5 CHAPTER 5: Epigenetic Modification in Biomarker Discovery

5.1 Introduction

There are many ways by which gene expression is controlled in eukaryotes, one such mechanism is methylation of DNA, which is a common epigenetic signaling tool that cells use to lock genes in the "off" position. More recent research on DNA methylation has led to understanding of gene regulation caused due to DNA methylation. Role of methylation in numerous cellular processes, including embryonic development, genomic imprinting, X-chromosome inactivation, and preservation of chromosome stability has been well established. Given the large number of processes in which methylation plays a part, it is perhaps not surprising that a lot of work have been done in linking aberrations in methylation to several human diseases including cancer. DNA methylation occurs at the cytosine bases of eukaryotic DNA, which are converted to 5-methylcytosine by DNA methyltransferase (DNMT) enzymes. The altered cytosine residues are usually immediately 5’ to a guanine nucleotide, resulting in two methylated cytosine residues sitting diagonally to each other on opposing DNA strands. In mammals, majority of methylation is found CpG Island. The methylation of these sequences can lead to inappropriate gene silencing, such as the silencing of tumor suppressor genes in cancer cells (Moore et al., 2013).

Methylation is believed to play a crucial role in repressing gene expression by altering the chromatin state. It is very well documented that binding of methyl CpG binding protein, recruits HDAC’s to the promoter region thereby altering the nucleosomes and the DNA is packaged into condensed form. The tissue specific pattern of gene expression correlates with methylation states. In cancer cells versus normal, the levels of methylation at specific promoters are altered. The methylation states of tumor suppressor genes and oncogenes are altered in cancer cells.

Given the critical role of DNA methylation in gene expression and cell differentiation, it seems obvious that aberration in methylation could give rise to various diseases, including cancer. A large amount of research on DNA methylation and disease has focused on cancer and tumor suppressor genes. Tumor suppressor genes are often silenced in cancer cells due to hypermethylation (Kazanets et al., 2016). In contrast, the genomes of cancer cells have been shown to be hypomethylated overall when compared to normal cells, with the exception of hypermethylation events at genes involved in cell cycle regulation, tumor cell invasion, DNA repair, and others events in which silencing propagates metastasis (Ehrlich, 2009). In colorectal cancer hypermethylation is detectable early and might serve as a biomarker for the disease (Rasmussen et al., 2016).
LncRNAs have cell-type-specific expression patterns, and increasing evidence has shown altered expression level of lncRNAs in various types of human cancer. Dysregulated lncRNAs may function as tumor suppressors or oncogenes, which may be used as potential early tumor diagnostic, metastatic or prognostic markers and molecular-targeted therapy sites in the future (Inamura, 2017). As we have discussed before that a lot of lncRNA have been shown to play a role in epigenetic regulation, but not much work have been done in understanding the methylation status of lncRNA in cancer. Here we have worked to understand the methylation status of lncRNA in prostate cancer when compared to normal and to correlate with their expression.

5.2 Results

5.2.1 Differential methylation analysis

We performed differential methylation analysis using the dataset SRP016558 (Lin et al., 2013). The dataset encompass of 7 primary prostate sample and 7 benign prostate tissue sample. Upon differential methylation analysis almost equal number of hypermethylated and hypomethylated CpG islands were observed. These were uniformly distributed across all the chromosomes with a few exception. For example chromosomes 1,4,8,16,20 and 22 were high in hypomethylated CpG islands whereas chromosome 17 was high in hypermethylated CpG islands (Figure 5.1). The hypomethylation of chromosome 8 might be the reason why it highly transcribed region in prostate cancer, harboring some of the hallmark coding and non-coding RNA like MYC and PCAT1 respectively. Clustering of cancer and normal samples was performed based on the CpG island methylation status. The methylation profiles from samples were clustered to create two clads cancer and normal with an exception of two normal samples going in the cancer clade (Figure 5.1). As these are matched normal samples, there is always a chance of cross contamination of cancer tissue while extracting normal sample.
5.2.2 Correlation between promoter methylation and gene expression

After obtaining the differentially methylated region, we intersected them with the promoter region of both coding and non-coding RNA. Previous work in Chapter 3 gives a list of 366 coding genes and 121 nlncRNA which are differentially expressed in prostate cancer (Bawa et al., 2015). In this study we did a correlation analysis to see how many of these have their promoter region differentially methylated in prostate cancer. Out of 366 differentially expressed coding genes, 122 had significant differential methylation in there CpG island promoter region. Among them 83 were hypermethylated and 39 were hypomethylated. Among the 83 hypermethylated promoter, a majority of genes (69) were significantly down-regulated based on our RNA Seq analysis. A few of the hypermethylated promoter regions (14) showed anti-correlation with their methylation status as they were up-regulated based on our RNA Seq analysis. 39 out of 122 differentially methylated promoter regions were hypomethylated. Among the hypomethylated promoter region, 26 were up-regulated based on our RNA Seq analysis showing a positive correlation between the methylation status and the gene expressions, whereas 13, were down-regulated showing a negative correlation. Appendix Section 8.1.1 summaries the gene expression and methylation status of the 366 coding gene.
Following the correlation of coding gene expression with their CpG island promoter methylation, we did similar correlation of 121 significantly differential expression lncRNA with their CpG island promoter methylation. As there was no well-defined promoter region for lncRNA, we considered 2000bp upstream to TSS for each lncRNA as its respective promoter region. Among the previously reported 121 differentially expressed lncRNA, 32 were significantly differentially methylated in this study. Among the 32 lncRNA with differential methylated region, 8 were hypermethylated with 6 being down-regulated and 2 being up-regulated. Out of 121 differentially expressed lncRNA, 24 were hypomethylated in this analysis. Interestingly all of them were up-regulated in our RNA Seq analysis. Appendix Section 8.1.2 summaries the gene expression and methylation status of the 121 lncRNA.

5.2.3 Validation of promoter methylation in selected lncRNA

The methylation status of few lncRNA was validated using Methylation Specific PCR (MS-PCR) in prostate cancer cell line (PC3). We selected those lncRNA which had a positive correlation between expression and methylation. Candidate lncRNA were selected for CpG island of length more than 100bps. MSP primers were designed for the four selected and methylation specific PCR was carried out. Figure 5.2 summarizes the MSP PCR results along with the RNASeq and BisulfiteSeq data for the selected lncRNA.

<table>
<thead>
<tr>
<th>Non-coding RNA ID</th>
<th>Tissue Specificity</th>
<th>Expression status using RNA-Seq</th>
<th>Methylation status using Bisulfite-Seq</th>
<th>Methylation status using MSP PCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCONS_00017728</td>
<td>Prostate</td>
<td>Down regulated</td>
<td>Hyper methylated</td>
<td><img src="image1" alt="MSP PCR Image" /></td>
</tr>
<tr>
<td>TCONS_00023188</td>
<td>Kidney</td>
<td>Up regulated</td>
<td>Hypo methylated</td>
<td><img src="image2" alt="MSP PCR Image" /></td>
</tr>
<tr>
<td>TCONS_00003042</td>
<td>-</td>
<td>Up regulated</td>
<td>Hypo methylated</td>
<td><img src="image3" alt="MSP PCR Image" /></td>
</tr>
<tr>
<td>TCONS_00029157</td>
<td>-</td>
<td>Down regulated</td>
<td>Hyper methylated</td>
<td><img src="image4" alt="MSP PCR Image" /></td>
</tr>
</tbody>
</table>

Figure 5.2 Methylation sensitive PCR

Summary of the expression and methylation status of a few selected lncRNA, along with its methylation status using MSP PCR.
5.3 Discussion

A lot of work has been done to elucidate the potential of non-coding RNAs in methylation. However, the role of methylation in the expression of non-coding RNAs is understudied. In this work, we have demonstrated that differential methylation also plays a role in regulating the expression of non-coding RNAs to affect the underlying biology. We have already reported in Chapter 3 that the majority of nIncRNA are up-regulated in prostate cancer. Our interest here was to check what percentage of nIncRNA over-expression is driven by differential methylation. For example, we have shown that among the 366 prostate specific differentially expressed coding genes that shows differential methylation at the promoter regions 77% of the methylation percentage correlated with the expression. Similarly, out of 121 prostate specific differentially expressed nIncRNA, which are differentially methylated in CpG island 93% show correlation between methylation status and expression. These observation suggest that methylation is also an important mechanism of regulation of nIncRNA in cancer.

Furthermore it validates our finding that majority of the nIncRNA up-regulated in cancer. As part of our prioritization, we have shown that both SIK1 known tumor suppressor factor and TCON_00029517 on the same locus on chromosome 21 are both hypermethylated and down-regulated in prostate cancer. This down-regulation and methylation status of the transcript pair was also validated using qRT-PCR in prostate cancer cell line. We have shown that this kind of integrated approach helps in further prioritization of the candidate transcript.

Recent reports from lung cancer have also shown that SIK1 and TCON_00029517 are both downregulated. The lower expression of the pair have been associated with lung node metastasis suggesting the regulatory role of TCON_00029517 in occurrence and progression of lung cancer. Using overexpression experiments, they have shown that increased level of TCON_00029517 effect cancer cell viability, suppress proliferation, migration and invasive abilities of lung cancer cell line. Though the molecular mechanism of TCON_00029517 is yet to be developed but based on this work and ours, the potential role of TCON_00029517 as a biomarker and therapeutic target for lung cancer can be seen (Yang et al., 2018). We believe that the molecular mechanism by which TCON_00029517 is downregulated is the hypomethylation of its promoter region.

5.4 Materials and Methods

5.4.1 Bisulfite sequencing
The importance of DNA methylation creates an urgent demand for effective methods with highly sensitivity and reliability to explore innovative diagnostic and therapeutic strategies. Bisulfite genomic sequencing is recognized as a revolution in DNA methylation analysis based on conversion of genomic DNA by using sodium bisulfite. Bisulfite genomic sequencing is regarded as a gold-standard technology for detection of DNA methylation because it provides a qualitative, quantitative and efficient approach to identify 5-methylcytosine at single base-pair resolution. This method was first introduced by (Frommer et al., 1992) and it is based on the finding that the amination reactions of cytosine and 5-methylcytosine (5mC) proceed with very different consequences after the treatment of sodium bisulfite. In this regard, cytosines in single-stranded DNA will be converted into uracil residues and recognized as thymine in subsequent PCR amplification and sequencing, however, 5mCs are immune to this conversion and remain as cytosines allowing 5mCs to be distinguished from unmethylated cytosines.

Over the past few years, several techniques have arisen based on the working basis of bisulfite including Methylation Specific PCR (MSP), Combined Bisulfite Restriction Analysis (COBRA), Methylation-sensitive Single Nucleotide Primer Extension (Ms-SNuPE) and several other techniques depending on different applications (Li and Tollefsbol, 2011).

5.4.2 Mapping of Bisulfite data using Bismark

Aberrant methylation has been associated with a variety of diseases, including cancer. Current massively parallel sequencing methods to study DNA methylation include enrichment-based methods such as methylated DNA immunoprecipitation (MeDIP-Seq) (Staunstrup et al., 2016) or methylated DNA binding domain sequencing (MBD-Seq) (Zhao et al., 2014), as well as direct sequencing of sodium bisulfite-treated DNA (BS-Seq) (Chen et al., 2018). A combination of bisulfite treatment of DNA and high-throughput sequencing (BS-Seq) can capture a snapshot of a cell's epigenomic state by revealing its genome-wide cytosine methylation at single base resolution.
Bismark is a flexible tool for the time-efficient analysis of BS-Seq data which performs both read mapping and methylation calling in a single convenient step (Krueger and Andrews, 2011). Its output discriminates between cytosines in CpG, CHG and CHH context and enables to visualize and interpret their methylation data soon after the sequencing run is completed. Bismark aims to find a unique alignment by running four alignment processes simultaneously. First, bisulfite reads are transformed into a C-to-T and G-to-A version (equivalent to a C-to-T conversion on the reverse strand). Then, each of them is aligned to equivalently pre-converted forms of the reference genome using four parallel instances of the short read aligner Bowtie. This read mapping enables Bismark to uniquely determine the strand origin of a bisulfite read. Consequently, Bismark can handle BS-Seq data from both directional and non-directional libraries. Since residual cytosines in the sequencing read are converted \textit{in silico} into a fully bisulfite-converted form before the alignment takes place, mapping performed in this manner handles partial methylation accurately and in an unbiased manner. In addition to the alignment process, Bismark determines the methylation state of each cytosine position in the read. DNA methylation in mammals is thought to occur predominantly at CpG dinucleotides; however, a certain amount of non-CpG methylation has been shown in embryonic stem cells. In plants, methylation is quite common in both the symmetric CpG or CHG, and asymmetric CHH context (whereby H can be either A, T or C). To enable methylation analysis in different sequence contexts and/or model organisms, methylation calls in Bismark take the surrounding sequence context into consideration and discriminate between cytosines in CpG, CHG and CHH context.

\subsection{Datasets used}

We used Bisulfite-seq dataset from NCBI public repository with accession ID SRP016558 (Lin et al., 2013). The dataset encompass of 7 primary prostate sample and 7 benign prostate tissue sample.

\subsection{Differential methylation analysis}
The raw reads were mapped onto hg19 reference genome using Bismark (Krueger and Andrews, 2011). The mapped file obtained after Bismark was given as an input to a R package called as methylKit (Akalin et al., 2012). Total number of bases that were hypermethylated or hypomethylated in prostate cancer were calculated. The coordinates of promoter region of coding genes were downloaded from Eukaryotic Promoter Database. The promoter region of noncoding RNA was not well defined so we considered 2000 base pair upstream of the transcription start site. We then looked for the CpG island where were differentially methylated in the promoter region of both coding and noncoding RNA. The coding and non-coding RNA whose CpG island where 75% more methylated in cancer samples versus normal were marked as hypermethylated. Whereas the coding and non-coding RNA who’s CpG island where 75% less methylated in cancer samples versus normal were marked as hypomethylated.

### 5.4.5 Primer design for MSP PCR

Primer design for PCR-based methylation analysis following bisulfite conversion of DNA is considerably more complex than primer design for regular PCR. The choice of the optimal primer set is critical to the performance and correct interpretation of the results. Most methodologies in methylation analysis utilize primers that theoretically amplify methylated and unmethylated templates at the same time. The proportional amplification of all templates is critical but difficult to achieve due to PCR bias favoring the amplification of the unmethylated template. In MSP, assessment of the methylation status of a given locus is determined by the CpG sites within the primer sequence. It is thus important to include several CpG sites towards the 3' end of the primers to ensure specific binding and subsequent amplification of only methylated variants of the template. Careful evaluation of MSP primers has to be performed prior to analyses to assure the specificity and exclude over interpretation of results. The primer designing for MSP requires certain standard guidelines to be followed: (i) A limited number of CpG dinucleotides (usually one) should be included in the primer sequence. (ii) The included CpGs should as far as possible from the 3’ end of the primers as otherwise the primers will be entirely selective for methylated templates and amplify only methylated sequence. (iii) The melting temperature (Tm) of the primer should be around 65°C in order to run the PCR at or near to 60°C. Running the PCR in this temperature range is important to ensure the specificity of the PCR reaction. (vi) The inclusion of one or more Ts originating from a non-CpG C at, or near the 3’ end of each primer is desirable to ensure amplification of only bisulfite modified DNA (Davidović et al., 2014). Keeping above parameters in mind, MSP PCR primers for nLncRNA were made and MSP PCR was done using bisulfite converted PC3 gDNA as a template.

### 5.4.6 Methylation specific PCR
Precise mapping of DNA methylation patterns in CpG islands has become essential for understanding diverse biological processes. MSP (methylation-specific PCR) (Herman et al., 1996), rapidly assess the methylation status of virtually any group of CpG sites within a CpG island, independent of the use of methylation-sensitive restriction enzymes. The assay entails initial modification of DNA by sodium bisulfite, converting all unmethylated, but not methylated, cytosines to uracil, and subsequent amplification with primers specific for methylated versus unmethylated DNA. MSP requires small quantities of DNA, is sensitive to 0.1% methylated alleles of a given CpG island locus, and can be performed on DNA extracted from paraffin-embedded samples. MSP eliminates the false positive results inherent to previous PCR-based approaches which relied on differential restriction enzyme cleavage to distinguish methylated from unmethylated DNA. Primers are designed to distinguish methylated from unmethylated DNA in bisulfite-modified DNA, taking advantage of the sequence differences resulting from bisulfite modification. Unmodified DNA or DNA incompletely reacted with bisulfite can also be distinguished, since marked sequence differences exist between these DNAs. The frequency of CpG sites in CpG islands renders this technique uniquely useful and extremely sensitive for such regions. We isolated gDNA from prostate cancer cell line (PC3) and subjected it to Bisulfite conversion using EZ DNA Methylation-Gold Kit (Zymo Research). Manufacture recommended protocol was used. Bisulfite converted DNA was purified using column and precipitated using 100% ethanol. PCR was carried using for selected nlnRNA using the MSP specific primers designed using parameters described in above section.