each replicate ($n = 4$ for each sample) were subjected to principal component analysis. SHMEC, THMEC-40d and THMEC-80d appear to have distinct epigenetic profiles as demonstrated by the separation of samples based on cell type.

d: Day; SHMEC: Semitransformed human mammary epithelial cells; THMEC: Transformed human mammary epithelial cells.

Figure 3. Mean β values

Distribution of average β values for each cell type for all 1996 CpG loci identified as differentially methylated through statistical filtering. A loss of DNA methylation occurs as an early event in oncogenesis during the transition from SHMEC to THMEC-40d. Error bars represent standard error of the mean.

** $p < 0.01$ vs THMEC-40d; **** $p < 0.0001$ vs SHMEC

d: Day; SHMEC: Semitransformed human mammary epithelial cells; THMEC: Transformed human mammary epithelial cells.

Figure 4. Venn diagram analysis for absolute methylation level difference >0.20

Venn diagram analysis was completed for CpG loci passing the biological filter of an absolute methylation level difference ($|\Delta\beta|$) >0.20, corresponding to a 20% or greater difference in methylation change. **(A)** SHMEC versus THMEC-40d data set (n = 1010 CpG sites) overlapped with SHMEC versus THMEC-80d data set (n = 1033 CpG sites). **(B)** SHMEC versus THMEC-80d data set overlapped with THMEC-40d versus THMEC-80d data set (n = 358 CpG sites). **(C)** All three data sets overlapped. Values indicate number of CpG loci.

d: Day; SHMEC: Semitransformed human mammary epithelial cells; THMEC: Transformed human mammary epithelial cells.

Figure 5. Methylation trends

(A) SHMEC versus THMEC-40d (total $n = 1010$ CpG loci). (B) THMEC-40d versus THMEC-80d (total $n = 358$ CpG loci). Time is represented on the x-axis while β value is on the y-axis. The number of CpG loci exhibiting a particular methylation trend is indicated next to each diagram. We were interested in isolating genes with DNA methylation changes that occurred during only one time-point transition over the course of oncogenesis. These are boxed for each data set comparison.

d: Day; SHMEC: Semitransformed human mammary epithelial cells; THMEC: Transformed human mammary epithelial cells.

Figure 6. Top Ingenuity Pathway Analysis biological functions

(A) The top biological functions for genes included in the SHMEC versus THMEC-40d data sets. During the onset of oncogenesis, differential methylation occurs in genes associated with cellular movement, cellular growth and proliferation and cell death. (B) The top biological functions for genes included in the THMEC-40d versus THMEC-80d data set. Differential methylation occurs in genes associated with tissue, cellular and embryonic development during late oncogenesis. The x-axis is represented as the $-\log(p\text{-value})$. The threshold is set at $p < 0.05$.

d: Day; SHMEC: Semitransformed human mammary epithelial cells; THMEC: Transformed human mammary epithelial cells.

Figure 7. Correlation between DNA methylation measures by Illumina average β and Pyrosequencing®
(A) *HMG3* ($r = 0.952$) and **(B) *ZNF577*** ($r = 0.825$). The x-axis is represented by the Illumina average β value and the y-axis corresponds to the methylation percentage from Pyrosequencing.

Table 1

Occurrence of hypo- or hyper-methylation location.

	Within CpG island; % (n)	Outside CpG island; % (n)
<i>Hypomethylated site location</i>		
SHMEC vs THMEC-40d	65.0 (549/844)	35.0 (295/844)
THMEC-40d vs THMEC-80d	53.1 (85/160)	46.9 (75/160)
<i>Hypermethylated site location</i>		
SHMEC vs THMEC-40d	53.0 (88/166)	47.0 (78/166)
THMEC-40d vs THMEC-80d	82.3 (163/198)	17.7 (35/198)

Data show number and percentage of CpG loci from indicated group comparisons exhibiting hypo- or hypermethylation and their location within or outside CpG islands. CpG islands are defined according to Illumina array criteria.

d: Day; SHMEC: Semitransformed human mammary epithelial cells; THMEC: Transformed human mammary epithelial cells.