
[All ETDs from UAB](#)

[UAB Theses & Dissertations](#)

2014

Compression

Bethany Anne Mitchell
University of Alabama at Birmingham

Follow this and additional works at: <https://digitalcommons.library.uab.edu/etd-collection>

Recommended Citation

Mitchell, Bethany Anne, "Compression" (2014). *All ETDs from UAB*. 2483.
<https://digitalcommons.library.uab.edu/etd-collection/2483>

This content has been accepted for inclusion by an authorized administrator of the UAB Digital Commons, and is provided as a free open access item. All inquiries regarding this item or the UAB Digital Commons should be directed to the [UAB Libraries Office of Scholarly Communication](#).

GENOME-WIDE DNA METHYLATION CHANGES DURING BREAST
TUMORIGENESIS

by

NATALIE E. MITCHELL

TRYGVE O. TOLLEFSBOL, COMMITTEE CHAIR
MOLLY S. BRAY
JAMES A. COKER

A THESIS

Submitted to the graduate faculty of the University of Alabama at Birmingham,
in partial fulfillment of the requirements for the degree of
Master of Science

BIRMINGHAM, ALABAMA

2012

Copyright by
Natalie E. Mitchell
2012

GENOME-WIDE DNA METHYLATION CHANGES DURING BREAST TUMORIGENESIS

NATALIE E. MITCHELL

BIOLOGY

ABSTRACT

Neoplastic reprogramming generates cancerous cells from normal derivatives, enabling researchers to create a model of the early stages of oncogenesis and analyze molecular changes in real-time. Aberrations in DNA methylation are considered a major mechanism contributing to the acquisition and maintenance of a cancerous phenotype; however, little research exists regarding gene-specific changes in DNA methylation during the transition from a normal to neoplastic cell. Human mammary epithelial cells (HMECs) underwent neoplastic transformation through the addition of three defined genetic elements to generate one line of semi-transformed, pre-malignant cells (SHMECs) and two separate lines of fully transformed, malignant cells (THMECs). We used an Illumina Infinium HumanMethylation27 BeadChip to interrogate 27,578 CpG loci and identify genes that exhibited differential methylation over three consecutive time points during breast tumorigenesis. We found 805 CpG loci exhibiting a single DNA methylation change from SHMEC to THMECs at 40 days (THMEC-40d). These genes were enriched for biological functions associated with cellular growth and proliferation. In contrast, 196 CpG sites were differentially methylated between THMEC-40d and THMEC-80d and were associated with major developmental processes. The data

generated provide novel insight into the both the timing and possible role of specific DNA methylation changes throughout the initiation and progression of breast oncogenesis. However, additional studies are necessary to fully explore how these specific DNA methylation alterations, in addition to other epigenetic aberrations, mechanistically contribute to differential gene and protein expression, enabling the development of breast cancer.

Key Words: breast cancer, neoplastic reprogramming, epigenetics, DNA methylation, Illumina Infinium HumanMethylation27 BeadChip, MIRA

ACKNOWLEDGEMENTS

I express my sincere gratitude and thanks to my mentor, Dr. Trygve Tollefsbol, for enabling me to freely conduct research as both an undergraduate and graduate student and instructing me to think independently. In addition, I appreciate the guidance and support of the Tollefsbol lab members during the past four years. I would also like to acknowledge my committee members, Dr. Molly Bray and Dr. James Coker, for their assistance in obtaining my degree. Finally, thank you to my family and friends for your constant encouragement.

TABLE OF CONTENTS

	<i>Page</i>
ABSTRACT	iii
ACKNOWLEDGMENTS	v
LIST OF TABLES	vii
LIST OF FIGURES	viii
LIST OF ABBREVIATIONS	ix
GENERAL INTRODUCTION	1
GENOME-WIDE DNA METHYLATION ANALYSIS OF HUMAN MAMMARY EPITHELIAL CELL NEOPLASTIC TRANSFORMATION	7
METHYLATED-CpG ISLAND RECOVERY ASSAY	38
GENERAL LIST OF REFERENCES	55
APPENDICES:	
A PYROSEQUENCING PRIMERS FOR CpG LOCI OF INTEREST	59
B CpG LOCI EXHIBITING DIFFERENTIAL METHYLATION ONLY BETWEEN SHMEC AND THMEC-40D.....	60
C CpG LOCI EXHIBITING DIFFERENTIAL METHYLATION ONLY BETWEEN THMEC-40D AND THMEC-80D	80

LIST OF TABLES

<i>Table</i>	<i>Page</i>
GENOME-WIDE DNA METHYLATION ANALYSIS OF HUMAN MAMMARY EPITHELIAL CELL NEOPLASTIC TRANSFORMATION	
1 Occurrence of hypo- or hypermethylation location	30

LIST OF FIGURES

Figure *Page*

INTRODUCTION

1 Neoplastic reprogramming of Human Mammary Epithelial Cells (HMECs)3

GENOME-WIDE DNA METHYLATION ANALYSIS OF HUMAN MAMMARY EPITHELIAL CELL NEOPLASTIC TRANSFORMATION

1 Schematic of experimental design filtering methods..... 31

2 Principal Component Analysis (PCA) of DNA methylation data 32

3 Mean beta (β) value 33

4 Venn diagram analysis for $|\Delta\beta| \geq 0.20$ 34

5 Methylation trends 35

6 Top IPA biological functions 36

7 Correlation between DNA methylation measure by Illumina average beta and
pyrosequencing 37

METHYLATED-CpG ISLAND RECOVERY ASSAY

1 Schematic of the steps necessary to complete MIRA43

ABBREVIATIONS

5mC	5-methylcytosine
ANOVA	analysis of variance
β value	beta value
ChIP	chromatin immunoprecipitation
CpG	cytosine phosphate guanosine
DNMT	DNA methyltransferase
ER	estrogen receptor
FDR	false discovery rate
GST	glutathione-S-transferase
HMEC	human mammary epithelial cell
hTERT	exogenous human telomerase reverse transcriptase
HUGO	Human Gene Organization
IPA	Ingenuity Pathway Analysis
Kb	Kilobase
LT	large T antigen
MBD	methyl-CpG binding domain
MeDIP	methylated DNA immunoprecipitation
MIRA	Methylated-CpG Island Recovery Assay
NGS	next-generation sequencing

PCA	Principal Component Analysis
PCR	polymerase chain reaction
PD	population doubling
pRB	retinoblastoma protein
RT	room temperature
SAM	S-adenylmethionine
SHMEC	semi-transformed HMEC (<i>SV40 ER, hTERT</i>)
st	small t antigen
SV40 ER	Simian Virus 40 early region
THMEC	transformed HMEC (<i>SV40 ER, hTERT, hRAS-V12</i>)
TSG	tumor suppressor gene
Tukey's HSD	Tukey's honestly significant difference
WGA	whole-genome tiling array
$ \Delta\beta $	absolute delta beta

INTRODUCTION

Oncogenesis

Oncogenesis, or the process by which a normal somatic cell transitions to a cancerous state, is one of the least understood aspects of cancer. Hanahan and Weinberg provided the first holistic definition of the disease by highlighting phenotypic characteristics acquired during the development of human tumors. Termed the hallmarks of cancer, these traits include: 1) limitless replicative potential, 2) evasion of apoptosis, 3) self-sufficiency in growth signals, 4) insensitivity to anti-growth signals, 5) sustained angiogenesis, and 6) invasion and metastasis [1]. Since the initial publication in 2000, our understanding of the disease has evolved, necessitating the addition of four new hallmarks in 2011: 7) abnormal metabolic pathways, 8) evasion of the immune system, 9) genome instabilities and 10) inflammation [2]. Despite recent technological advances in the area of cancer research, limited resources currently exist to investigate the timing of these hallmark molecular changes during the progression of the disease.

Neoplastic Reprogramming

The process of oncogenesis cannot be studied *in vivo*; however, the method of neoplastic reprogramming can be used to analyze oncogenesis in real-time *in vitro*. Weinberg and colleagues demonstrated that normal human epithelial cells and fibroblast cells could undergo the process of neoplastic reprogramming through retroviral-mediated

addition of three defined genetic elements- Simian Virus 40 Early Region (*SV40 ER*), exogenous human telomerase reverse transcriptase (*hTERT*) and an oncogenic allele of hRAS (*hRAS-V12*). The addition of these elements mimics the activation or inactivation of pathways that typically occur during oncogenesis *in vivo* [3,4]. *SV40 ER* sequesters p53 and retinoblastoma (pRB), thereby knocking out the defense mechanisms of the cell [5,6]. *hTERT* promotes an increase in telomerase activity, allowing the cell to replicate indefinitely [7]. Activated *hRAS-V12* mimics a sustained supply of growth signals, permitting cellular division and the acquisition of a malignant phenotype [8].

SV40 ER codes for both Large T (LT) and small t (st) antigens. LT is capable of transforming a variety of cell types due to its disruption of certain key tumor suppressors and cell cycle regulatory proteins, like pRB and p53 proteins [9,10]. Although it was originally assumed that st played only a minor role in this process, it has recently been elucidated that its presence is required for the complete transformation of normal cells into their cancerous derivatives. Additionally, expression of st stimulates cell proliferation and permits anchorage-independent growth [5]. The introduction of exogenous *hTERT*, which encodes the catalytic subunit of the telomerase holoenzyme, results in an up-regulation of telomerase activity, conferring the ability for uncontrolled cellular proliferation and immortalization [11]. Ectopic *hTERT* expression enables transformed cells to bypass crisis, a proliferative barrier after senescence characterized by widespread cell death [4]. The final step in the transformation process is the addition of an oncogenic form of RAS, known as *hRAS-V12*. The RAS family functions to transduce signals from tyrosine kinase receptors that promote cell survival and proliferation. Mutation of RAS causes insensitivity to GTPase-activating proteins and is one of the

most common genetic alterations found in human cancers. *hRAS-V12* prevents inhibitory signals from regulating RAS-mediated pathways [12].

Although neoplastic reprogramming was first accomplished in human epithelial and fibroblast cells, the procedure has also been completed to generate breast cancer cells from human mammary epithelial cells (HMECs) [13,14]. HMECs are normal, somatic cells with a finite lifespan that are unable to form tumors. Neoplastically transformed HMECs display most of the characteristics of breast cancer cells, including the ability to form tumors in nude mice. HMECs that have been transformed with *SV40ER* and *hTERT*, termed SHMECs, are pre-malignant breast cells. The addition of *hRAS-V12* to SHMECs generates fully malignant, transformed HMECs (THMECs) [14,15]. Figure 1 is a schematic of the neoplastic reprogramming model.

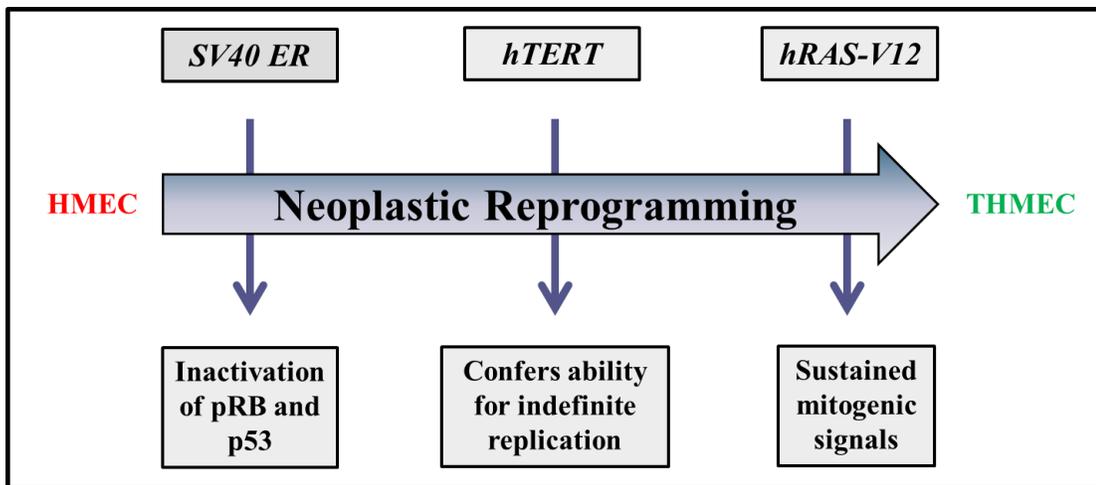


Figure 1. Neoplastic Reprogramming of Human Mammary Epithelial Cells (HMECs). The stable addition of three genetic elements (*SV40 ER*, *hTERT* and *hRAS-V12*) to HMECs generates fully transformed breast cancer cells (THMECs). THMECs are malignant breast cancer cells.

It is important to note that *SV40ER* only causes human breast cancer in approximately 22% of breast carcinoma cases; however, breast cancers frequently carry mutations that deregulate pRB and p53 pathways [16,17]. In addition, while mutated forms of RAS are activated in only 5% of breast cancers, mutations in the RAS signaling pathway are commonly seen in breast cancer. Although 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 [14]. The necessity for research regarding the mechanisms of breast cancer development is obvious- it is the second leading cause of cancer deaths in women; one in eight females will be diagnosed with the disease in their lifetime [18].

Epigenetic Modifications

Epigenetic modifications describe the heritable changes in gene expression that occur without a change in the primary DNA sequence, allowing modifications to pass from one generation of cells to the next in the same organism [19]. Changes in DNA methylation patterns and modifications of chromatin are two classical examples of epigenetic processes that contribute to aberrance, loss or gain of gene function [20]. DNA methylation is carried out by the DNA methyltransferase (DNMT) family of enzymes. DNMT1, DNMT3a and DNMT3b typically catalyze the transfer of a methyl group from S-adenylmethionine (SAM) to the 5-carbon of cytosine, producing a “fifth base,” 5-methylcytosine (5mC) [21,22]. DNMT1 is considered a maintenance enzyme that preserves DNA methylation after cellular replication. DNMT1 has a strong preference to

bind to hemi-methylated DNA [21]. DNMT3a and DNMT3b are *de novo* methyltransferases that enable DNA methylation patterns early in development; they show no binding preference to unmethylated versus methylated DNA [23]. DNMT1 is responsible for the majority of DNA methylation in cancer cells and maintains abnormal promoter methylation in neoplastic cells [24].

In humans, the 5mC modification generally occurs on cytosines that precede a guanosine (CpG dinucleotide) [25]. The distribution of CpG dinucleotides in the genome is asymmetric. Regions of DNA where a large density of CpG dinucleotides exist are known as CpG islands, which are found in the proximal promoter region in nearly half of the genes in the mammalian genome. These regions are typically unmethylated in normal cells. In contrast, roughly 80% of the CpG dinucleotides unassociated with CpG islands are heavily methylated. Compared to their normal cell counterparts, cancer cells display abnormal patterns of DNA methylation. Virtually all types of cancer have both gains in DNA methylation in CpG islands and losses of DNA methylation in regions outside of CpG islands [24]. During oncogenesis, genome-wide hypomethylation combined with site-specific hypermethylation leads to alterations in the expression of genes associated with tumor development [26,27]. CpG hypermethylation of promoter regions is integral for the transcriptional silencing of certain tumor suppressor genes (TSGs), leading to the acquisition of the typical hallmarks of a cancer cell, including limitless replicative potential and sustained angiogenesis [1,28].

The possible dynamic role that exists between epigenetic modifications and transcriptional regulation provides the basic foundation for explicating cancer mechanisms. Controlling access to DNA is a way to repress or activate transcription of

genes necessary for the progression of the disease. The means by which a normal cell transitions to a cancerous state may partially be understood by an in-depth analysis of changes in genome-wide DNA methylation patterns throughout the course of oncogenesis.

Microarray Technology

Recent advances in technology have changed the manner in which cancer can be assessed. Instead of analyzing alterations in DNA methylation on a gene-by-gene basis, cancer epigenetics can now be studied on a genome-wide basis. Although important aspects of cancer biology have been gained through smaller-scale studies, many questions still remain unanswered about epigenetics. For example, how many genes undergo epigenetic changes in breast cancer? What molecular mechanisms regulate the altered epigenetic profile [28]? In addition, predicting exactly when these epigenetic alterations occur has been difficult for researchers to pinpoint. Several existing technologies, including Illumina's Infinium HumanMethylation27 BeadChip and Methylated-CpG Island Recovery Assay (MIRA), are capable of generating the data necessary to explicate these complex questions. Chapter 1 will focus on the use of a methylomics array to assess changes in DNA methylation in a breast neoplastic transformation model, in order to elucidate the timing of epigenetic modifications during the progression of breast cancer. Chapter 2 provides a detailed explanation of the MIRA procedure.

GENOME-WIDE DNA METHYLATION ANALYSIS OF HUMAN MAMMARY
EPITHELIAL CELL NEOPLASTIC TRANSFORMATION

by

NATALIE E. MITCHELL, MACKENZIE L. WILSON, MOLLY S. BRAY, DAVID K.
CROSSMAN, TRYGVE O. TOLLEFSBOL

Submitted to *BMC Genomics*

Format adapted for thesis

ABSTRACT

Background: Neoplastic transformation can generate cancerous cells from normal cells through the addition of three defined genetic elements- Simian Virus 40 Early Region (*SV40 ER*), human telomerase reverse transcriptase (*hTERT*) and *hRAS-V12*, providing one of the few existing opportunities to analyze molecular changes in real-time during the initiation and progression of breast cancer. Human mammary epithelial cells (HMECs) were induced to undergo neoplastic reprogramming, generating one line of semi-transformed, pre-malignant cells (SHMECs) and two separate, temporal lines of fully transformed, malignant cells (THMECs). An Illumina Infinium HumanMethylation27 BeadChip was used to analyze alterations in DNA methylation in 27,578 CpG loci at three consecutive time points over the course of an eighty day (d) transformation period.

Results: A paired t-test ($p < 0.0001$) revealed the mean β value for SHMEC 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 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 distinctly different than in the later, more progressive stages of the transformation process.

Conclusions: These results suggest that the timing of major methylomic changes may be important in directing the cell toward a progressively more cancerous phenotype and that global hypomethylation may drive the initial transition to an oncogenic state.

Additionally, gene-specific hypermethylation appears to silence developmentally-related genes leading to breast oncogenesis de-differentiation. Future studies regarding the

implications of these DNA methylation alterations in conjunction with assessment of genome-wide chromatin modifications will provide unparalleled understanding of the epigenetic aberrations necessary for breast tumorigenesis.

Key words: breast cancer, neoplastic reprogramming, epigenetics, DNA methylation, Illumina Infinium HumanMethylation27 BeadChip

INTRODUCTION

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, 2]. 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 [3]. Although the procedure has been completed in epithelial and fibroblast cells [3, 4], it has also been extended to

generate several other cancer types, including breast cancer, through the use of human mammary epithelial cells (HMECs) [5, 6]. 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 up-regulation of Ras-associated pathways. In addition, this model 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 [6].

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. We have previously identified both genes and proteins associated with an increase in metastatic potential that displayed differential expression throughout the course of an eighty day transformation period [7]. Because changes to the epigenome contribute to the cancerous phenotype by altering the regulation of gene and protein expression, we elected to also investigate the epigenetic regulatory mechanisms occurring during the initiation and progression of human mammary cell neoplastic induction [8, 9].

DNA methylation has been shown to play a crucial role in defining the cell type and maintaining both the normal and diseased state [10-12]. However, compared to their normal cell counterparts, cancer cells display abnormal patterns of DNA methylation, generally characterized by widespread hypomethylation and site-specific hypermethylation [13]. 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 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 AND METHODS

Generation of the neoplastic reprogramming model and cell culture

SHMECs (HMECs transformed with *SV40 ER* and *hTERT*) and THMECs (SHMECs transformed with *hRAS-V12*) were generated and maintained as previously described [6, 7]. SHMECs are pre-malignant breast cells, while THMECs are fully transformed and malignant breast cancer cells, capable of producing tumors *in vivo* [7]. 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; Valencia, CA*). 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. (*San Diego, CA*). Briefly, following bisulfite treatment (*Zymo, Inc.; Seattle, WA*), 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; San Diego, CA*). Quality standards for bisulfite conversion, staining, single base extension, hybridization, stringency, and non-specific 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.; San Diego, CA*). The relative methylation value for each CpG locus was expressed as a beta (β) value. The β value was calculated by dividing the average signal for the methylated probe by the total signal for intensity [14]. 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 [15].

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 beta value means was

performed using an analysis of variance (ANOVA) with Benjamini-Hochberg multiple testing correction [false discovery rate (FDR) corrected $p < 0.01$], generating a list of CpG loci ($n= 1,996$) with statistically different β values [16]. Post hoc pairwise comparisons between cell lines were performed using a Tukey's HSD procedure. Three separate lists of differentially methylated CpG loci were of interest- (i) SHMEC versus THMEC-40d, (ii) SHMEC versus THMEC-80d and (iii) 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 [$\Delta\beta$] ($|\Delta\beta|$) of at least 0.20 ($|\Delta\beta| \geq 0.20$), 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. 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; Santa Clara, CA*).

Ingenuity Pathway Analysis (IPA)

Core bioinformatics analyses were performed using IPA (*Ingenuity Systems, www.ingenuity.com*). IPA contains a curated database of networks and biological relationships based on original peer-reviewed articles. Data sets containing Human Gene Organization (HUGO) 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 (Appendix A) that 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 [17, 18]. The Target ID for each CpG loci is indicated in parenthesis. The pyrosequencing reaction and quantitation of DNA methylation was performed using the PyroMarkHS96 sequencer and PyroMark MD software (*Qiagen; Valencia, CA*).

RESULTS

Principal Component Analysis (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 to 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.

Identification of statistically different β values

An ANOVA with Benjamini-Hochberg multiple testing correction (FDR corrected $p < 0.01$) was used to identify CpG loci that had statistically different β values over the course of neoplastic induction [16]. 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 around 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 to the difference between SHMEC and THMEC-40d, suggesting the loss of promoter region DNA methylation as an early event during oncogenesis. Although genome-wide hypomethylation is most commonly associated with repetitive elements in most cancers, the loss of DNA methylation may also occur at gene promoters, resulting in over-expression of proto-oncogenes [19, 20].

Venn diagram analysis of differentially methylated CpG loci

The list of 1,996 CpG loci was subjected to post hoc pairwise comparisons between cell lines using a Tukey's HSD 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 1,033 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- and hypermethylation and CpG site location

A total of 1,010 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 to a value of 47% for sites outside CpG islands. In 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=1,010), 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 6 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 hypermethylation 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, Appendix B) and THMEC-40d versus THMEC-80d (n=196 CpG loci, Appendix C) gene lists, containing Human Genome Organization (HUGO) 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 [18, 21]. *HGMN3* ($r=0.952$) and *ZNF577* ($r=0.825$) both displayed strong correlations between array data and percent methylation (Figure 7).

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 [13, 22]. 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 [13, 23-25]. 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 [6]. We have chosen three time points in the neoplastic progression to breast cancer to identify and analyze changes in DNA methylation patterns. Because both the sequence and timing of genome-wide epigenetic modifications are crucial in understanding the progression of breast cancer, the use of neoplastic transformation in a time-specific manner provides 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 to the SHMEC line. This additionally 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 1,996 CpG loci displaying statistically different β values over the course of neoplastic induction. SHMEC had the highest total average β value at around 0.245, with THMEC-40d and THMEC-80d having significantly lower mean β values, around 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. Approximately 50% of breast cancer cases show reduced 5-methylcytosine content compared with normal tissue counterparts. In addition, among varying solid tumor types, global DNA hypomethylation occurs most frequently in breast cancer [26-28]. Fackler et al. report a higher frequency of hypomethylation in ER-negative breast tumors as compared to ER-positive tumors, suggesting that the reduced methylation may account for the increased aggressiveness of this cancer subtype [29].

We then applied a stringent $|\Delta\beta|$ biological filter and determined that 1,010 CpG loci were differentially methylated from SHMEC to THMEC-40d compared to 358 differentially methylated sites during the transition from THMEC-40d to THMEC-80d. Categorizing whether the CpG sites were hypo- or hypermethylated 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 wide-spread occurrence during the initiation of breast cancer. Mechanistically, the observed hypomethylation may have a role in carcinogenesis through activation of oncogenes. 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 other systems [30, 31]. 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 hypermethylation 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 to be true of both colon and esophageal cancer as well [11, 32].

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 determine the biological significance of genes exhibiting these methylation changes. Genes with early differential methylation from SHMEC to THMEC-40d were most closely associated with the processes of cellular movement and cellular growth and proliferation. This is consistent with our prior proteomics analysis completed on the same neoplastic reprogramming model, which identified 9 differentially expressed proteins over the course of oncogenesis. All were associated with maintaining an early metastatic state through motility and migration [7].

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 [33, 34]. 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 [35]. 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 de-differentiation. 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% to 88% [36]. 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 [37]. Also, direct interactions between PcG proteins and DNA methyltransferase (DNMT) enzymes have been described, highlighting a possible mechanism for DNMT recruitment [38, 39].

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 (PRC2), which consists of three essential core subunits: EZH2, EED, and SUZ12 and functions to mainly trimethylate H3K27 (*data not shown*) [40, 41]. Additional studies will be necessary to fully explore the timing and significance of PRC2 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. We have identified potential lists of differentially methylated CpG loci for additional analysis. Future studies on this subset of genes will be aimed toward assessing other aspects of epigenetic control, including chromatin modifications, as well the effect of these multiple aberrations on gene transcription.

ACKNOWLEDGEMENTS

This work was funded by grants to TOT from the American Institute for Cancer Research and the National Institutes of Health (CA 129415). We thank Dr. J. Tyson DeAngelis for

providing the cell lines. We 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. We also appreciate the help of Dr. Tabitha Hardy in proofreading the manuscript.

REFERENCES

1. DeSantis C, Siegel R, Bandi P, Jemal A: **Breast cancer statistics, 2011.** *CA Cancer J Clin* 2011, **61**(6):409-418.
2. Howlader N, Noone AM, Krapcho M, Neyman N, Aminou R, et al. (1975-2008) **SEER Cancer Statistics Review.** National Cancer Institute. Bethesda, MD.
3. 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.
4. 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.
5. Liu J, Yang G, Thompson-Lanza JA, Glassman A, Hayes K, Patterson A, Marquez RT, Auersperg N, Yu Y, Hahn WC *et al*: **A genetically defined model for human ovarian cancer.** *Cancer Res* 2004, **64**(5):1655-1663.
6. Elenbaas B, Spirio L, Koerner F, Fleming MD, Zimonjic DB, Donaher JL, Popescu NC, Hahn WC, Weinberg RA: **Human breast cancer cells generated**

- by oncogenic transformation of primary mammary epithelial cells.** *Genes Dev* 2001, **15**(1):50-65.
7. 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.
 8. Katto J, Mahlknecht U: **Epigenetic regulation of cellular adhesion in cancer.** *Carcinogenesis* 2011, **32**(10):1414-1418.
 9. Baylin SB, Jones PA: **A decade of exploring the cancer epigenome - biological and translational implications.** *Nat Rev Cancer* 2011, **11**(10):726-734.
 10. Doi A, Park IH, Wen B, Murakami P, Aryee MJ, Irizarry R, Herb B, Ladd-Acosta C, Rho J, Loewer S *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.
 11. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M *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.
 12. Anderson RM, Bosch JA, Goll MG, Hesselson D, Dong PD, Shin D, Chi NC, Shin CH, Schlegel A, Halpern M *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.

13. Herman JG, Baylin SB: **Gene silencing in cancer in association with promoter hypermethylation.** *N Engl J Med* 2003, **349**(21):2042-2054.
14. Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, Gunderson KL: **Genome-wide DNA methylation profiling using Infinium® assay.** *Epigenomics* 2009, **1**(1):177-200.
15. Sun Z, Chai HS, Wu Y, White WM, Donkena KV, Klein CJ, Garovic VD, Therneau TM, Kocher JP: **Batch effect correction for genome-wide methylation data with Illumina Infinium platform.** *BMC Med Genomics* 2011, **4**:84.
16. Storey JD, Tibshirani R: **Statistical significance for genomewide studies.** *Proc Natl Acad Sci U S A* 2003, **100**(16):9440-9445.
17. Rajendram R, Ferreira JC, Grafodatskaya D, Choufani S, Chiang T, Pu S, Butcher DT, Wodak SJ, Weksberg R: **Assessment of methylation level prediction accuracy in methyl-DNA immunoprecipitation and sodium bisulfite based microarray platforms.** *Epigenetics* 2011, **6**(4):410-415.
18. Hanna CW, Bloom MS, Robinson WP, Kim D, Parsons PJ, vom Saal FS, Taylor JA, Steuerwald AJ, Fujimoto VY: **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.
19. Cho YH, Yazici H, Wu HC, Terry MB, Gonzalez K, Qu M, Dalay N, Santella RM: **Aberrant promoter hypermethylation and genomic hypomethylation in**

- tumor, adjacent normal tissues and blood from breast cancer patients.**
Anticancer Res 2010, **30**(7):2489-2496.
20. Dunn BK: **Hypomethylation: one side of a larger picture.** *Ann N Y Acad Sci* 2003, **983**:28-42.
21. Shen J, Wang S, Zhang YJ, Kappil M, Wu HC, Kibriya MG, Wang Q, Jasmine F, Ahsan H, Lee PH *et al*: **Genome-wide DNA methylation profiles in hepatocellular carcinoma.** *Hepatology* 2012.
22. Esteller M: **Cancer epigenomics: DNA methylomes and histone-modification maps.** *Nat Rev Genet* 2007, **8**(4):286-298.
23. Holliday R: **Epigenetics: a historical overview.** *Epigenetics* 2006, **1**(2):76-80.
24. Herman JG, Merlo A, Mao L, Lapidus RG, Issa JP, Davidson NE, Sidransky D, Baylin SB: **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.
25. Jones PA, Baylin SB: **The fundamental role of epigenetic events in cancer.** *Nat Rev Genet* 2002, **3**(6):415-428.
26. Bernardino J, Roux C, Almeida A, Vogt N, Gibaud A, Gerbault-Seureau M, Magdelenat H, Bourgeois CA, Malfoy B, Dutrillaux B: **DNA hypomethylation in breast cancer: an independent parameter of tumor progression?** *Cancer Genet Cytogenet* 1997, **97**(2):83-89.
27. Soares J, Pinto AE, Cunha CV, André S, Barão I, Sousa JM, Cravo M: **Global DNA hypomethylation in breast carcinoma: correlation with prognostic factors and tumor progression.** *Cancer* 1999, **85**(1):112-118.

28. Veeck J, Esteller M: **Breast cancer epigenetics: from DNA methylation to microRNAs.** *J Mammary Gland Biol Neoplasia* 2010, **15**(1):5-17.
29. Fackler MJ, Umbrecht CB, Williams D, Argani P, Cruz LA, Merino VF, Teo WW, Zhang Z, Huang P, Visvanathan K *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.
30. Tan J, Gu Y, Zhang X, You S, Lu X, Chen S, Han X, Sun Y: **Hypermethylation of CpG islands is more prevalent than hypomethylation across the entire genome in breast carcinogenesis.** *Clin Exp Med* 2012.
31. Park SY, Kwon HJ, Lee HE, Ryu HS, Kim SW, Kim JH, Kim IA, Jung N, Cho NY, Kang GH: **Promoter CpG island hypermethylation during breast cancer progression.** *Virchows Arch* 2011, **458**(1):73-84.
32. Alvarez H, Opalinska J, Zhou L, Sohal D, Fazzari MJ, Yu Y, Montagna C, Montgomery EA, Canto M, Dunbar KB *et al*: **Widespread hypomethylation occurs early and synergizes with gene amplification during esophageal carcinogenesis.** *PLoS Genet* 2011, **7**(3):e1001356.
33. Rauch T, Wang Z, Zhang X, Zhong X, Wu X, Lau SK, Kernstine KH, Riggs AD, Pfeifer GP: **Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay.** *Proc Natl Acad Sci U S A* 2007, **104**(13):5527-5532.
34. Novak P, Jensen T, Oshiro MM, Wozniak RJ, Nouzova M, Watts GS, Klimecki WT, Kim C, Futscher BW: **Epigenetic inactivation of the HOXA gene cluster in breast cancer.** *Cancer Res* 2006, **66**(22):10664-10670.

35. Helman E, Naxerova K, Kohane IS: **DNA hypermethylation in lung cancer is targeted at differentiation-associated genes.** *Oncogene* 2012, **31**(9):1181-1188.
36. van Vlodrop IJ, Niessen HE, Derks S, Baldewijns MM, van Criekinge W, Herman JG, van Engeland M: **Analysis of promoter CpG island hypermethylation in cancer: location, location, location!** *Clin Cancer Res* 2011, **17**(13):4225-4231.
37. Easwaran H, Johnstone S, Vanneste L, Ohm J, Mosbrugger T, Wang Q, Aryee MJ, Joyce P, Ahuja N, Weisenberger D *et al*: **A DNA hypermethylation module for the stem/progenitor cell signature of cancer.** *Genome Res* 2012.
38. Mohammad HP, Cai Y, McGarvey KM, Easwaran H, Van Neste L, Ohm JE, O'Hagan HM, Baylin SB: **Polycomb CBX7 promotes initiation of heritable repression of genes frequently silenced with cancer-specific DNA hypermethylation.** *Cancer Res* 2009, **69**(15):6322-6330.
39. Viré E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C, Morey L, Van Eynde A, Bernard D, Vanderwinden JM *et al*: **The Polycomb group protein EZH2 directly controls DNA methylation.** *Nature* 2006, **439**(7078):871-874.
40. Ben-Porath I, Thomson MW, Carey VJ, Ge R, Bell GW, Regev A, Weinberg RA: **An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors.** *Nat Genet* 2008, **40**(5):499-507.
41. Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K *et al*: **Control of developmental regulators by Polycomb in human embryonic stem cells.** *Cell* 2006, **125**(2):301-313.

Table 1. Occurrence of hypo- or hypermethylation location. Data show number and percentage of CpG loci from indicated group comparisons exhibiting hypomethylation or hypermethylation and their location within or outside CpG islands.

A. Hypomethylated Site Location

	Within CpG Island	Outside CpG Island
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)

B. Hypermethylated Site Location

	Within CpG Island	Outside CpG Island
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)

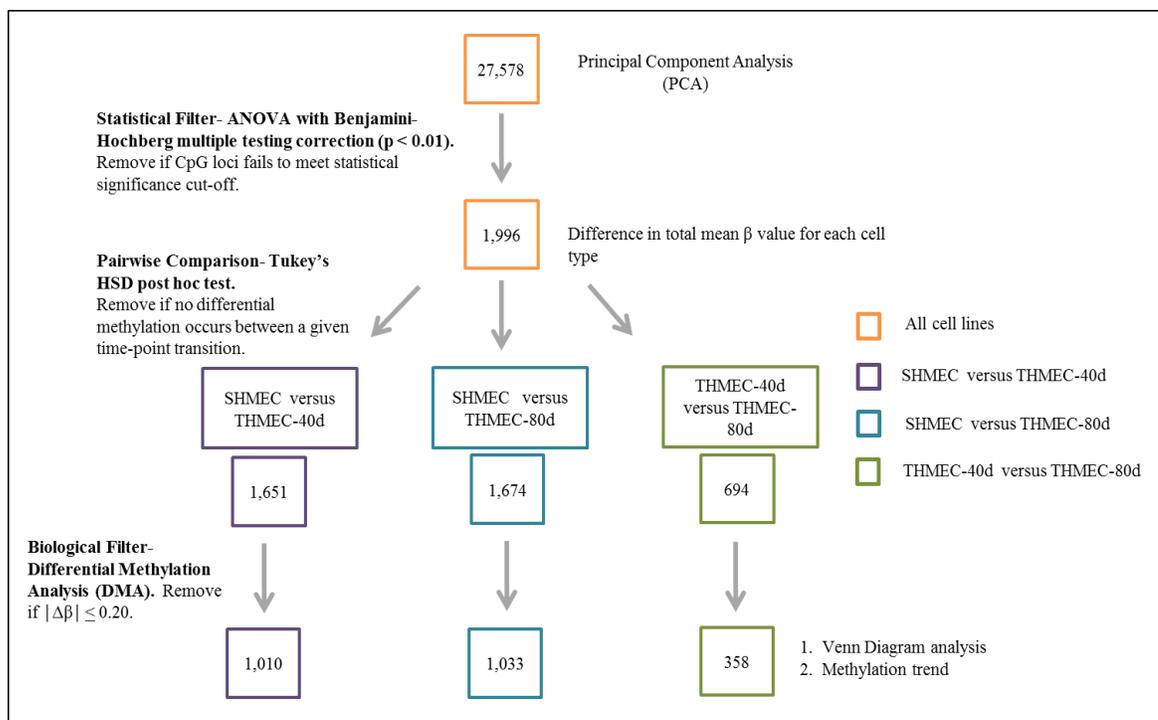


Figure 1. Schematic of experimental design filtering methods. The statistical and biological filters applied to data appear on the left side of the diagram, whereas the types of analyses completed are listed to the right. Boxes contain the number of CpG loci represented in each data set. All cell lines 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.

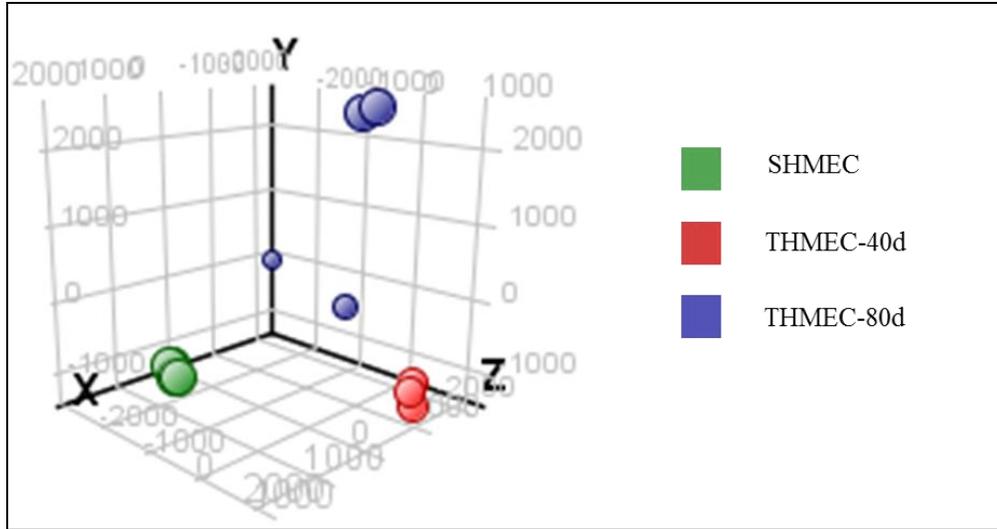


Figure 2. Principal Component Analysis (PCA) of DNA methylation data. The methylation data for all 27,578 CpG loci from each replicate (n=4 for each sample) were subjected to PCA. SHMEC, THMEC-40d and THMEC-80d appear to have distinct epigenetic profiles as demonstrated by the separation of samples based on cell type. SHMEC replicates are indicated by green circles, THMEC-40d in red and THMEC-80d in blue.

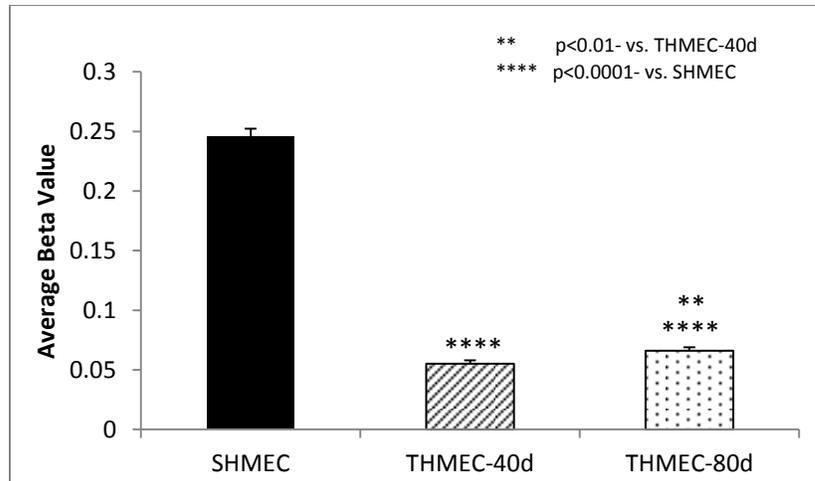


Figure 3. Mean beta (β) value. Distribution of average β value for each cell type for all 1,996 CpG loci identified as differentially methylated through statistical filtering. A loss of DNA methylation occurs as an early event in oncogenesis during the transition from SHMEC to THMEC-40d. Error bars represent standard error of the mean (SEM).

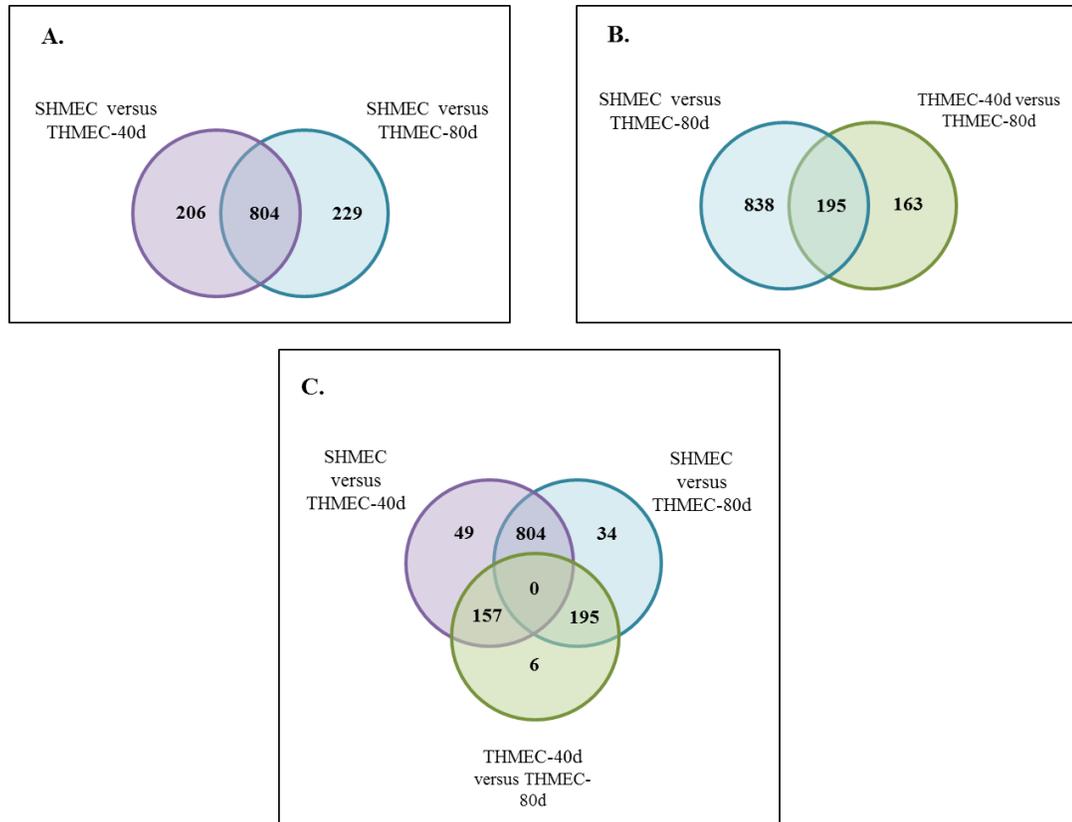


Figure 4. Venn diagram analysis for $|\Delta\beta| \geq 0.20$. Venn Diagram Analysis was completed for CpG loci passing the biological filter of a $|\Delta\beta| \geq 0.20$, corresponding to a 20% or greater difference in methylation change. A.) SHMEC versus THMEC-40d data set (n= 1,010 CpG sites) overlapped with SHMEC versus THMEC-80d data set (1,033 CpG sites). B.) SHMEC versus THMEC-80d data set overlapped with THMEC-40d versus THMEC-80d data set (358 CpG sites). C.) All three data sets overlapped. Values indicate number of CpG loci.

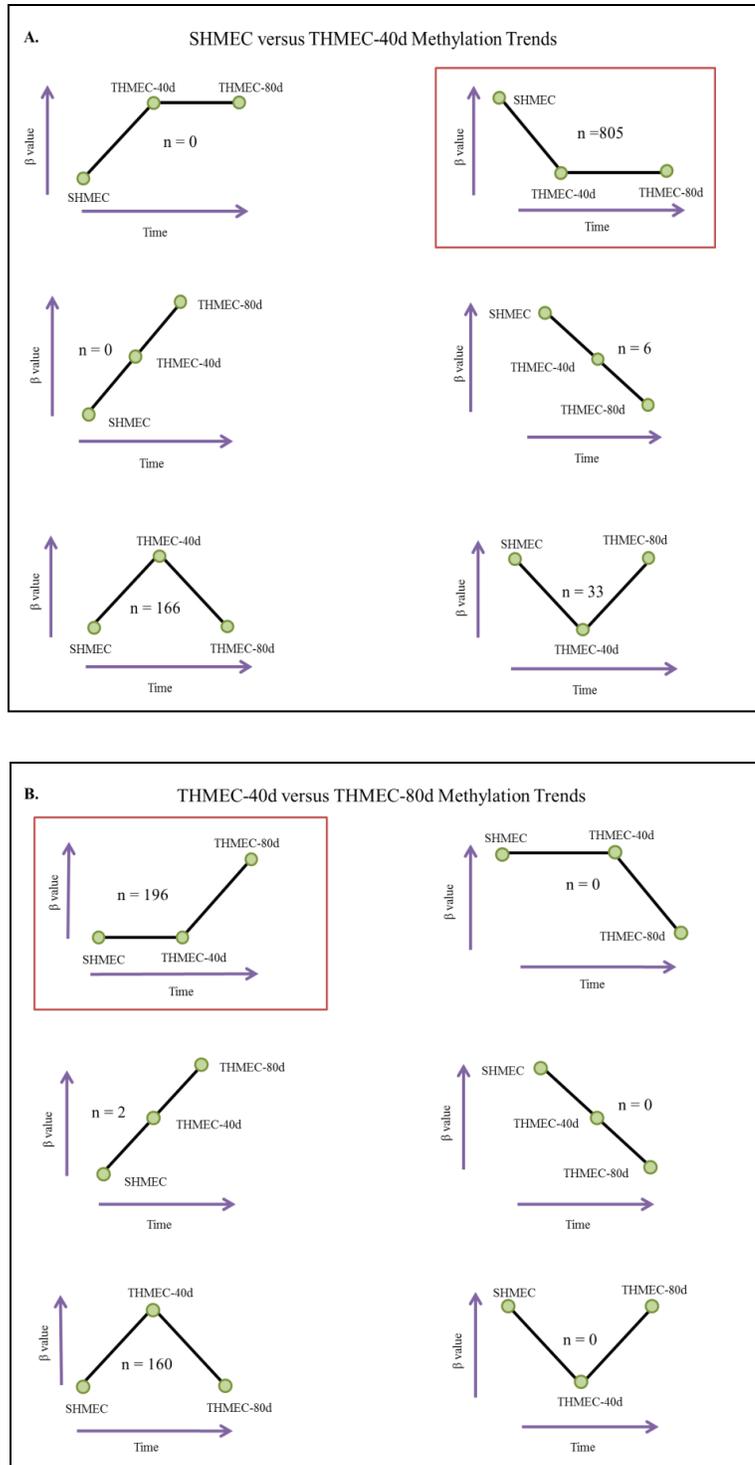


Figure 5. Methylation trends. A.) SHMEC versus THMEC-40d (total $n = 1,010$ CpG loci). B.) THMEC-40d versus THMEC-80d (total $n = 358$ CpG loci). Time is represented on the x-axis while β value is on the y-axis. The number of CpG loci exhibiting a particular methylation trend is indicated next to each 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.

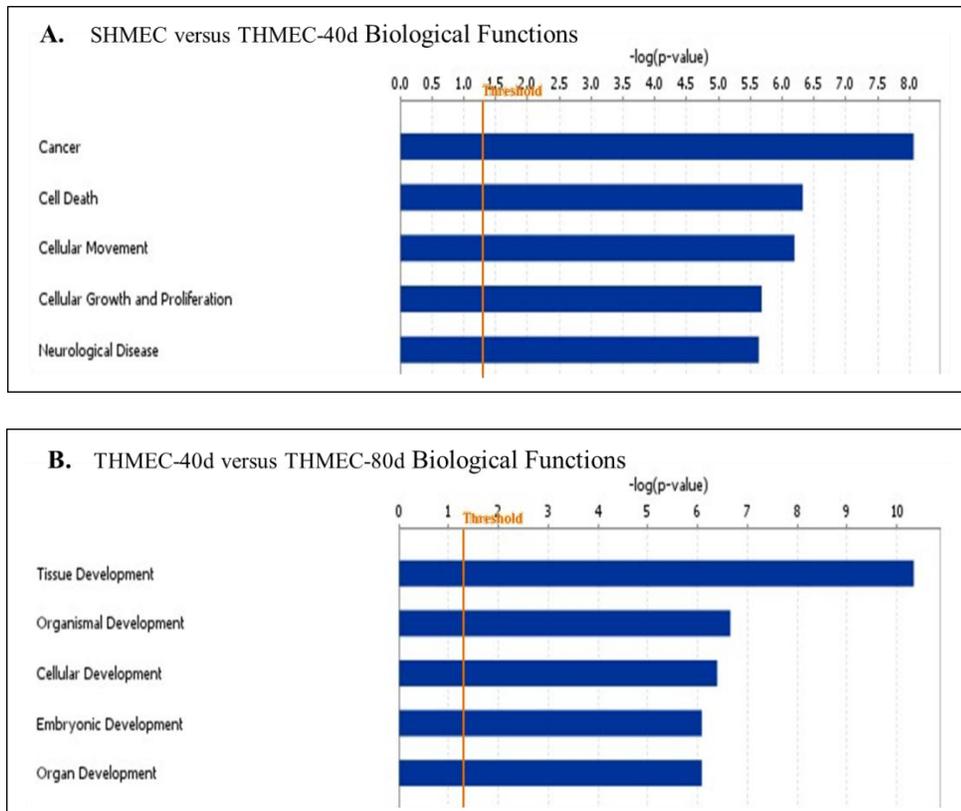


Figure 6. Top IPA biological functions. A.) The top biological functions for genes included in the SHMEC versus THMEC-40d data set. During the onset of oncogenesis, differential methylation occurs in genes associated with cellular movement and cellular growth and proliferation. B.) The top biological 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 $p < 0.05$.

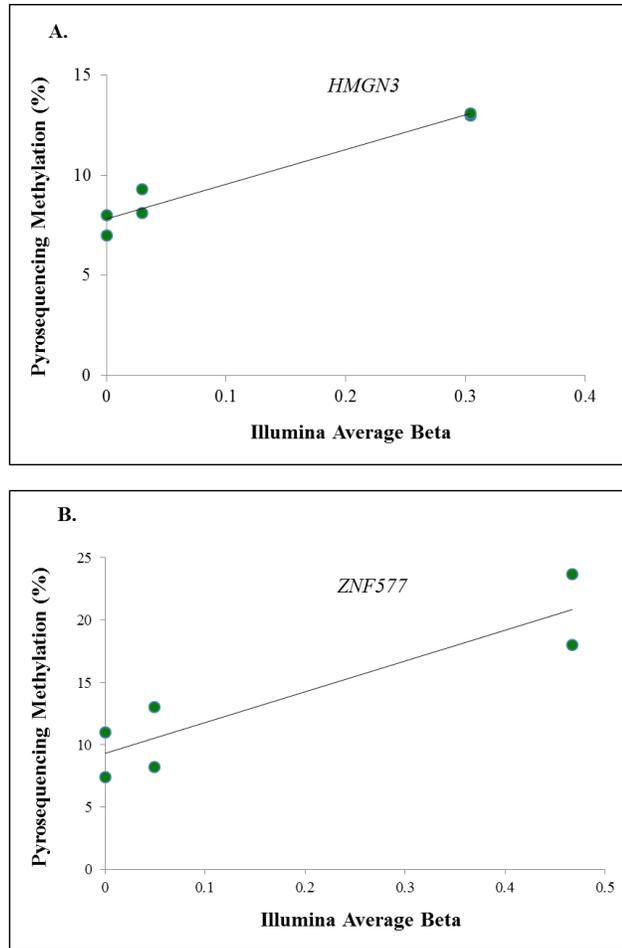


Figure 7. Correlation between DNA methylation measure by Illumina average beta and pyrosequencing. A.) *HMGN3* ($r=0.952$) and B.) *ZNF577* ($r=0.825$). The x-axis is represented by the Illumina Average Beta Value and the y-axis corresponds to the % methylation from pyrosequencing.

METHYLATED-CpG ISLAND RECOVERY ASSAY

by

NATALIE MITCHELL, J. TYSON DEANGELIS, TRYGVE O. TOLLEFSBOL

METHODS IN MOLECULAR BIOLOGY

Copyright

2011

By

HUMANA PRESS

Used by permission

Format adapted for thesis

ABSTRACT

Alterations in DNA methylation patterns are implicated in playing a major role in the development of cancer, thus highlighting the need to continually develop new technologies to analyze epigenetic marks. Methylated-CpG Island Recovery Assay (MIRA), based on the high affinity of the MBD2b/MBD3L1 complex for double-stranded methylated DNA, allows for the recovery of methylated DNA without the use of bisulfite conversion or antibody recognition. MIRA is capable of detecting low-density methylation of a single methylated CpG nucleotide. This technique can be used in conjunction with microarrays or next-generation sequencing to analyze recovered methylated DNA on a genome-wide scale.

Key words: DNA methylation, Methylated-CpG island recovery assay, Methyl-CpG binding protein, Microarray, Next-generation sequencing

1. INTRODUCTION

DNA methylation at CpG dinucleotides is an important epigenetic modification in mammalian cells. Epigenetic modifications describe the heritable changes in gene expression that occur without a change in the primary DNA sequence [1]. Regions of DNA where a large density of CpG dinucleotides exist are known as CpG islands which are found in the proximal promoter region in nearly half of the genes in the mammalian genome [2]. In normal cells, hypermethylated promoters serve to maintain genes in an inactive state that are not required for that specific cell type. During oncogenesis, genome-wide hypomethylation combined with site-specific hypermethylation leads to alterations in the expression of genes associated with tumor development [3,4]. Because changes in DNA methylation have been implicated in playing a central role in tumorigenesis, it is important that techniques be available to readily analyze changes in DNA methylation at CpG islands [5,6]. While numerous methods exist for analyzing DNA methylation, the Methylated-CpG Island Recovery Assay (MIRA) is of high interest. It is important to note that MIRA is sensitive enough to detect low-density methylation of a single methylated CpG nucleotide; however, it is the spacing between individual methylated CpG nucleotides that possibly limits the sensitivity of MIRA [7]. This technique can be used in combination with a variety of microarrays to analyze changes in the methylome [8].

In 1975, the first assay was developed to analyze DNA methylation at specific nucleotide sequences using methylation-sensitive restriction enzymes called isoschizomers. Isoschizomers are capable of cutting the same unmethylated DNA sequence; however, only one of the two restriction endonucleases will cut the nucleotide

sequence if it is methylated. Thus, the presence of DNA methylation at specific cytosine residues can affect the results obtained using restriction endonucleases. Methylation of a specific genomic sequence can be observed through Southern blot hybridization [9].

Since the initial attempt of locating DNA methylation in the mid- 1970's, a wide variety of other methods have been created. Although many techniques for detecting methylated CpG islands exist, sodium bisulfite treatment of genomic DNA has become the preferred method for studying DNA methylation. Bisulfite treatment of genomic DNA converts unmethylated cytosines to uracils by deamination and leaves methylated cytosines unaffected. After the area of interest is amplified through PCR, it is usually cloned and sequenced. Using PCR, this technique analyzes only a small region of the genome, obviously presenting a large limitation of the procedure [10,11]. In addition, the loss of a base through bisulfite sequencing compromises the complexity of data generated and requires comparison to the original sample sequence. Although bisulfite treatment of DNA has been incredibly valuable over the last few decades, the inability to compare DNA methylation density between large discrete genomic regions at the single nucleotide level has limited its usefulness in the post-genomic era.

The first next-generation assay for studying genome-wide DNA methylation was the methylated DNA immunoprecipitation microarray (MeDIP) technique. MeDIP is capable of detecting DNA methylation at CpG islands across the entire genome. MeDIP relies on antibodies that recognize 5-methylcytosine (5mC) to immunoprecipitate single-stranded DNA that contains heavily methylated CpG sites. The methylated DNA can then be further analyzed with DNA microarrays or parallel sequencing techniques by simply quantifying the regions that were recovered. Because MeDIP relies on genomic

regions that are heavily methylated, regions of DNA with a low CpG dinucleotide density may not be accurately analyzed by MeDIP despite the fact that DNA methylation changes may exist. In addition, antibodies can be expensive and rather variable in their recognition response [12].

The latest and most efficient means of analyzing genomic DNA methylation patterns is MIRA. MIRA does not rely on bisulfite conversion of DNA or antibody recognition of single-stranded DNA. MIRA can analyze DNA methylation patterns in double-stranded DNA and has a higher sensitivity and specificity than the aforementioned methods. Mammalian genomes encode several proteins that are capable of binding to methylated CpG dinucleotides and serve to repress transcription of the downstream gene [13,14]. While the first methyl-CpG binding (MBD) protein discovered was MeCP2, MIRA is based on the recognition of methylated CpG dinucleotides by the methyl-CpG-binding domain protein-2b (MBD2b). The binding of MBD2b to methylated DNA is sequence-independent, making it a more attractive candidate for isolating methylated DNA than MeCP2 [15,16]. MBD2b has one of the highest affinities for binding methylated DNA; however, the binding affinity is further increased by the addition of the methyl-CpG-binding domain protein 3-like-1 (MBD3L1) [7,15,17]. MBD3L1 binds with MBD2b *in vivo* and *in vitro* to promote the formation of methylated-DNA-binding complexes [17]. This complex is the central core that allows MIRA to successfully function.

The first step in MIRA is to fragment double-stranded DNA to an average length of 0.35 Kb by either sonication or by restriction endonuclease digestion. Methylated DNA fragments are then bound by the glutathione-S-transferase (GST)-tagged

recombinant MBD2b protein/His-tagged MBD3L1 complex. The DNA-protein complex is then captured using glutathione beads in an affinity column. Following capture, DNA is eluted from DNA-protein complexes and amplified with PCR for site-specific analysis [7]. See **Fig. 1** for an overview of the MIRA procedure.

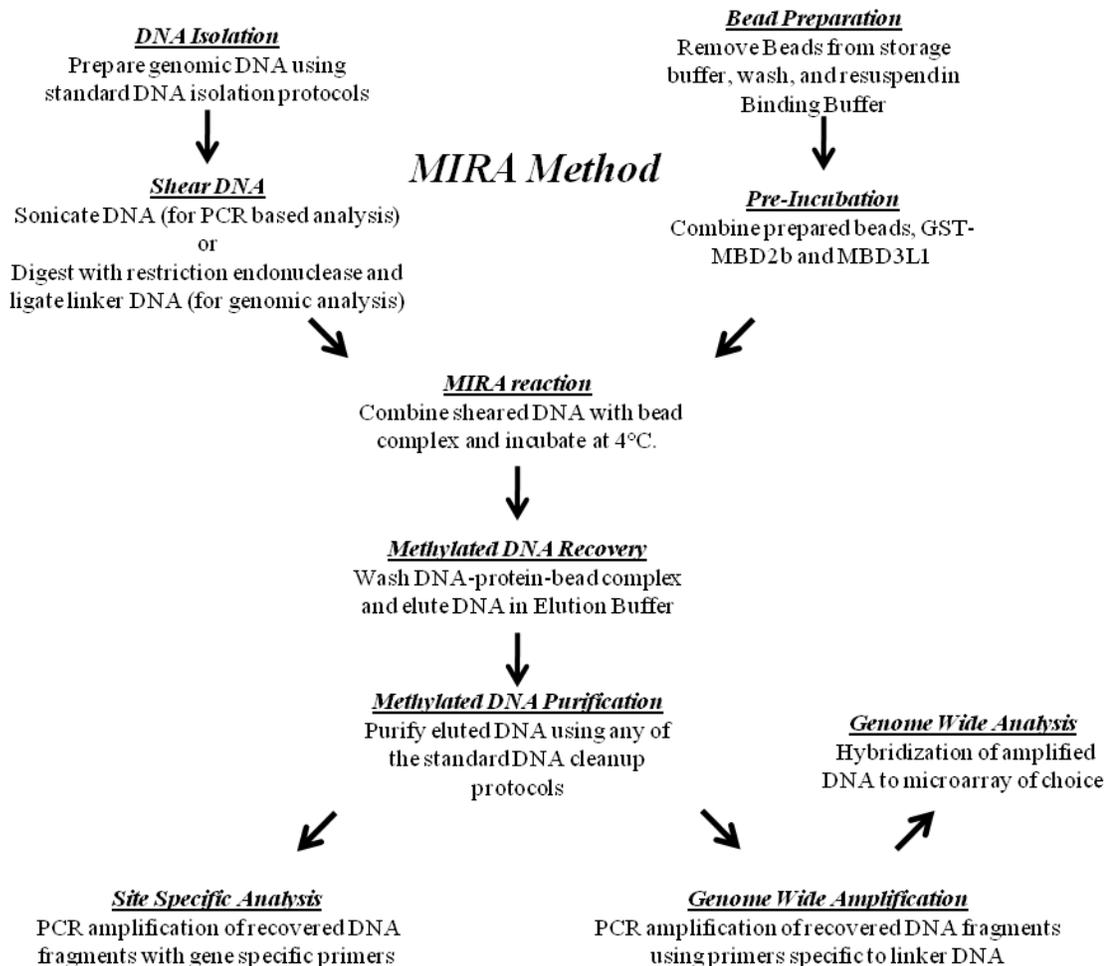


Figure 1. Schematic of the steps necessary to complete MIRA. The products of DNA shearing and the bead/protein complex are combined to perform the MIRA reaction. After the methylated DNA is recovered, purified and amplified with PCR, it can be analyzed via a variety of assays.

MIRA has several advantages over other current DNA methylation analysis techniques and may be applicable in a variety of clinical and diagnostic situations,

including characterization of DNA methylation in tissues or body fluids. Unlike MeDIP, MIRA works on double-stranded DNA and does not require the use of antibodies against 5-methylcytosine that is specific to single-stranded DNA. In addition, MIRA does not require the occurrence of methylation-sensitive restriction sites within the targeted sequence. Because it is not extremely laborious, MIRA is useful in genome-wide methylation analysis in cancers. MIRA may also be used to detect cell-type dependent differences in DNA methylation for a large number of genes through subsequent microarray analysis. In addition, the number of false-positives generated by the MIRA method is extremely low, making it an even more attractive candidate for analyzing DNA methylation at CpG dinucleotides [7,8].

Recently, Active Motif has developed a commercially available kit (*MethylCollector Ultra*) based on the MIRA assay. DNA is sheared and incubated with the His-tagged MBD2b/MBD3L1 protein complexes. The protein/DNA complex is captured with nickel-coated magnetic beads. Unmethylated DNA and protein are washed away and methylated DNA is eluted and ready for analysis. Active Motif has also developed an *UnMethylCollector* kit that recovers only unmethylated DNA. The combination of these two kits provides a very powerful tool able to analyze genome-wide DNA methylation patterns. The use of *UnMethylCollector* further validates experimentation results obtained by *MethylCollector Ultra*. Instead of assuming the status of DNA is unmethylated by negative identification with the *MethylCollector Ultra* kit, *UnMethylCollector* provides a positive method of ensuring collected data is unmethylated.

Multiple commercially available arrays exist for the analysis of recovered DNA fragments; this chapter will focus on describing methods to analyze the products of MIRA and current trends in epigenetic research.

2. MATERIALS

1. 1 mg purified glutathione-S-transferase-tagged recombinant Methyl-CpG-binding domain (MBD) protein-2b (GSTMBD2b) (see Note 1).
2. 1 mg purified His-tagged MBD3L1 (see Note 1).
3. Glutathione sepharose CL-4B matrix (Amersham Biosciences).
4. Binding buffer (10 mM Tris-HCl (pH 7.5), 50 mM NaCl, 1 mmol/L EDTA, 1 mM DTT, 3 mM MgCl₂, 0.1% Triton-X100, 5% glycerol, 25 mg/mL bovine serum albumin, and 1.25 mg sonicated JM110 (*dcm* minus) bacterial DNA).
5. Wash buffer (binding buffer containing 700 mM NaCl).
6. Guanidinium hydrochloride-containing Elution buffer.
7. Qiaquick PCR purification kit.
8. Genomic DNA.
9. Ultrasonic homogenizer with a microtip.
10. Benchtop microcentrifuge.

3. METHODS

3.1 DNA preparation

1. Genomic DNA from cell cultures or tissue samples should be prepared using any of the standard protocols.

2. Shear genomic DNA using Ultrasonic homogenizer to an average length of 0.35 Kb by sonicating each sample 5 times for 10 sec at an output setting of 30%. Sonicate samples in an ice bath to prevent overheating. See **Note 4.2** for additional information regarding shearing and linker ligation for microarray analysis.

3.2 Bead preparation

1. Centrifuge 100 µl of GST-sepharose slurry for 5 min at 500g. See **Note 4.3** for additional information bead preparation.
2. Carefully remove supernatant.
3. Wash beads 3 times with 500 µL of binding buffer.
4. Resuspend beads in 400 µL of binding buffer.

3.3 Preincubation

1. Combine 1 µg of GST-MBD2b and 1 µg of MBD3L1 with GST sepharose beads prepared in the previous step and incubate for 20 min at 4°C on a rotating platform. (See **Note 4.4**)

3.4 MIRA reaction

1. Combine 500 ng of sonicated DNA with binding reaction from the previous step and incubate for 4 h at 4°C on a rotating platform. (See **Note 4.5**)
2. Pellet beads by centrifugation at 500g for 5 min.
3. Wash beads 3 times with 500 µL of Wash Buffer. For each wash, incubate at room temperature (RT) for 5 min on a rotating platform followed by centrifugation at 500g for 5 min.

3.5 Elution and DNA purification

1. Following the third and final wash, resuspend beads in 100 μ L of Elution Buffer and incubate at 50°C for 1 h.
2. Pellet beads by centrifugation at 500g for 5 min.
3. Carefully remove supernatant and transfer to a fresh tube.
4. Purify methylated genomic DNA using the *Qiaquick PCR Purification Kit* following the manufacturer's protocol.

3.6 Amplification of Recovered DNA

1. To analyze a specific region of genomic, amplify with PCR using gene primers.
2. For microarray analysis, amplify all recovered DNA using primers specific to linker DNA. (See **Note 4.6**)

3.7. Analyzing the Genomic Products of MIRA

The MIRA-chip method can be used to determine the DNA methylation status of large numbers of genes and chromosomal regions of both normal and cancerous tissues. Rauch and colleagues characterized the DNA methylation status of the human HOX gene clusters by combining the MIRA method with genome-wide CpG island tiling arrays [18]. Site-specific PCR amplification can also be used in conjunction with MIRA to analyze DNA methylation changes at a particular gene of interest.

It was originally assumed that functionally important DNA methylation changes occurred in promoter regions and CpG islands. However, as recently noted by Feinberg and colleagues, many methylation alterations occur in sequences up to 2 Kb from CpG islands; such areas are termed CpG island shores. It is methylation changes at CpG shores that have been shown to discriminate normal tissue from cancerous tissue more so

than CpG island methylation status [19]. Thus, a distinct limitation exists in using only CpG island tiling arrays in conjunction with MIRA to analyze DNA methylation changes as all possible sites where methylation may occur are not included in the array.

This problem is partially solved through the use of a human promoter array, which allows the user to analyze promoter DNA from over 25,000 genes. The Affymetrix product covers 7.5 Kb upstream and 2.45 Kb downstream of the 5' transcriptional start site. For over 1300 cancer-associated genes, the promoter coverage was expanded to include 10 Kb upstream through 2.45 Kb downstream. Thus, the ability to analyze DNA methylation at CpG shores is possible; however, it does not test for all genomic regions, including exonic and intronic DNA.

Of further interest, however, is whole-genome tiling arrays (WGAs) which interrogate an entire genome in an unbiased fashion through the use of non-overlapping or partially overlapping probes spaced at regular intervals. The technique is unbiased due to the ability to include as yet unidentified genes. WGAs have never been used in conjunction with MIRA but have been used with MeDIP to map the *Arabidopsis thaliana* methylome [20]. After immunoprecipitation with an anti-5-methyl cytosine antibody, the complex is hybridized to a tiling array to map the methylome. However, the extremely high costs of WGAs make them an impractical choice for use in mammalian genomes. In addition, because WGAs have such a high level of sensitivity and assume no underlying gene models or annotations, results may contain an overlap of noise and signal [21]. It is important to note that Rauch and colleagues recently used MIRA in combination with whole-genome tiling arrays to characterize the entire B cell methylome of an individual human at 100-bp resolution [22]. Although progress is being made on current

microarray technologies, most study of DNA methylation patterns in mammalian genomes is currently limited to CpG islands and promoters.

Several techniques have been developed to analyze DNA methylation patterns on a genome-wide scale. Current methods involve either cleavage by methylation-sensitive restriction endonucleases [3], bisulfite sequencing [23] or precipitation of methylated DNA with an antibody [12]. However, these microarray methods are not capable of providing a high-resolution DNA methylation map of the mammalian genome.

Next-generation sequencing (NGS) allows for the sequencing of short DNA fragments at a cost much lower than the traditional Sanger method of sequencing. It is capable of single-nucleotide resolution and consequently is challenging microarrays as the tool of choice for analyzing the genome. Known as ChIP sequencing (ChIP-seq), NGS has been combined with chromatin-immunoprecipitation (ChIP) to sequence the genomic DNA fragments bound by transcription factors [24]. In addition, MIRA-seq provides extremely quantitative data about genome-wide methylation. MIRA enriched fragments undergo bisulfite conversion and analysis by NGS [25]. Thus, NGS has future implications of further analyzing epigenetic marks of the genome.

4. NOTES

1. Plasmids expressing both of the recombinant MBD proteins used in the MIRA assay are available upon request by contacting the original authors (see Ref. 7).
2. If microarrays are to be used to identify methylated genomic regions, genomic DNA should be digested with MseI and linker DNA ligated. Ligation of linker DNA allows for amplification of recovered DNA prior to microarray hybridization.

3. GST-beads are shipped in an ethanol slurry and bead preparation is performed to remove all traces of ethanol that may interfere in the MIRA reaction.
4. The high binding affinity of GST-MBD2b and MBD3L1 allows formation of the MBD2b-MBD3L1 complex *in vitro*. This step is critical to the recovery of methylated DNA and steps should be taken to ensure this part of the procedure is performed properly.
5. To ensure even mixing, samples should be lightly mixed by hand at each hour interval during the course of the 4 h incubation.
6. For genome wide analysis of recovered DNA, a much greater amount of DNA is required. Linker DNA that is ligated following enzymatic digestion can be used to amplify all recovered DNA.

ACKNOWLEDGEMENTS

This work was supported in part by a grant from the National Cancer Institute (RO1 CA129415). JTD was supported by an NCI Cancer Prevention and Control Training Program (R25CA047888).

REFERENCES

1. Holliday, R. (1994) Epigenetics: an overview. *Dev Genet*, **15**, 453-7.
2. Bird, A. (1986) CpG-rich islands and the function of DNA methylation. *Nature*, **321**, 209-13.
3. Costello, J.F., Fruhwald, M.C., Smiraglia, D.J., Rush, L.J., Robertson, G.P., Gao, X., Wright, F.A., Feramisco, J.D., Peltomaki, P., Lang, J.C., Schuller, D.E., Yu,

- L., Bloomfield, C.D., Caligiuri, M.A., Yates, A., Nishikawa, R., Su Huang, H., Petrelli, N.J., Zhang, X., O'Dorisio, M.S., Held, W.A., Cavenee, W.K. and Plass, C. (2000) Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat Genet.*, **24**, 132.
4. Jones, P. and Baylin, S. (2002) The fundamental role of epigenetic events in cancer. *Nat Rev Genet.*, **3**, 415-28.
 5. Fraga, M. and Esteller, M. (2002) DNA methylation: a profile of methods and applications. *Biotechniques*, **33**, 632, 634, 636-49.
 6. Shiraishi, M., Oates, A. and Sekiya, T. (2002) An overview of the analysis of DNA methylation in mammalian genomes. *Biol Chem*, **383**, 893-906.
 7. Rauch, T. and Pfeifer, G. (2005) Methylated-CpG island recovery assay: a new technique for the rapid detection of methylated-CpG islands in cancer. *Lab Invest*, **85**, 1172-80.
 8. Rauch, T., Li, H., Wu, X. and Pfeifer, G. (2006) MIRA-assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain-containing genes in lung cancer cells. *Cancer Res*, **66**, 7939-47.
 9. Singer, J., Roberts-Ems, J. and Riggs, A. (1979) Methylation of mouse liver DNA studied by means of the restriction enzymes msp I and hpa II. *Science*, **203**, 1019-21.
 10. Clark, S., Harrison, J., Paul, C. and Frommer, M. (1994) High sensitivity mapping of methylated cytosines. *Nucleic Acids Res*, **22**, 2990-7.

11. Frommer, M., McDonald, L., Millar, D., Collis, C., Watt, F., Grigg, G., Molloy, P. and Paul, C. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A*, **89**, 1827-31.
12. Weber, M., Davies, J., Wittig, D., Oakeley, E., Haase, M., Lam, W. and Schübeler, D. (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet*, **37**, 853-62.
13. Wade, P. (2001) Methyl CpG binding proteins: coupling chromatin architecture to gene regulation. *Oncogene*, **20**, 3166-73.
14. Hendrich, B. and Bird, A. (2000) Mammalian methyltransferases and methyl-CpG-binding domains: proteins involved in DNA methylation. *Curr Top Microbiol Immunol*, **249**, 55-74.
15. Fraga, M., Ballestar, E., Montoya, G., Taysavang, P., Wade, P. and Esteller, M. (2003) The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. *Nucleic Acids Res*, **31**, 1765-74.
16. Klose, R., Sarraf, S., Schmiedeberg, L., McDermott, S., Stancheva, I. and Bird, A. (2005) DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. *Mol Cell*, **19**, 667-78.
17. Jiang, C., Jin, S. and Pfeifer, G. (2004) MBD3L1 is a transcriptional repressor that interacts with methyl-CpG-binding protein 2 (MBD2) and components of the NuRD complex. *J Biol Chem*, **279**, 52456-64.

18. Rauch, T., Wang, Z., Zhang, X., Zhong, X., Wu, X., Lau, S., Kernstine, K., Riggs, A. and Pfeifer, G. (2007) Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay. *Proc Natl Acad Sci U S A*, **104**, 5527-32.
19. Irizarry, R., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., Ji, H., Potash, J., Sabunciyan, S. and Feinberg, A. (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet*, **41**, 178-86.
20. Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S., Chen, H., Henderson, I., Shinn, P., Pellegrini, M., Jacobsen, S. and Ecker, J. (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell*, **126**, 1189-201.
21. Gregory, B. and Belostotsky, D. (2009) Whole-genome microarrays: applications and technical issues. *Methods Mol Biol*, **553**, 39-56.
22. Rauch, T., Wu, X., Zhong, X., Riggs, A. and Pfeifer, G. (2009) A human B cell methylome at 100-base pair resolution. *Proc Natl Acad Sci U S A*, **106**, 671-8.
23. Meissner, A., Mikkelsen, T., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B., Nusbaum, C., Jaffe, D., Gnirke, A., Jaenisch, R. and Lander, E. (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, **454**, 766-70.
24. Hurd, P. and Nelson, C. (2009) Advantages of next-generation sequencing versus the microarray in epigenetic research. *Brief Funct Genomic Proteomic*, **8**, 174-83.

25. Pfeifer, G.P. and Rauch, T.A. (2009) DNA methylation patterns in lung carcinomas. *Semin Cancer Biol*, **19**, 181-7.

GENERAL LIST OF REFERENCES

1. Hanahan D, Weinberg RA (2000) The hallmarks of cancer. *Cell* 100: 57-70.
2. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144: 646-674.
3. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, et al. (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* 41: 178-186.
4. Hahn WC, Counter CM, Lundberg AS, Beijersbergen RL, Brooks MW, et al. (1999) Creation of human tumour cells with defined genetic elements. *Nature* 400: 464-468.
5. Hahn W, Dessain S, Brooks M, King J, Elenbaas B, et al. (2002) Enumeration of the simian virus 40 early region elements necessary for human cell transformation. *Mol Cell Biol* 22: 2111-2123.
6. DeCaprio JA, Ludlow JW, Figge J, Shew JY, Huang CM, et al. (1988) SV40 large tumor antigen forms a specific complex with the product of the retinoblastoma susceptibility gene. *Cell* 54: 275-283.
7. Counter C, Hahn W, Wei W, Caddle S, Beijersbergen R, et al. (1998) Dissociation among in vitro telomerase activity, telomere maintenance, and cellular immortalization. *Proc Natl Acad Sci U S A* 95: 14723-14728.

8. Knauf JA, Ouyang B, Knudsen ES, Fukasawa K, Babcock G, et al. (2006) Oncogenic RAS induces accelerated transition through G2/M and promotes defects in the G2 DNA damage and mitotic spindle checkpoints. *J Biol Chem* 281: 3800-3809.
9. Ahuja D, Sáenz-Robles M, Pipas J (2005) SV40 large T antigen targets multiple cellular pathways to elicit cellular transformation. *Oncogene* 24: 7729-7745.
10. Hahn WC, Dessain SK, Brooks MW, King JE, Elenbaas B, et al. (2002) Enumeration of the simian virus 40 early region elements necessary for human cell transformation. *Mol Cell Biol* 22: 2111-2123.
11. Bodnar AG, Ouellette M, Frolkis M, Holt SE, Chiu CP, et al. (1998) Extension of life-span by introduction of telomerase into normal human cells. *Science* 279: 349-352.
12. Roberts P, Der C (2007) Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene* 26: 3291-3310.
13. Hahn W, Counter C, Lundberg A, Beijersbergen R, Brooks M, et al. (1999) Creation of human tumour cells with defined genetic elements. *Nature* 400: 464-468.
14. Elenbaas B, Spirio L, Koerner F, Fleming MD, Zimonjic DB, et al. (2001) Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. *Genes Dev* 15: 50-65.
15. DeAngelis JT, Li Y, Mitchell N, Wilson L, Kim H, et al. (2011) 2D difference gel electrophoresis analysis of different time points during the course of neoplastic transformation of human mammary epithelial cells. *J Proteome Res* 10: 447-458.

16. Varley JM, Armour J, Swallow JE, Jeffreys AJ, Ponder BA, et al. (1989) The retinoblastoma gene is frequently altered leading to loss of expression in primary breast tumours. *Oncogene* 4: 725-729.
17. Hachana M, Trimeche M, Ziadi S, Amara K, Korbi S (2009) Evidence for a role of the Simian Virus 40 in human breast carcinomas. *Breast Cancer Res Treat* 113: 43-58.
18. Society AC (2009) *Cancer Facts and Figures*. Atlanta: American Cancer Society.
19. Holliday R (1994) Epigenetics: an overview. *Dev Genet* 15: 453-457.
20. Holliday R (2006) Epigenetics: a historical overview. *Epigenetics* 1: 76-80.
21. Turek-Plewa J, Jagodziński PP (2005) The role of mammalian DNA methyltransferases in the regulation of gene expression. *Cell Mol Biol Lett* 10: 631-647.
22. Bestor TH (2000) The DNA methyltransferases of mammals. *Hum Mol Genet* 9: 2395-2402.
23. Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99: 247-257.
24. Herman JG, Baylin SB (2003) Gene silencing in cancer in association with promoter hypermethylation. *N Engl J Med* 349: 2042-2054.
25. Esteller M (2005) Aberrant DNA methylation as a cancer-inducing mechanism. *Annu Rev Pharmacol Toxicol* 45: 629-656.

26. Costello JF, Fruhwald MC, Smiraglia DJ, Rush LJ, Robertson GP, et al. (2000) Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat Genet*, 24: 132.
27. Jones P, Baylin S (2002) The fundamental role of epigenetic events in cancer. *Nat Rev Genet* 3: 415-428.
28. Esteller M (2007) Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet* 8: 286-298.

APPENDIX A

PYROSEQUENCING PRIMERS FOR CpG LOCI OF INTEREST

Gene Symbol	Gene Name	Pyrosequencing Primers
<i>ZNF577</i>	zinc finger protein 577	<i>Forward: 5'-GGTTGAAAATGGTTGTTTTTAGTAAG-3'</i> <i>Reverse: 5'-CCAAACCCATAACCTATCTCCACTAC-3'</i> <i>Sequencing: 5'-GTAGATATATTGTATTATAAAGGGA-3'</i>
<i>HMGN3</i>	high mobility group nucleosomal binding domain 3	<i>Forward: 5'-GGTTTGGAGGTTATTAGTTGAATTG-3'</i> <i>Reverse: 5'-ATTTCCACTTCACCCAAAAAATCTACC-3'</i> <i>Sequencing: 5'-AGTTGAATTGGGTATGT-3'</i>

APPENDIX B

CpG LOCI EXHIBITING DIFFERENTIAL METHYLATION ONLY BETWEEN
SHMEC AND THMEC-40D

TargetID	SYMBOL	GENE_ID
cg00042156	LOC84856	84856
cg00069261	CDR2	1039
cg00231140	ACAA2	10449
cg00236832	RARA	5914
cg00308665	HTR2A	3356
cg00321478	CRB1	23418
cg00338702	CHFR	55743
cg00347729	MMP10	4319
cg00401678	EMR3	84658
cg00431114	TTPAL	79183
cg00462994	NAGS	162417
cg00463848	KRT2	3849
cg00557354	ARHGEF7	8874
cg00616135	LACTB	114294
cg00729275	C18orf16	147429
cg00769520	ST8SIA1	6489
cg00779924	FLJ45983	399717
cg00792687	C8orf41	80185
cg00796728	FUT4	2526
cg00832994	ATPAF1	64756
cg00876704	PKN1	5585
cg00888007	CCND2	894
cg00893242	SND1	27044
cg00920960	XIAP	331
cg00983899	CABP2	51475
cg00995327	CHST2	9435
cg01033938	ADAMTS8	11095
cg01043330	RIC3	79608
cg01136458	CSMD1	64478

cg01173186	NRSN1	140767
cg01200060	SCRT2	85508
cg01204985	LILRA4	23547
cg01214847	TMPRSS3	64699
cg01240931	APC	324
cg01267315	RNF220	55182
cg01277844	PYGO2	90780
cg01309671	FAM122C	159091
cg01335367	C12orf34	84915
cg01346718	CSNK1E	1454
cg01386493	ADRB3	155
cg01418124	ARHGAP24	83478
cg01437411	ATN1	1822
cg01442426	XCR1	2829
cg01495509	KCNMB1	3779
cg01511567	SSRP1	6749
cg01519742	JAKMIP1	152789
cg01520924	HSPA2	3306
cg01581111	RBM18	92400
cg01635061	COL11A2	1302
cg01731685	IL17RA	23765
cg01869233	LRRN4	164312
cg01888601	AGBL2	79841
cg01919488	ZNF638	27332
cg01939428	UBE3C	9690
cg01947224	ZNF416	55659
cg01968178	REEP1	65055
cg01968793	CNNM1	26507
cg02017109	WDR41	55255
cg02019333	UPK1B	7348
cg02033116	GPATCH1	55094
cg02089348	TMEM129	92305
cg02148642	RGPD5	84220
cg02168291	CDH13	1012
cg02181506	SERPINA1	5265
cg02197293	FBXO27	126433
cg02238747	RRP12	23223
cg02276793	SEC62	7095
cg02286642	ZNF254	9534
cg02305261	KCNG1	3755
cg02317742	ATOH8	84913

cg02326006	ARHGEF9	23229
cg02449608	C19orf18	147685
cg02493771	KRTAP13-2	337959
cg02523617	UBE2K	3093
cg02533173	BRD4	23476
cg02545106	CCDC110	256309
cg02635407	SH3TC1	54436
cg02668984	PDK2	5164
cg02671171	RPH3AL	9501
cg02681442	FOXG1	2290
cg02712878	FAIM	55179
cg02740947	RAD51L3	5892
cg02818322	MAGEC3	139081
cg02833725	ISG20L2	81875
cg02854288	FHIT	2272
cg02883230	PARP6	56965
cg02916816	JPH4	84502
cg02941816	MGMT	4255
cg02945646	AP1G2	8906
cg02953306	CHRNA2	1135
cg02957086	AXIN2	8313
cg02959669	ZNF256	10172
cg02966851	LYRM4	57128
cg03014957	DEFB118	117285
cg03017264	NFAM1	150372
cg03020597	SLITRK2	84631
cg03041841	TPRN	286262
cg03214697	SEC14L4	284904
cg03218374	ANGPT4	51378
cg03221436	CCDC120	90060
cg03221619	FCER2	2208
cg03266453	EN1	2019
cg03273615	RBM41	55285
cg03297731	GDPD3	79153
cg03301801	FNDC7	163479
cg03330678	9-Sep	10801
cg03405173	ZNF441	126068
cg03469082	GABRG1	2565
cg03532879	ADIG	149685
cg03548857	FFAR2	2867
cg03586879	RBFOX1	54715

cg03609102	MUC5B	4587
cg03672021	FLNA	2316
cg03679734	CKMT1B	1159
cg03724463	CCK	885
cg03731616	MYLK	4638
cg03740216	ICAM4	3386
cg03743584	PRAP1	118471
cg03780733	PCDHB7	56129
cg03791917	BTK	695
cg03811411	SGCD	6444
cg03812679	CSF3	1440
cg03872376	ZP4	57829
cg03884783	ZNF569	148266
cg03890877	TWF2	11344
cg03923464	ABCB9	23457
cg03954858	FAM83F	113828
cg03972838	SMCR7L	54471
cg03976567	AKD1	221264
cg04001802	OLAH	55301
cg04010530	NR4A2	4929
cg04014889	MAGEL2	54551
cg04086012	TRIML1	339976
cg04099420	RIPK1	8737
cg04126335	ITGA10	8515
cg04155793	CDK10	8558
cg04217218	ZDHHC21	340481
cg04278702	HTR1E	3354
cg04297093	KRTAP19-3	337970
cg04301375	ZNF558	148156
cg04386405	ZNF83	55769
cg04389838	ZNF501	115560
cg04435420	SGCD	6444
cg04452713	DST	667
cg04460372	ST6GALNAC6	30815
cg04498739	PDK4	5166
cg04531254	TMEM136	219902
cg04555771	CACNA2D2	9254
cg04575343	SLC2A13	114134
cg04662594	EPB49	2039
cg04698187	LRRN2	10446
cg04720330	PHLDA2	7262

cg04749104	STEAP3	55240
cg04756629	LGALS17A	400696
cg04761824	1-Dec	50514
cg04797496	PCDH12	51294
cg04828792	MX2	4600
cg04848452	EIF1B	10289
cg04880063	SFXN3	81855
cg04900486	RFX5	5993
cg04912050	IDUA	3425
cg04914105	PTPN9	5780
cg04949741	EFEMP2	30008
cg04991214	PFDN2	5202
cg05060602	DCTD	1635
cg05102817	RBM18	92400
cg05221264	CELA2A	63036
cg05228408	CLCN6	1185
cg05245861	C18orf34	374864
cg05250458	ZNF177	7730
cg05252264	FCAR	2204
cg05279864	ABCB4	5244
cg05337441	APOB	338
cg05345286	MDFI	4188
cg05354432	LEF1	51176
cg05379350	GIT1	28964
cg05412136	CLTA	1211
cg05421688	FAM163A	148753
cg05434957	ICA1	3382
cg05461841	ZG16B	124220
cg05471521	AURKA	6790
cg05532892	PSMA1	5682
cg05556717	CCL26	10344
cg05570980	C3orf52	79669
cg05592527	ARMCX6	54470
cg05603623	TIMM44	10469
cg05627103	KIAA1984	84960
cg05650171	PFDN2	5202
cg05664072	PER2	8864
cg05782445	KCNK3	3777
cg05835416	CYP2C18	1562
cg05898524	LMNA	4000
cg05955301	PRELP	5549

cg06051662	PDZD4	57595
cg06092815	SPHKAP	80309
cg06117855	CLEC3B	7123
cg06132342	KRT85	3891
cg06148264	SERPINB1	1992
cg06154597	TMEM175	84286
cg06156376	SHOX2	6474
cg06189133	INSL3	3640
cg06196379	TREM1	54210
cg06221259	REEP2	51308
cg06244906	ZIM2	23619
cg06245154	ERCC2	2068
cg06254453	HIST1H3F	8968
cg06311778	TENC1	23371
cg06351503	RDBP	7936
cg06457357	KLF8	11279
cg06532176	WVOX	51741
cg06539804	CPXM1	56265
cg06577005	DHDDS	79947
cg06646021	RAB4A	5867
cg06661994	C20orf195	79025
cg06692957	GNAQ	2776
cg06701500	YWHAQ	10971
cg06746171	ATOX1	475
cg06816106	SPDYA	245711
cg06836338	C17orf39	79018
cg06914598	RBAK	57786
cg06940792	MEGF10	84466
cg06960698	FGF14	2259
cg07014174	KRTAP11-1	337880
cg07066326	CHID1	66005
cg07109801	NDUFAF3	25915
cg07126559	SGCG	6445
cg07147350	FMR1	2332
cg07159490	NXPH4	11247
cg07191600	TRERF1	55809
cg07218880	UPF3A	65110
cg07237830	BSCL2	26580
cg07242414	BMP1	649
cg07331806	AIFM3	150209
cg07374637	KRTAP19-5	337972

cg07405796	TRIM40	135644
cg07426848	S100A3	6274
cg07443748	CCT8L2	150160
cg07465609	EYS	346007
cg07525077	RNASE3	6037
cg07546360	LOC400931	400931
cg07607921	ZZEF1	23140
cg07638935	KAZALD1	81621
cg07651857	RNF4	6047
cg07657236	KCNH8	131096
cg07686479	LMX1A	4009
cg07690018	CPNE7	27132
cg07747970	LHX1	3975
cg07786760	DCHS2	54798
cg07820214	RUVBL1	8607
cg07841014	REG1B	5968
cg07873488	MUC17	140453
cg07879977	OR1F1	4992
cg07884019	HMP19	51617
cg07889201	ZNF343	79175
cg07905963	CYP2A13	1553
cg07915343	LDHAL6A	160287
cg07940804	ACTRT1	139741
cg07952391	THNSL2	55258
cg07971188	HCP5	10866
cg07973395	BLM	641
cg07980518	RNF152	220441
cg08008233	TCEB3C	162699
cg08077345	PLEKHA4	57664
cg08093211	CSTF2	1478
cg08126211	KAAG1	353219
cg08132711	FGFBP2	83888
cg08214957	CDR1	1038
cg08244028	MSH3	4437
cg08250444	EFNA1	1942
cg08269321	RPRML	388394
cg08321346	ANKMY1	51281
cg08341667	ZNF488	118738
cg08353146	TMCC2	9911
cg08359956	TMEM176B	28959
cg08377000	TIGD2	166815

cg08403419	RXFP4	339403
cg08431931	WBP2NL	164684
cg08453021	ELMO1	9844
cg08458487	SFTPD	6441
cg08462247	MYEOV2	150678
cg08521225	TTR	7276
cg08525145	RXFP4	339403
cg08536841	IZUMO1	284359
cg08558340	SRRT	51593
cg08618113	CHP	11261
cg08623383	C1orf38	9473
cg08675664	ALDH7A1	501
cg08713365	NRSN2	80023
cg08748415	CLIC5	53405
cg08797471	DAPK1	1612
cg08878323	DLX5	1749
cg08897388	LAMA4	3910
cg08923450	HEY2	23493
cg08977371	CDH13	1012
cg09142399	CRYZ	1429
cg09197965	TUBA4A	7277
cg09279263	TADA3	10474
cg09313705	HOXB2	3212
cg09440289	PTPRO	5800
cg09588210	FLJ39739	388685
cg09601629	CCDC105	126402
cg09750083	ETAA1	54465
cg09809672	EDARADD	128178
cg09868882	GRM8	2918
cg09879797	WSB2	55884
cg09915099	AGPAT6	137964
cg09936839	SIRT6	51548
cg09949775	COMP	1311
cg10003443	FOXA2	3170
cg10011232	FKBP8	23770
cg10057065	C1orf104	284618
cg10065416	MPP3	4356
cg10078829	KLK4	9622
cg10129493	CD33	945
cg10149329	MFSD8	256471
cg10182321	STK32B	55351

cg10183248	UBTF	7343
cg10217449	GNPDA2	132789
cg10238818	CYYR1	116159
cg10249734	SECTM1	6398
cg10257049	C5orf4	10826
cg10316764	TAOK2	9344
cg10334385	C2orf50	130813
cg10358212	ATMIN	23300
cg10455133	ARAF	369
cg10467098	C11orf68	83638
cg10500167	ATP2B2	491
cg10519140	SLC29A1	2030
cg10521852	LPAR2	9170
cg10574499	NRN1L	123904
cg10605520	HRH3	11255
cg10631471	ABCC11	85320
cg10694152	SLC15A1	6564
cg10725344	FAM180A	389558
cg10763059	PTPN12	5782
cg10822172	CREB5	9586
cg10847798	LAMC1	3915
cg10857774	DMPK	1760
cg10862587	GLB1L	79411
cg10883303	HOXA13	3209
cg10894453	ITM2A	9452
cg10918419	C8orf55	51337
cg10922280	DPEP2	64174
cg10971790	FUZ	80199
cg10976172	ZNF785	146540
cg11105610	LGALS3BP	3959
cg11120551	CHD1L	9557
cg11177693	ZNF513	130557
cg11295113	FOLR2	2350
cg11368578	CDK5	1020
cg11368946	GCAT	23464
cg11371394	TGFBRAP1	9392
cg11387131	KCNJ5	3762
cg11395610	RAB28	9364
cg11408517	LATS1	9113
cg11503744	RUSC1	23623
cg11536940	PGCP	10404

cg11552293	COL4A6	1288
cg11647681	PCDHGA12	26025
cg11648289	EN2	2020
cg11661234	RLIM	51132
cg11668923	ADAM12	8038
cg11695266	DNAJB6	10049
cg11706163	C9	735
cg11714014	GABRQ	55879
cg11724759	RPE65	6121
cg11764747	EFNB1	1947
cg11768886	STK32B	55351
cg11828089	CALN1	83698
cg11876919	HMG20A	10363
cg11884546	ITGAX	3687
cg11933375	SLC38A7	55238
cg12005098	SLC16A12	387700
cg12058490	A2M	2
cg12064213	WWC3	55841
cg12114524	TNNI1	7135
cg12207371	CD163	9332
cg12365667	HLA-DPB1	3115
cg12382941	SYTL4	94121
cg12402251	CALB1	793
cg12437481	MRPL28	10573
cg12438037	C9orf116	138162
cg12485020	CHRDL1	91851
cg12555334	PSPH	5723
cg12680609	ZFP41	286128
cg12687990	PTENP1	11191
cg12741488	FLYWCH2	114984
cg12783776	SERPING1	710
cg12800028	GPR6	2830
cg12815142	SPAG7	9552
cg12847373	EDNRB	1910
cg12867068	CDKL1	8814
cg12894126	UCK1	83549
cg12902039	OCA2	4948
cg12914657	GIMAP1	170575
cg12928668	SEMA6A	57556
cg12938998	GAB3	139716
cg12955583	KNDC1	85442

cg12968903	TMEM154	201799
cg13005002	RFPL3	10738
cg13078388	CIRBP	1153
cg13140267	SNRNP200	23020
cg13150596	GIPC1	10755
cg13240311	PKIG	11142
cg13240598	FLVCR1	28982
cg13246269	AQP7	364
cg13299148	COL16A1	1307
cg13315690	OVCA2	124641
cg13325529	TMEM107	84314
cg13338132	FAM71C	196472
cg13340269	TMEM35	59353
cg13348944	SCN2B	6327
cg13397379	OR2C3	81472
cg13435381	LGR5	8549
cg13470920	VAV1	7409
cg13520715	C21orf129	150135
cg13588354	MC3R	4159
cg13590277	SYNPO	11346
cg13615963	CCR6	1235
cg13619408	SARS	6301
cg13804316	XKR8	55113
cg13824734	PSMB3	5691
cg13834567	TMEM145	284339
cg13842648	IL9	3578
cg13858139	ZNF43	7594
cg13901526	GYPC	2995
cg13950558	BLOC1S3	388552
cg13969584	AFG3L1P	172
cg13980719	TNP1	7141
cg14021073	BRSK2	9024
cg14036856	C1orf210	149466
cg14089692	ADAM33	80332
cg14155397	MAP2K1	5604
cg14186992	ZBTB48	3104
cg14209518	NNMT	4837
cg14217157	WHSC2	7469
cg14221171	TAC1	6863
cg14236389	C20orf197	284756
cg14243026	HMP19	51617

cg14244577	DDX19B	11269
cg14262937	OPRM1	4988
cg14329976	NEIL2	252969
cg14353201	ALDH3B2	222
cg14473016	RASD2	23551
cg14534464	NFE2L3	9603
cg14542839	PAMR1	25891
cg14561282	ADAM29	11086
cg14562990	LDOC1	23641
cg14586939	NPY2R	4887
cg14592099	PCDHB13	56123
cg14702264	C7orf42	55069
cg14787704	SREK1	140890
cg14870271	LGALS3BP	3959
cg14894216	CATSPER1	117144
cg15003158	CCDC125	202243
cg15009813	TNFSF12	8742
cg15043057	CTPS	1503
cg15107670	WT1	7490
cg15195412	CX3CL1	6376
cg15210427	CST9L	128821
cg15241084	TLR7	51284
cg15319457	HES6	55502
cg15375586	TBX22	50945
cg15377518	ZEB2	9839
cg15442702	FGFR4	2264
cg15454483	SNX12	29934
cg15543551	FGF12	2257
cg15590526	ZNF124	7678
cg15639045	SLC16A11	162515
cg15649452	ATOH8	84913
cg15696627	MSX1	4487
cg15755084	MSX1	4487
cg15822346	SLC16A10	117247
cg15842884	GPCPD1	56261
cg15905979	ST18	9705
cg15931943	FHIT	2272
cg15982419	RALBP1	10928
cg16082125	USP11	8237
cg16098981	TMEM90B	79953
cg16204757	MSN	4478

cg16215361	MUC15	143662
cg16240480	EDARADD	128178
cg16321029	SNRPN	6638
cg16334795	BACE2	25825
cg16354207	UBE2F	140739
cg16357921	CDK20	23552
cg16444968	CPNE7	27132
cg16464322	HNRNPL	3191
cg16511708	SS18	6760
cg16725130	MMP19	4327
cg16745616	ANGPTL4	51129
cg16786458	PPARGC1B	133522
cg16884569	RASSF2	9770
cg16891895	RGS16	6004
cg16929104	GDF15	9518
cg16948369	KLHL4	56062
cg16983159	TMEM173	340061
cg17001430	KIF25	3834
cg17001652	CXCL13	10563
cg17016000	RIN2	54453
cg17031773	KLK9	284366
cg17043922	C6orf106	64771
cg17054360	MTERF	7978
cg17177660	VEGFC	7424
cg17192247	MAPRE3	22924
cg17205788	TBX22	50945
cg17241310	BARHL2	343472
cg17243643	RDH5	5959
cg17252960	ID4	3400
cg17267907	DEFA1	1667
cg17269277	FGFR3	2261
cg17342759	PPP2R2B	5521
cg17357062	FCN1	2219
cg17371081	NELL1	4745
cg17387870	CHFR	55743
cg17408686	CHCHD6	84303
cg17441778	MYO7A	4647
cg17468440	PPP2R1B	5519
cg17474651	MAGEC3	139081
cg17589341	SLC14A1	6563
cg17593391	RRAGC	64121

cg17607973	MEPCE	56257
cg17612991	C3	718
cg17662177	CFP	5199
cg17675882	NDRG3	57446
cg17685628	NKX2-8	26257
cg17687962	KLK3	354
cg17808849	HERPUD1	9709
cg17827767	LRIT1	26103
cg18056600	ZMYND15	84225
cg18108623	SLFN11	91607
cg18159180	CUL7	9820
cg18219951	GABRG2	2566
cg18230771	ACAN	176
cg18239753	KHDRBS2	202559
cg18259391	SERBP1	26135
cg18383391	MON1B	22879
cg18397523	EVX1	2128
cg18530716	SLC16A11	162515
cg18587271	GPT2	84706
cg18592174	CHAT	1103
cg18602314	GFPT2	9945
cg18712919	BRE	9577
cg18771300	RHOJ	57381
cg18804206	FLII	2314
cg18807515	PRAMEF2	65122
cg18829411	HMG1	3150
cg18833140	HABP2	3026
cg18840461	ANKRD43	134548
cg18843688	NUFIP1	26747
cg18869368	MAGEH1	28986
cg18871276	LRRC9	341883
cg18885999	CMPK1	51727
cg18888520	ZSCAN18	65982
cg19014419	ZNF300	91975
cg19094438	BCAR1	9564
cg19114595	DIRAS3	9077
cg19135982	SKP2	6502
cg19220825	AQP5	362
cg19224164	NOP14	8602
cg19241744	SLC29A3	55315
cg19325985	HOXD13	3239

cg19355190	EGR2	1959
cg19378133	RBFOX1	54715
cg19486673	LILRA2	11027
cg19514928	TMEM56	148534
cg19573166	SLC22A17	51310
cg19574623	SSPN	8082
cg19598544	TTYH2	94015
cg19614321	RASSF2	9770
cg19630689	SERPINA12	145264
cg19635712	EPB41L4B	54566
cg19682367	TPRG1	285386
cg19730092	ADAMTS14	140766
cg19777470	CRABP1	1381
cg19831077	WDR86	349136
cg19868730	POPDC2	64091
cg19899621	ZNF561	93134
cg19914607	SLC38A3	10991
cg19965810	KCNH7	90134
cg19976467	TMCO4	255104
cg19998328	C19orf52	90580
cg20017995	MED22	6837
cg20029201	BCL9L	283149
cg20052718	TWIST1	7291
cg20057066	FLJ42957	400077
cg20083730	MT1E	4493
cg20092728	SLC5A12	159963
cg20161179	MSX1	4487
cg20240860	ACCS	84680
cg20261915	GLP2R	9340
cg20287640	IRF1	3659
cg20312687	DEFB118	117285
cg20357628	PHACTR3	116154
cg20427865	CX3CL1	6376
cg20437604	ANXA9	8416
cg20450764	TSPAN7	7102
cg20451369	C10orf57	80195
cg20542190	LILRA4	23547
cg20550118	CRABP1	1381
cg20557202	SLC5A5	6528
cg20588069	MSX1	4487
cg20611872	XAGE3	170626

cg20684549	LPPR3	79948
cg20807545	ADAMTS18	170692
cg20812929	DHRS4L2	317749
cg20831708	SEC31B	25956
cg20955688	TMEM71	137835
cg21053323	SUMO3	6612
cg21080294	PRPS1	5631
cg21146268	HIST1H2BE	8344
cg21165219	FAM38B	63895
cg21180599	TLE6	79816
cg21229859	MYEF2	50804
cg21274570	CCL25	6370
cg21372914	CLEC4M	10332
cg21440587	AIF1	199
cg21488617	DNALI1	7802
cg21501064	6-Sep	23157
cg21519900	C20orf186	149954
cg21578541	TLR9	54106
cg21621248	LRRTM1	347730
cg21660727	PCGF3	10336
cg21690806	SLC9A8	23315
cg21692936	SLC3A2	6520
cg21787291	NKAIN2	154215
cg21790626	ZNF154	7710
cg21832243	TTC3	7267
cg21912567	GADD45G	10912
cg21918500	ZNF124	7678
cg21966410	AR	367
cg22062068	RBMXL2	27288
cg22175764	ANKRD60	140731
cg22187630	CACNA1A	773
cg22195268	FAM104A	84923
cg22197830	TXNDC15	79770
cg22199080	AGAP2	116986
cg22264436	SOST	50964
cg22315542	LCORL	254251
cg22374901	TESC	54997
cg22375610	APOBEC2	10930
cg22392666	FXYP7	53822
cg22395019	GALNT14	79623
cg22419075	DPP7	29952

cg22461835	ADRA1A	148
cg22462235	LEFTY2	7044
cg22467567	IGFBP5	3488
cg22472290	ZNF577	84765
cg22475430	ATN1	1822
cg22499964	EGLN2	112398
cg22527345	CCDC71	64925
cg22534509	GPR81	27198
cg22627950	TMED4	222068
cg22689690	ADCY6	112
cg22757447	IMPACT	55364
cg22759686	ARSD	414
cg22764341	ATP10D	57205
cg22775000	TMEFF1	8577
cg22794078	C15orf44	81556
cg22799321	TFPI2	7980
cg22833027	CHRD	8646
cg22876908	DGKA	1606
cg22889448	EMR1	2015
cg22930187	ARTN	9048
cg22960185	TMEM38A	79041
cg22968401	PIPOX	51268
cg22998840	TCEAL3	85012
cg23001650	B4GALT2	8704
cg23003832	ZDHHC12	84885
cg23013864	CAP2	10486
cg23020402	TDP2	51567
cg23026995	XPNPEP2	7512
cg23050981	L1CAM	3897
cg23054437	MOSC2	54996
cg23060239	RUSC1	23623
cg23178308	C21orf124	85006
cg23208152	ZMYND10	51364
cg23286660	PLOD1	5351
cg23290344	NEFM	4741
cg23303408	POU4F3	5459
cg23309825	MYL6B	140465
cg23317501	UGT3A1	133688
cg23320649	C3orf18	51161
cg23338195	SLC30A8	169026
cg23384620	IL1RAPL2	26280

cg23428445	GPR37	2861
cg23473904	COL6A2	1292
cg23563234	PCDHGB7	56099
cg23606079	CRISPLD2	83716
cg23654219	TCF19	6941
cg23673107	ZNF491	126069
cg23679724	CTSZ	1522
cg23680518	SBSN	374897
cg23699324	CTNNA2	1496
cg23828595	PRKG1	5592
cg23835812	CHMP2B	25978
cg23850272	ZFP112	7771
cg23906760	UBL5	59286
cg23912721	HACL1	26061
cg23916845	DFNB31	25861
cg23959705	TNFRSF9	3604
cg23984130	IGKV7-3	28905
cg24058120	MFI2	4241
cg24194132	B4GALNT4	338707
cg24206256	MAT2B	27430
cg24207176	ABCG2	9429
cg24286301	GUCY2F	2986
cg24346637	UBL4B	164153
cg24356797	UBA52	7311
cg24357161	RCVRN	5957
cg24365867	MRVI1	10335
cg24475171	C9orf78	51759
cg24484296	ZNF320	162967
cg24509668	SSPN	8082
cg24512303	HNF4G	3174
cg24623694	PRX	57716
cg24642468	C16orf78	123970
cg24642820	NUP210	23225
cg24646414	GATA4	2626
cg24662718	VAV3	10451
cg24710073	TBP	6908
cg24710886	PUM1	9698
cg24743283	GNA12	2768
cg24759821	PAH	5053
cg24812167	SNAPC4	6621
cg24816866	PARK2	5071

cg24861272	ZEB1	6935
cg24890043	TMEM163	81615
cg24919972	FYCO1	79443
cg25048564	HUNK	30811
cg25063710	BRWD3	254065
cg25195673	GFPT1	2673
cg25229706	SAMD8	142891
cg25234611	VWA1	64856
cg25268718	PSME1	5720
cg25277950	EML2	24139
cg25355803	MARVELD1	83742
cg25366404	MOCS1	4337
cg25383093	PGM3	5238
cg25427638	CYP2A7	1549
cg25430696	TSR1	55720
cg25446086	SYT11	23208
cg25455753	C15orf32	145858
cg25463428	TBC1D10B	26000
cg25554036	WFS1	7466
cg25569462	TRIML2	205860
cg25580018	ADAMTSL2	9719
cg25607161	OR1A1	8383
cg25608041	TBC1D1	23216
cg25635352	B3GALNT1	8706
cg25687894	ACLY	47
cg25711779	EFEMP1	2202
cg25714610	FAM3A	60343
cg25766774	ZDHHC3	51304
cg25787984	DKK3	27122
cg25808906	DGKG	1608
cg25816468	PBXIP1	57326
cg25843439	OR10H3	26532
cg25856811	SPRR3	6707
cg25864727	NCRNA00158	54072
cg25870263	FNDC5	252995
cg25920792	HTRA1	5654
cg26018901	CORIN	10699
cg26091142	ISLR2	57611
cg26124016	RARB	5915
cg26202340	TIRAP	114609
cg26209058	EIF5A2	56648

cg26233914	ITGAX	3687
cg26256793	COL11A1	1301
cg26270746	GSX1	219409
cg26293512	TEPP	374739
cg26299767	REM1	28954
cg26333317	METTL7A	25840
cg26333591	PTPN4	5775
cg26379475	SH2D1B	117157
cg26412379	NLGN4X	57502
cg26436315	KLF15	28999
cg26446827	ZNF133	7692
cg26473272	SYT8	90019
cg26504021	IRX2	153572
cg26551843	RPS6KL1	83694
cg26606552	CCDC22	28952
cg26624134	KCNJ5	3762
cg26738010	CETN1	1068
cg26850754	CD8B	926
cg26898336	TEKT3	64518
cg27000831	CCL8	6355
cg27025856	HDAC11	79885
cg27033479	TMEM44	93109
cg27069753	CELA3B	23436
cg27187881	NAGA	4668
cg27223047	FBN2	2201
cg27264345	REEP2	51308
cg27279652	ITPR2	3709
cg27285599	ZNF750	79755
cg27305303	OTOF	9381
cg27420123	FSHB	2488
cg27442349	NFKBIB	4793
cg27456885	KRTAP22-1	337979
cg27458888	UBE3A	7337
cg27478659	KRT16	3868
cg27560292	TAF10	6881
cg27562023	RPH3AL	9501
cg27619475	SLC16A5	9121
cg27622610	OR1G1	8390

APPENDIX C

CpG LOCI EXHIBITING DIFFERENTIAL METHYLATION ONLY BETWEEN
THMEC-40D AND THMEC-80D

TargetID	SYMBOL	GENE_ID
cg00059740	C7orf49	1299
cg00404599	TSC22D3	4833
cg00445824	ISYNA1	245973
cg00520135	TPM1	51305
cg00837103	COL9A3	126129
cg00933411	DLC1	3784
cg01049530	BMP3	158471
cg01056568	GLP2R	7490
cg01169778	GBGT1	112937
cg01517728	SOCS6	6870
cg01594262	YIF1B	796
cg01617750	CMTM8	26297
cg01717150	KCNK9	79778
cg01820777	MICALL2	8322
cg02026235	RHBDL1	2261
cg02189785	CLIC3	1812
cg02204046	MYCN	3000
cg02244695	TMEM176A	2253
cg02271621	DUSP8	2691
cg02501779	CBLN4	283078
cg02620769	CCDC65	221421
cg02672220	SIM2	10683
cg02717866	C11orf66	84969
cg02827112	SMARCAD1	796
cg02849695	CCDC19	115557
cg02990612	GATA4	3394
cg02994956	NEFH	3351
cg03000846	RAC3	9306
cg03139057	DLL3	6493

cg03544320	CRMP1	84765
cg03588357	GPR68	1299
cg03617456	EXOSC10	517
cg03993463	KCNJ15	56916
cg04180868	PRTFDC1	1831
cg04263186	TACR3	10008
cg04369341	TOX2	54848
cg04411625	CRIP1	6588
cg04549333	ALX4	3207
cg04600618	RSPH9	388759
cg04987894	GSTM5	10257
cg05221167	ZNF560	90522
cg05222924	WT1	129804
cg05454446	BMS1	5999
cg05924583	TP73	5079
cg05961212	ADPRH	162517
cg06469542	ISM2	84969
cg06497752	COL9A3	91133
cg06498267	HCN1	148930
cg06621126	HSF4	145501
cg06637774	P2RY6	147837
cg06722633	GRIK3	79957
cg06760035	HOXB4	79944
cg06825142	DRD4	8115
cg06882877	PAQR6	84229
cg06905514	CAMK2B	60529
cg06945625	SERPINB6	78996
cg07184439	MGAT2	140689
cg07195282	ZNF563	55365
cg07360076	FGFR3	10518
cg07636178	HIST1H3C	79006
cg07699362	GPX3	5394
cg07778029	HOXA9	8352
cg07816439	C1QTNF2	2878
cg07845392	SLC25A10	2834
cg07880854	FBLN7	8382
cg07895149	CALHM2	51305
cg08108311	WNK4	10395
cg08260959	HIST1H4F	26086
cg08460435	HENMT1	3784
cg08614481	HTR1B	5947

cg09068492	CALCA	51296
cg09160477	SUSD3	358
cg09210315	SLCO4A1	144347
cg09212468	CCDC135	64399
cg09427311	ANGPTL2	9607
cg09501516	RGS4	9002
cg09827833	TEK	1890
cg09872616	MEST	116443
cg10059959	PAX5	51477
cg10163825	CCDC78	91179
cg10467022	CALCA	5269
cg10720654	PTENP1	9324
cg10737521	TBC1D9B	136371
cg11003133	AIM2	1272
cg11027330	METRNL	2899
cg11086066	NFATC2	348807
cg11102794	LMO1	55687
cg11319389	TOX2	9509
cg11376198	AKR7L	284618
cg11418559	ZNF593	3854
cg11452221	ARHGEF25	84698
cg11498156	TLX1	4773
cg11504740	GPR152	7975
cg11630242	AKAP10	23555
cg11787522	STRA6	53346
cg12237269	SLN	79661
cg12281657	C14orf50	26301
cg12448933	RAB37	4613
cg12600197	NEIL1	4744
cg12610070	TSPAN15	2626
cg12683929	BLOC1S1	3188
cg12795208	KRT6B	2647
cg13081704	KCNQ1DN	51706
cg13176979	HMG3	3553
cg13282837	TCL1A	3772
cg13289321	ASB10	28231
cg13707560	NME5	4232
cg13723482	HNRNP2	126206
cg13749822	HHIP	1745
cg14409941	ADAMTS2	85478
cg14614211	MKX	8352

cg14696396	TM6SF1	2946
cg14785479	SCARF2	360
cg14882700	OTOP1	60529
cg14934821	GPSM1	2199
cg15236866	DLX1	115106
cg15748507	PRLHR	147381
cg15760840	HOXA11	3306
cg15836722	IL1B	25790
cg15942481	CAPS2	1468
cg16014085	ZNF48	27254
cg16268563	ATP6V1G2	2949
cg16584573	FGF8	142680
cg16604516	FBLN2	3549
cg16670497	GSTM2	651
cg16713727	SLC25A34	30812
cg16731240	ZNF577	1850
cg16834431	ATPIF1	10672
cg16998353	TRIM3	5305
cg17049328	GHRH	23061
cg17285325	TYMP	5031
cg17307280	DRD1	9790
cg17688525	L3MBTL4	7010
cg17820828	KCNQ1	9022
cg17834752	KCNK9	51042
cg17853587	NDST3	5458
cg18106189	C18orf1	6300
cg18149207	RORC	220004
cg18275051	CYB5R1	6097
cg18307767	HAUS1	348980
cg18411891	FZD4	11211
cg18568653	SERGEF	124093
cg18676162	NME4	7349
cg18794577	GRIN3A	145376
cg19006008	F2RL3	9348
cg19044630	PRUNE2	55539
cg19144013	PIR	9028
cg19145398	FOXS1	3205
cg19358442	ALX4	1815
cg19674669	GLB1L3	3299
cg19778698	MAFK	114898
cg20028470	UCN	152189

cg20176648	AQP1	9340
cg20189937	L2HGDH	3205
cg20312228	CCDC37	3214
cg20498685	TWIST1	753
cg20723355	FBXO39	3195
cg20761322	CIB2	4004
cg20798152	CARTPT	65266
cg21453309	FAM101A	60529
cg21604803	CPT1C	246181
cg21696393	SOX8	197407
cg21902544	CBLN2	1003
cg21974239	MAPK12	133060
cg21992250	SLC15A3	390212
cg22234962	C1orf104	10612
cg22243662	ATP6V1C2	51063
cg22319147	CDH5	11216
cg22546318	ATP5G2	23452
cg23054883	FZD10	11191
cg23124755	AQP3	7168
cg23186643	NPB	534
cg23189044	KCNE3	4247
cg23363832	RBP1	284723
cg23695504	C1orf229	7161
cg24165760	TRAF3	326624
cg24199834	POU4F2	7187
cg24476569	ABCC4	113802
cg24642523	HSPA2	64220
cg24747122	GNA13	8361
cg24826867	IRF8	8111
cg25047280	HOXA9	256933
cg25370441	ARHGEF38	7291
cg25438963	HIST1H3C	9447
cg25447894	CSDC2	1400
cg25465406	GUCY2D	2307
cg25908985	IHH	8544
cg26117023	DOK1	141
cg26264314	NLRP5	1796
cg26365854	ALX4	93974
cg26368842	PIP4K2A	5881
cg26660631	KNCN	1396
cg26681123	SLC34A3	147741

cg27032184	TRMU	816
cg27352992	CNTN1	56952
cg27491887	KCNQ1	203328