

Exploring the Epigenome and Transcriptome to Identify Biomarkers in Ovarian Cancer

Maegha Tipirneni¹ and Dr. Fatima Merchant²

¹Friendswood High School, Friendswood, TX, USA

²Department of Biomedical Engineering, University of Houston, Houston, TX, USA

ABSTRACT

DNA methylation plays a vital role in cancer by regulating gene transcription. Alteration in DNA methylation among some genes has been associated with various tumors and this change in methylation signature among various genes can be potentially used as a biomarker for cancer detection. In the current work the Cancer Genome Atlas (TCGA) database was assessed for alterations in DNA methylation of 1816 genes in 600 ovarian cancer patients and 8 healthy subjects. Clustering of methylation patterns showed 5 major patient clusters. Alteration in methylation and gene expression of known biomarkers BRCA1, IQGAP1, BLCAP, MSH2, HSF1 along with novel biomarkers hSulf1 and HMSH2 was assessed in the ovarian cancer and normal subjects. The methylation levels were differentially expressed compared to healthy subjects. Compared to normal subjects, in ovarian cancer patients hSulf1, BRCA1, HSF1, BLCAP were hypomethylated while genes IQGAP1 and MSH2 were hypermethylated. The methylation levels for the specific genes considered biomarkers above were correlated with their mRNA expression level to see if there is an association between mRNA expression and methylation levels. Genes hSulf1, BRCA1, HSF1, IQGAP1 were negatively correlated with their methylation levels while MSH2 and BLCAP were positively correlated in ovarian cancer patients. Our study showed specific gene methylation clusters among the ovarian cancer patients suggesting this could potentially be used as a signature. In Cancer patients, methylation was significantly altered, and this alteration was correlated with its gene transcription.

Introduction

Every year around the world more than one hundred thousand women die of ovarian cancer. Ovarian cancer is the fifth leading cause of cancer death among women in the U.S, with 22,530 new cases and 13,980 deaths predicted for. Most deaths are of patients presenting with advanced stage, high grade serous ovarian cancer (~70%).

DNA methylation is a process where methyl-groups are added to a nucleotide, usually a cytosine followed by a guanine called a CpG site. DNA methylation is being extensively studied as a mechanism in tumor development and several studies have shown that differentially methylated genes are potential cancer driver genes. It has important implications in diagnostic, prognostic, screening and in predicting chemotherapy resistance in ovarian cancer.

Sulfatase 1 (hsulf 1) a gene encoding endosulfatase is known to regulate important growth factors like basic fibroblast growth factor (bFGF) in humans. This study seeks to evaluate the epigenetic regulation of hSulf1 and HMSH2 in ovarian cancer. Methylation and gene expression status of hsulf1 and HMSH2 in ovarian tumor datasets will be compared to normal data sets. Understanding the methylation status of genes has the potential for use as a biomarker for cancer. Therefore, it is critical to identify methylated genes that are correlated across human tumors that are used in research and drug discovery.

Problem

The purpose of this study is to determine methylation expression and incidence across the genome as well as for targeted markers in publicly available human ovarian cancer database to evaluate if methylation can be used as a robust marker for identifying and classifying cancer and to evaluate if mRNA (gene) expression correlates with methylation changes in various ovarian cancer patients. The question that this study seeks to address is: Do the epigenetic regulation and gene expression of recently identified markers hSulf1 and HSMH2 differ from ovarian cancer to normal samples?

Hypothesis

If the tumor and normal datasets for ovarian cancer from The Cancer Genome Atlas (TCGA) are compared, then the genes hSulf-1 and HSMH2 will show differential methylation and gene expression across tumors and normal datasets.

Methods

The methylation expression of 600 ovarian cancer patients from the National Cancer Institute's (NCI), The Cancer Genome Atlas (TCGA) was assessed. MethylMix was used to evaluate methylation changes. MethylMix 2.15.1 is an R Platform 3.5 bioinformatics tool that facilitates the analysis of methylation-driven genes in pan-cancer studies from TCGA. MethylMix identifies DNA methylation driven genes by modeling DNA methylation data in cancer versus normal and identifies homogeneous subpopulations. In addition, matched gene expression data was used to identify transcriptionally predictive DNA methylation events.[5][6] Table 1 shows the methylation and RNA seq data accessed from TCGA. Table 2 The table below shows demographic and clinical characteristics of the patients. Methylation data and RNA-seq data was assessed for 600 patients but demographic data was limited to 488 patients.

Table 1. The Cancer Genome Atlas (TCGA).

<https://portal.gdc.cancer.gov/projects/TCGA-OV> (Assessed December 2019).

TCGA cancer code	Number of cancer samples	Number of normal samples	Number of DNA Methylation samples	Number of DNA Methylation available sample data	Number of RNA sequencing samples
OV	608	8	600	582	541

Table 2. Demographic and Clinical Characteristics of Patients from TCGA GDC Portal.

Characteristic	Number of patients (n=488)
Age	
Mean, years (STD)	60.2(11.4)
range	27-87
Tumor stage #	
II	24(5%)
III	381(79%)
IV	79(16%)
Tumor grade	
2	57(12%)
3	419(88%)
Histology	
Serous	487(99%)
Unknown	1(1%)
Surgical Outcome	
Optimal(<1cm)	313(64%)
Suboptimal(>1cm)	119(24%)
Unknown	56(12%)
Platinum Status	
Sensitive	197(40%)
Resistant	90(18%)
Unknown	201(42%)
Recurrent Disease	
No	137(28%)
Yes	349(71%)
Unknown	2(1%)
Vital Status	
Alive	215(44%)
Dead	268(55%)
Unknown	5(1%)
Median PFS,Months(+SE)	16.8(0.8)
Median OS,Months(+SE)	43.6(2.2)
Adjuvant chemotherapy regimen	
Single agent platinum	12(2%)
Platinum/Taxane doublet	325(67%)
Other Platinum doublet	13(3%)
Platinum /Taxane triplet	72(15%)
Unknown	66(13%)

Procedure

1. Install R framework on intel-4 core processor
2. Install MethylMix on the R. Download all the related packages specified in the Package imports.
3. Download DNA methylation datasets and gene expression data from TCGA via MethylMix with the GetData function.

4. Run MethylMix functions on the downloaded human ovarian datasets the methylation, gene expression, and sample annotation data.
 - a. Run function ClusterProbes
 - b. Run function Preprocess DNAmethylation
 - c. Run function Preprocess GeneExpression
 - d. Run function MethylMix

At this stage, the epigenetic state i.e., hyper and hypomethylation state of this gene and if it causes up or downregulation will be known.

5. Use the MethylMixResults function to create the result matrix.
6. Make hyper and hypomethylation plots for the genes using the MethylMix PlotModel function: hsul1-1, HMSH2, BRCA1, RASSF1a, OPCML, P16INK4a, KLF11, ARH1, GBGT1, PDLIM2, HOXA9, RUNX3, TFP12, HOXA11, IQGAP1, PRAME. These are some of the other genes known to play a pivotal role and considered biomarkers for ovarian cancer.
7. Check the inverse correlation plots of the Gene expression data Vs. the DNA methylation for the genes listed in Step 6.

Statistical Analysis

Data normalization for understanding hyper and hypo methylation was done by taking differential methylation of each gene and dividing it by the mean of the normal samples. Normalization allowed for the visualization of the methylation genes on a scatter plot on a scale between -1 to +1. The mean methylation levels in cancer patients were evaluated and compared with normal controls. Occurrence data was plotted as a histogram to visualize methylation changes between cancer patients vs normal controls. One way ANOVA was used to evaluate the statistical significance of the methylation values between ovarian cancer patients vs normal controls. P value of 0.05 was considered statistically significant. Normality of the methylation changes of specific genes of interest was evaluated using Shapiro Wilks test. Pearson correlation was used to quantify correlation between gene expression (mRNA) and methylation changes.

Results

Profiling the methylation changes in ovarian cancer patients

MethylMix, a univariate mixture model developed in R, an open-source platform was used to quantify methylation changes in each CPG site associated with its nearest gene. The methylation data of ~600 patients was accessed from cancer genome atlas project. 1816 genes were also found to be differentially methylated. Differentially methylation was quantified as change in cancer methylation state versus normal state of the gene. Figure 1 shows the methylation changes of 1816 genes. The data was normalized between -1 to +1. The Genes in red that are between 0 to +1 state represent hyper methylation and genes with methylation normalized to -1 to 0 represent hypomethylation state of the genes.

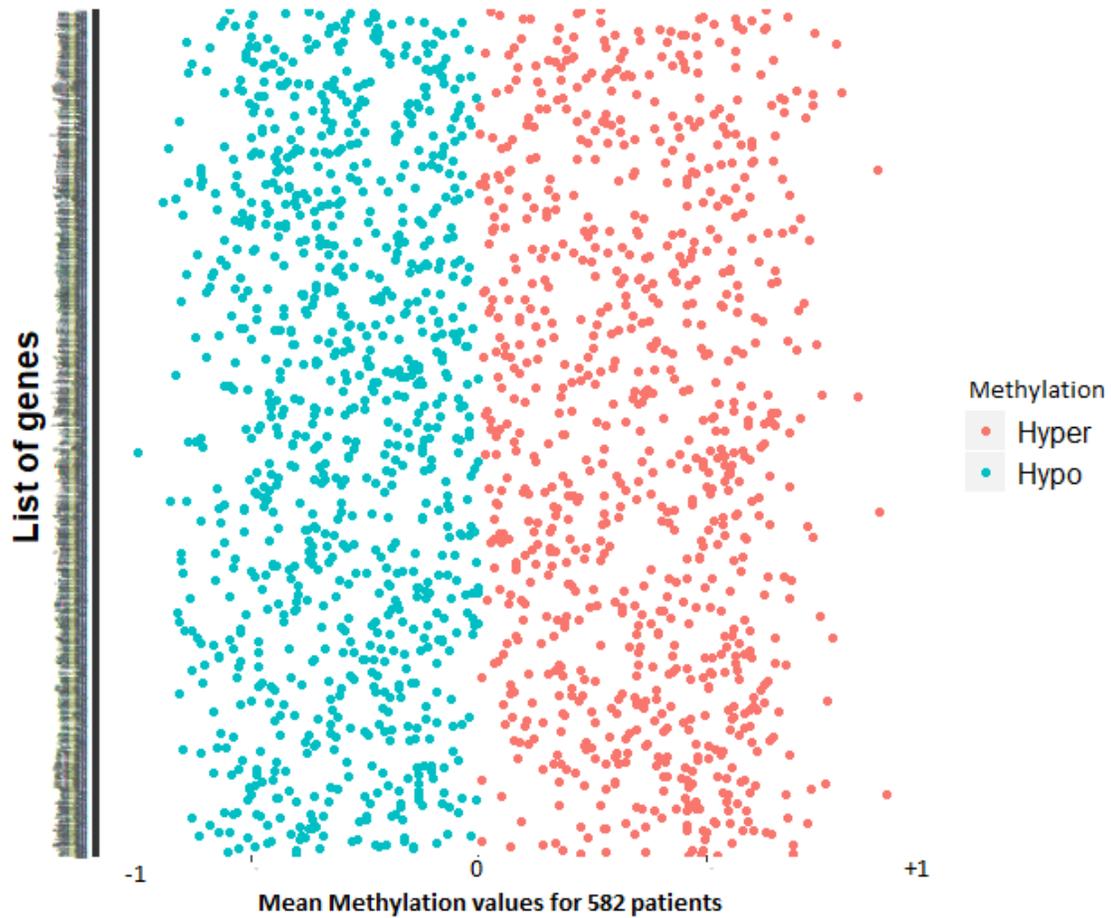


Figure 1. Scatter Plot of All Genes- Mean Methylation Values for 582 patients. The scatterplot shows the methylation values of 1816 genes among 582 patients with ovarian cancer. The scaled mean methylation values range from 0 to +/- 1 Red /blue indicating hyper and hypomethylation as shown.

Consensus clustering to identify clusters in ovarian cancer patients

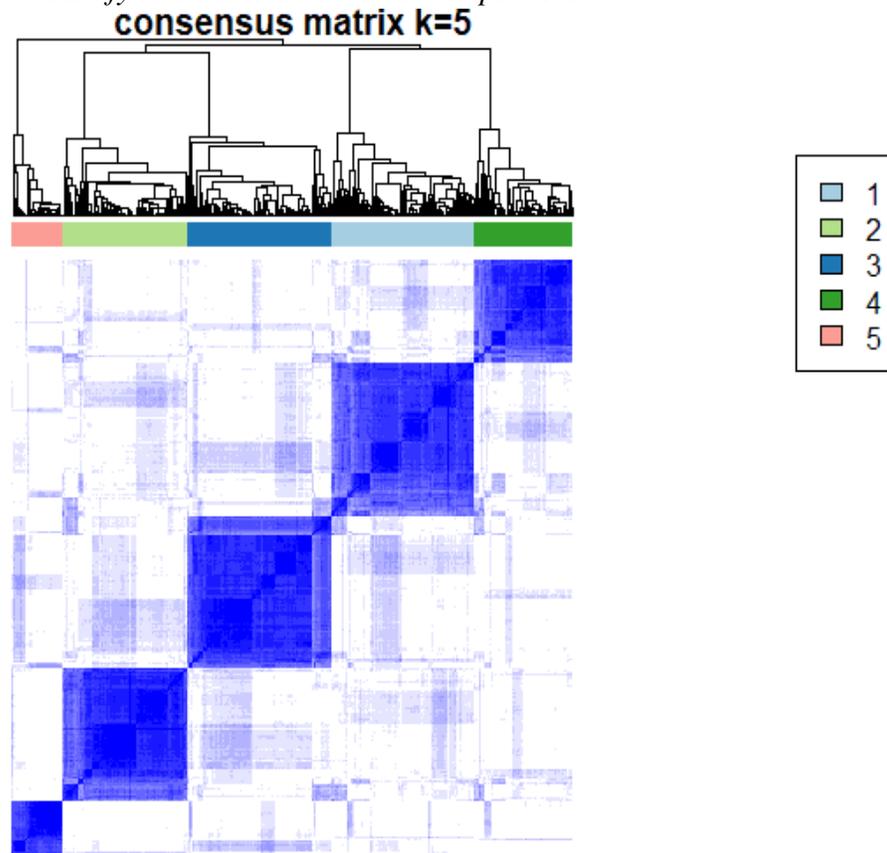


Figure 2. Clustering Chart- The clustering of various methylated patterns for each gene showed 5 major patient data clusters or groups.

To identify the cancer subgroups among the patients, consensus clustering of the ovarian cancer patient data was performed based on the differentially methylated gene expression pattern. As shown in figure 2, clustering of various methylated patterns using consensus clustering showed 5 major patient data groups. The blue clusters are high consensus clusters while white represents low consensus clusters. The disease state or severity information for these ovarian cancer patients was not available. The 5 patient clusters could be representing various ovarian cancer severity.

Alteration in methylation levels of key genes in ovarian cancer patients

To understand if methylation state can be used as a biomarker to identify ovarian cancer in the state or severity of the disease, key genes or biomarkers for ovarian cancer were identified from literature and their methylation changes were quantified in the ovarian cancer patients and compared with normal controls. BRCA1, IQGAP1, BLCAP, and HSF1 were genes previously identified as biomarkers for ovarian cancer. Apart from these the methylation changes of hSULF1 and MSH2 two novel genes that were shown to potentially have a role in ovarian cancer metastasis were also investigated.

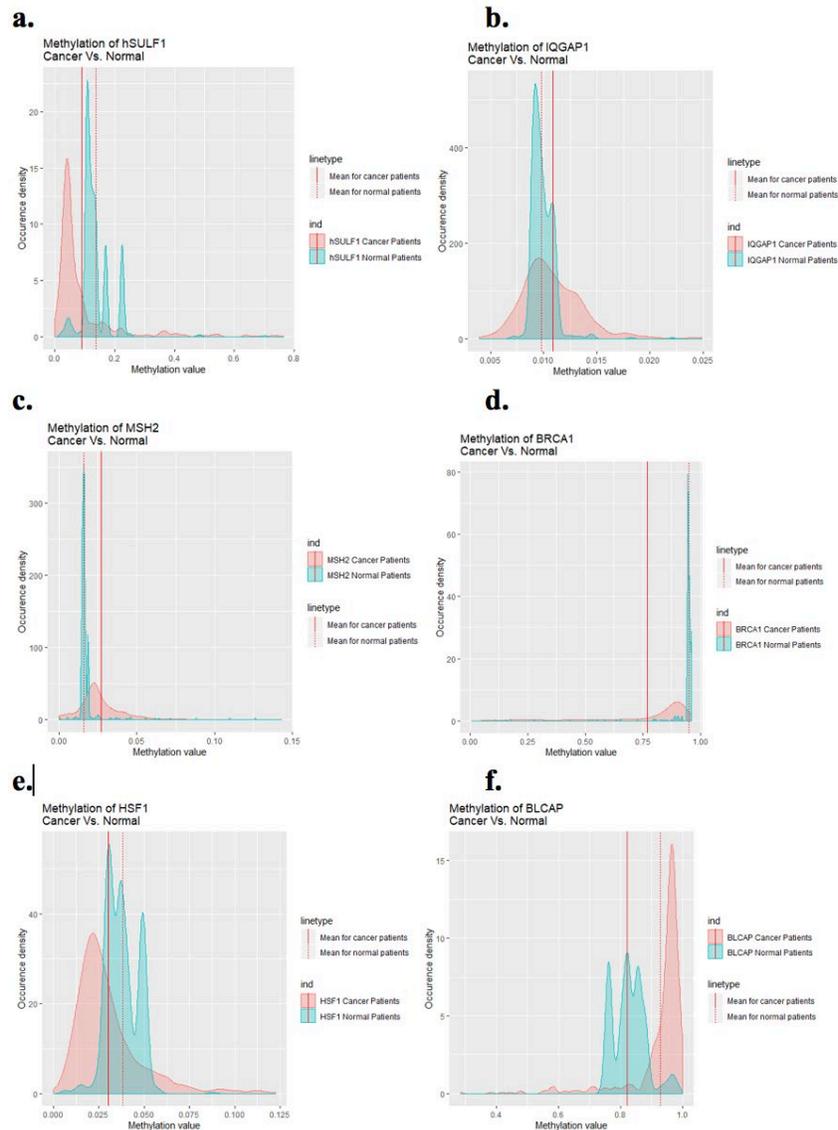


Figure 3A. Methylation Density Curve for the Gene hSulf1- This figure shows the occurrence density plotted against the methylation value of the gene hSulf1.

The cancer samples are plotted in red, and the normal samples are plotted in blue. The dotted line shows the mean for the normal samples and the straight line shows the mean for the cancer samples. The occurrence density is on a scale from 0 to 20 while the methylation value is scaled from 0 to 0.8. hSulf1, in ovarian cancer patients, was hypomethylated compared to normal controls.

Figure 3B. Methylation Density Curve for the Gene IQGAP1-This figure shows the occurrence density plotted against the methylation value of the gene IQGAP1.

The cancer samples are plotted in red and the normal samples are plotted in blue. The dotted line shows the mean for the normal samples and the straight line shows the mean for the cancer samples. The occurrence density is on a scale from 0 to 500 while the methylation value is scaled from 0 to 0.025. IQGAP1, in ovarian cancer patients, was hypermethylated compared to normal controls.

Figure 3C. Methylation Density Curve for the Gene MSH2- This figure shows the occurrence density plotted against the methylation value of the gene MSH2.

The cancer samples are plotted in red and the normal samples are plotted in blue. The dotted line shows the mean for the normal samples and the straight line shows the mean for the cancer samples. The occurrence density is on a scale from 0 to 400 while the methylation value is scaled from 0 to 0.15. MSH2, in ovarian cancer patients, was hypermethylated compared to normal controls.

Figure 3D. Methylation Density Curve for the Gene BRCA1- This figure shows the occurrence density plotted against the methylation value of the gene BRCA1.

The cancer samples are plotted in red and the normal samples are plotted in blue. The dotted line shows the mean for the normal samples and the straight line shows the mean for the cancer samples. The occurrence density is on a scale from 0 to 80 while the methylation value is scaled from 0 to 1. BRCA 1, in ovarian cancer patients, was hypomethylated compared to normal controls.

Figure 3E. Methylation Density Curve for the Gene HSF1- This figure shows the occurrence density plotted against the methylation value of the gene HSF1.

The cancer samples are plotted in red and the normal samples are plotted in blue. The dotted line shows the mean for the normal samples and the straight line shows the mean for the cancer samples. The occurrence density is on a scale from 0 to 60 while the methylation value is scaled from 0 to 0.125. HSF1, in ovarian cancer patients, was hypomethylated compared to normal controls.

Figure 3F. Methylation Density Curve for the Gene BLCAP- This figure shows the occurrence density plotted against the methylation value of the gene HSF1.

The cancer samples are plotted in red and the normal samples are plotted in blue. The dotted line shows the mean for the normal samples and the straight line shows the mean for the cancer samples. The occurrence density is on a scale from 0 to 15 while the methylation value is scaled from 0 to 1. BLCAP, in ovarian cancer patients, was hypomethylated compared to normal controls.

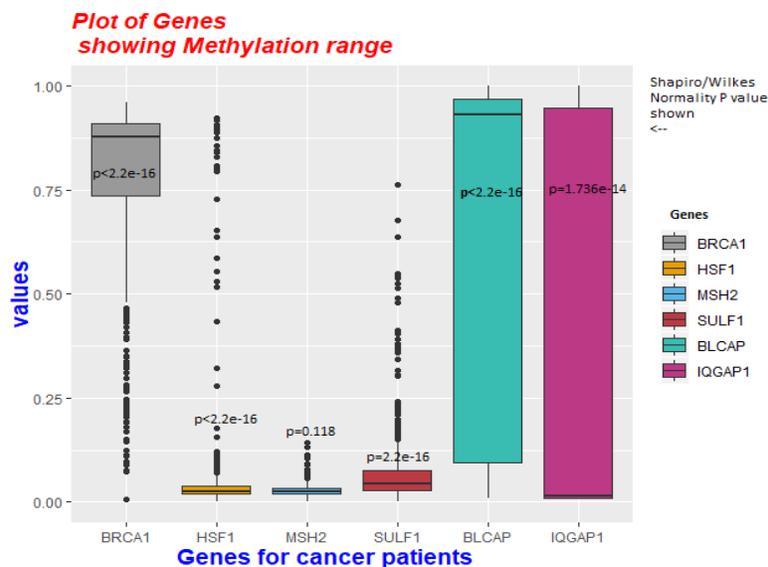


Figure 4. Methylation Density Curve for the Gene BLCAP- This figure shows how statistical significance of the change in methylation from cancer samples to normal samples of each gene.

The box plot was made using a Shapiro-Wilks test for normality in p-value Compared to normal subjects, in ovarian cancer patients hSulf1, BRCA1, HSF1, BLCAP were hypomethylated while genes IQGAP1 and MSH2 were hypermethylated as shown in figure 4.

Correlation of mRNA transcriptional response with methylation changes in ovarian cancer patients

In addition to the methylation data, the cancer genome atlas also has RNA seq data for the same tumor tissue. The RNA seq data was used to analyze the correlation between gene expression and methylation changes to see if methylation influenced the transcriptional response of the gene in the tumor tissue. As shown in the gene expression and methylation correlation graphs in Figure 5 A, B, C, D, E, HSF1 and IQGAP1 had strong inverse correlation while the genes BLCAP, BRCA1, and MSH2 had weak correlation. The gene hSulf1 had no statistically significant correlation as shown in figure 5A.

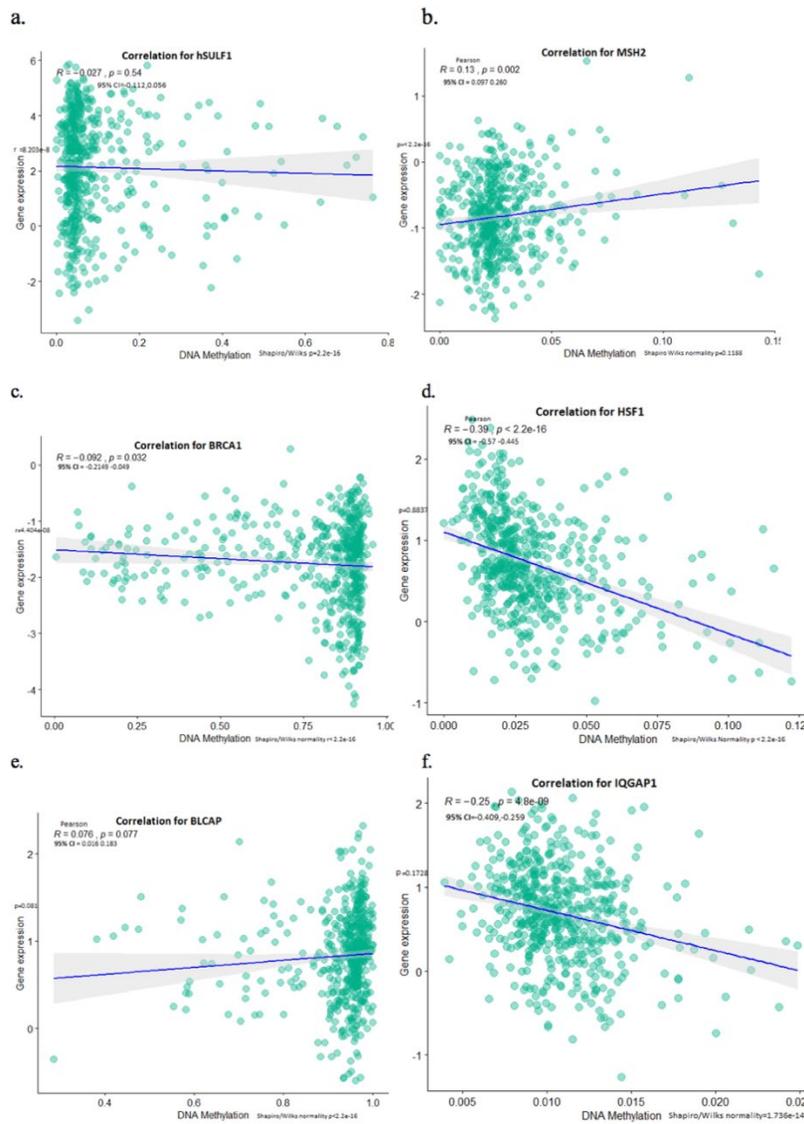


Figure 5A. Gene Expression and Methylation Correlation Graph for hSulf1-This figure shows the gene expression plotted against the DNA methylation for the gene hSulf1.

The gene expression is on a scale from 0 to 6 while the DNA methylation is on a scale from 0 to 0.8. A Pearson correlation test was performed to test the statistical significance of the correlation between the gene expression and DNA methylation. The coefficient is -0.027 indicating no statistically significant correlation.

Figure 5B. Gene Expression and Methylation Correlation Graph for MSH2- This figure shows the gene expression plotted against the DNA methylation for the gene MSH2.

The gene expression is on a scale from -2 to 1 while the DNA methylation is on a scale from 0 to 0.15 . A Pearson correlation test was performed to test the statistical significance of the correlation between the gene expression and DNA methylation. The coefficient is 0.13 indicating a weak statistically significant correlation.

Figure 5C. Gene Expression and Methylation Correlation Graph for BRCA1- This figure shows the gene expression plotted against the DNA methylation for the gene BRCA1.

The gene expression is on a scale from -4 to 0 while the DNA methylation is on a scale from 0 to 1 . A Pearson correlation test was performed to test the statistical significance of the correlation between the gene expression and DNA methylation. The coefficient is -0.092 indicating a weak statistically significant correlation.

Figure 5D. Gene Expression and Methylation Correlation Graph for HSF1- This figure shows the gene expression plotted against the DNA methylation for the gene HSF1.

The gene expression is on a scale from -1 to 2 while the DNA methylation is on a scale from 0 to 0.125 . A Pearson correlation test was performed to test the statistical significance of the correlation between the gene expression and DNA methylation. The coefficient is -0.39 indicating a strong statistically significant correlation.

Figure 5E. Gene Expression and Methylation Correlation Graph for BLCAP- This figure shows the gene expression plotted against the DNA methylation for the gene BLCAP.

The gene expression is on a scale from -1 to 2 while the DNA methylation is on a scale from 0 to 1 . A Pearson correlation test was performed to test the statistical significance of the correlation between the gene expression and DNA methylation. The coefficient is 0.076 indicating a weak statistically significant correlation.

Figure 5F. Gene Expression and Methylation Correlation Graph for IQGAP1- This figure shows the gene expression plotted against the DNA methylation for the gene IQGAP1.

The gene expression is on a scale from -1 to 2 while the DNA methylation is on a scale from 0 to 0.025 . A Pearson correlation test was performed to test the statistical significance of the correlation between the gene expression and DNA methylation. The coefficient is -0.25 indicating a strong statistically significant correlation.

Conclusion

The data analysis performed on the ovarian cancer tumor samples indicated that there was an approximately equal amount of hyper and hypo methylated genes among the 1816 genes screened in ovarian cancer patients. These gene methylation patterns, when clustered, showed primarily 5 different clusters or groups in the 600 ovarian cancer patients. In the ovarian cancer patients compared to the healthy subjects, BRCA1, IQGAP1, BLCAP, and HSF1 were found to be significantly differentially methylated. With regards to the two novel biomarkers hSulf1 and HMSH2, the hypothesis was partially supported as both genes were differentially methylated however, HMSH2 did not have a statistically significant p value in ovarian cancer patients compared to normal controls. The gene hSulf1 did not have very strong correlation with gene expression. The dysregulation in methylation patterns of cancer patients correlated with the gene expression levels (mRNA expression) for BRCA1, IQGAP1, BLCAP, and HSF1. The relationship i.e., direct, or inverse correlation between the methylation patterns and gene expression was dependent on the genes hyper or hypo methylation state. The correlation of methylation and mRNA expression suggests methylation has a

transcriptional control on key ovarian cancer biomarker genes. The consensus clustering data suggests that DNA methylation pattern among the patient genes could be used not only to detect ovarian cancer but also potentially determine the state or severity of the disease.

Future Applications

The hypermethylated gene HSF1 identified in this work may be a promising drug target and biomarker.^[7] This gene has been shown in several recent studies to regulate pathways related to cancer cell growth and survival. The hypermethylated gene IQGAP1 that has been identified in this study is a novel cancer antigen and currently a candidate that is being pursued for development of clinical diagnostics.^[8] Further studies would be needed starting with in vitro ovarian cancer cells to validate the dysregulation of methylation identified in the current study. Computationally further work using outlier detection and pathway analysis can help to further understand how these genes are signaling and modulating the cellular function.

Acknowledgements

I am thankful to Dr. Fatima Merchant of the University of Houston, Dept of Biomedical Engineering for helping me understand the concept of methylation and its role in cancer progression. I would like to thank her for her insight on this project.

References

1. Howlader N, Noone AM, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). SEER Cancer Statistics Review, 1975-2016, National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/csr/1975_2016/, based on November 2018 SEER data submission, posted to the SEER web site, April 2019.
2. Hentze, J.L., Høgdall, C.K., & Høgdall, E.V. (2019). Methylation and ovarian cancer: Can DNA methylation be of diagnostic use? (Review). *Molecular and Clinical Oncology*, (10), Retrieved September 14, 2019, 323-330. <https://doi.org/10.3892/mco.2019.1800>
3. Xu, G., Ji, W., Su, Y., Xu, Y., Yan, Y., Shen, S., ... Su, C. (2014). Sulfatase 1 (hSulf-1) reverses basic fibroblast growth factor-stimulated signaling and inhibits growth of hepatocellular carcinoma in animal model. *Oncotarget*, 5(13), 5029–5039. Retrieved October 26, 2019 doi:10.18632/oncotarget.2078
4. *American Journal of Epidemiology*, Volume 156, Issue 10, 15 November 2002, Pages 885–902, <https://doi.org/10.1093/aje/kwf139>
5. Pierre-Louis Cedoz, Marcos Prunello, Kevin Brennan, Olivier Gevaert, September 2018 MethylMix 2.0: an R package for identifying DNA methylation genes, *Bioinformatics*, (34), 17, 3044–3046, Retrieved September 20, 2019, <https://doi.org/10.1093/bioinformatics/bty156>
6. Gevaert, O., Tibshirani, R. & Plevritis, S.K. Pancancer analysis of DNA methylation-driven genes using MethylMix. *Genome Biol* 16, 17 (2015) doi:10.1186/s13059-014-0579-8

7. Sharma, C., & Seo, Y. H. (2018). Small Molecule Inhibitors of HSF1-Activated Pathways as Potential Next-Generation Anticancer Therapeutics. *Molecules (Basel, Switzerland)*, 23(11), 2757. doi:10.3390/molecules23112757
8. Rotoli, D., Morales, M., Maeso, M., García, M., Gutierrez, R., Valladares, F. ... Martín-Vasallo, P. (2017). Alterations in IQGAP1 expression and localization in colorectal carcinoma and liver metastases following oxaliplatin-based chemotherapy. *Oncology Letters*, 14, 2621-2628. <https://doi.org/10.3892/ol.2017.6525>
9. Reyes, H.D., Devor, E.J., Warriar, A. *et al.* Differential DNA methylation in high-grade serous ovarian cancer (HGSOC) is associated with tumor behavior. *Sci Rep* **9**, 17996 (2019) doi:10.1038/s41598-019-54401-w