NELSON, JONATHAN M., M.S. Identification of Drug Sensitive Gene Motifs Using "Epigenetic Profiles" Derived From Bioinformatics Databases. (2016)
Directed by Dr. Karen Katula 78 pp.

The use of epigenetic modifying drugs such as DNA methyltransferase inhibitors (DNMTi) and histone deacetylase inhibitors (HDACi) is becoming more common in the treatment of cancer. Currently, there is a profound interest in determining predictive biomarkers for patient response and the efficacy of known and novel drugs. There are likely distinct "epigenetic profiles" defined by the location and abundance of DNA methylation patterns and histone modifications. Here we propose to investigate the response of a selected subset of genes to particular DNMTi and HDACi treatments, in two human cancer cell lines, colorectal carcinoma HCT-116 and liver adenocarcinoma HepG2. In this study we identified unique epigenetic profiles based on microarray and bioinformatics derived epigenetic data that are predictive of the response to epigenetic drug treatment. Microarray studies were used to identify re-activated genes common in two different cancer cell types treated with epigenetic drugs. Bioinformatics data was compiled on these genes and correlated against re-expression to construct the genes' "epigenetic profile". We then verified the response of the select group of genes in HCT-116 and HepG2 upon treatment at varying concentrations of epigenetic drugs and illustrated that selective reactivation of the target gene. Additionally, two novel genes were introduced and one selectively activated over another.

Further research would prove invaluable for the medical and drug development communities, as a more extensive model would certainly be of use to determining patient response to drug treatment based on their individual epigenetic profile and leading to more successful novel drug design.

IDENTIFICATION OF DRUG SENSITIVE GENE MOTIFS USING “EPIGENETIC PROFILES” DERIVED FROM BIOINFORMATICS DATABASES

by

Jonathan M. Nelson

A Thesis Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Greensboro
2016

Approved by

_____
Committee Chair

To my amazing and intelligent son, Cassius and my wonderful and loving fiancée, Josie.

APPROVAL PAGE

This thesis written by Jonathan M. Nelson has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair _____

Committee Members _____

_____

_____
Date of Acceptance by Committee

_____
Date of Final Oral Examination

ACKNOWLEDGEMENTS

My deep gratitude first goes to Dr. Karen Katula who guided, counselled, and challenged me through my research and coursework. Having worked with her for the last 4 years, Dr. Katula has taught me not only how to perform science, but also how to ask the important questions that lead to further understanding. Her insights into my methods have polished my skills and move past hurdles that would have otherwise slowed me down.

This work would have been much more difficult had it not been for the knowledgeable support of Dr. Joel Parker and Sara Selitsky who aided in the computations necessary for the microarrays and elastic net regression.

I would also like to thank Dr. David Remington who provided me with both incredibly informative bioinformatics lectures which determined the direction of this thesis and who also extended a great deal of help when using the R statistical software. Additionally, I would like to extend my thanks to Dr. Amy Adamson who further refined my writing and grant writing techniques.

Furthermore, none of this would have been possible had it not been for the help of my family. To that end I extend my thanks to grandparents and parents for their support with helping to care for my son, to Josie Hug, my closest friend and future wife whose support ensured both sanity and joy, and to my son Cassius Nelson who always made sure that I did my homework.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

ABBREVIATIONS

ENCODE: encyclopedia of DNA elements

5-AZA: 5-azacytidine

SAHA: Vorinostat

DNMT: DNA methyltransferase

DNMTi: DNA methyltransferase inhibitor

HDAC: histone deacetylase

HDACi: histone deacetylase inhibitor

KMT: Lysine methyltransferase

TSS: transcription start site

PCA: principal component analysis

DMSO: dimethyl sulfoxide

FBS: fetal bovine serum

PCR: polymerase chain reaction

qPCR: quantitative real-time polymerase chain reaction

STRING: search tool for the retrieval of interacting genes/proteins

SAM: Significance analysis of microarray

FDR: False discovery rate

CHAPTER I

CANCER AND EPIGENETICS

Cancer is the second leading cause of death in the United States (Siegel et al., 2015). Each year cancer contributes to the death of approximately 8.2 million individuals worldwide accounting for 21.7% of deaths due to non-communicable diseases (WHO, 2014). The American Cancer Society estimates that almost 600,000 Americans will die in 2016 due to cancer while, simultaneously, there will be over 1.6 million new cases of cancer reported by the US alone. These same predictions suggest that roughly an equal number of cases will be discovered in both men and women, but females are projected to have a higher survival rate. From 2009 – 2012, incidence rates in women have stabilized and decreased by 3.1% in men (Siegel et al., 2016). While some studies suggest that the mortality rates of cancer are decreasing (Ryerson et al., 2016) others point out that these statistics are heavily skewed by early detection methods and do not necessarily represent realistic changes in survival rates (Cho et al., 2014). Even with our advances in technology and medicine, cancer is still a very real threat.

It was initially suspected that human cells would divide infinitely. However, early *in vitro* studies of human cell lines found that there is a finite number of times that a cell can replicate; this has become known as the Hayflick limit (Shay & Wright, 2000) . Under normal circumstances cells reach this point, somewhere between 40 and 60 divisions, and become senescent. Many different cellular mechanisms help to reinforce this limited replication. In cancer, cells manage to bypass the preventative measures intended to force a cell into senescence. This evasion of senescence, is one of the six

fundamental traits proposed as the hallmarks of cancer (Hanahan & Weinberg, 2011). In addition to enabling replicative immortality, the list of required steps for a cell to progress to cancer include: sustaining proliferative signaling, resisting cell death, evasion of growth suppressors, activation of invasion and metastasis, and inducing angiogenesis (Hanahan & Weinberg, 2011). Studies have shown that transforming growth factor-β inhibits proliferation while deletion of the receptor or perturbations to its signaling pathway can result in continuous cellular division (N. Cheng et al., 2012). Most cells that escape the Hayflick limit or manage to divide continually will eventually incur enough DNA damage causing the cell to undergo apoptosis. It is therefore necessary for a cancerous cell to develop mutations in apoptotic pathways (Adams & Cory, 2010). In order to be able to move away from its original location a cancer cell normally disrupts the E-cadherin pathway, which would normally keep it anchored to surrounding cells, thereby allowing it to metastasize into foreign areas (Berx & van Roy, 2009). As a cell metastasizes from one location to another, it will continue to grow. When a cell mass reaches sufficient size, it can no longer transport enough nutrients to the centralized cells and must begin recruiting new blood vessels. This aids not only in the ability to gather nutrients, but also to obtain oxygen and rid the cells of their metabolic waste products (Wicki & Christofori, 2008). Together, these traits result in a successful cancer cell, uninhibited by limitations of normal cells and capable of wreaking havoc on the body.

Normal cells don't acquire all of the hallmarks of cancer at a single time. Rather, a normal cell will move from its typical state to that of aggressive or invasive cancer progressively, sometimes not even successfully completing the transition. Between the normal cell and the far end of the cancer spectrum exist multiple stages that are defined

by the amount of cancerous characteristics that the cell possesses. Cancer progression and the appearance of the hallmarks of cancer involve changes in multiple genes that can broadly be classified as oncogenes and tumor suppressors. The oncogenes are normal cellular genes that are involved in cell proliferation. These include growth receptors (Yu et al., 2014), signaling pathway proteins (Liang et al., 2016), and transcription factors (Jacques et al., 2016). Oncogenes are gain of function mutations such that the protein is generally in an active state or overexpressed (D. Jiang et al., 2015). The Ras protein and its pathway components are classic examples of proto-oncogenes, known to be mutated to oncogenes. In contrast, tumor suppressors are genes that normally function in growth control and cell cycle checkpoint and are loss of function mutations in cancer cells (Peyser et al., 2015). P53, which responds to DNA damage, stops the cell cycle for either DNA repair or else initiates apoptosis, is a classic example of a tumor suppressor (Ferraiuolo etal., 2016). Other genes mutated during cancer progression are outside these two general categories and include DNA repair genes and genes involved in genomic stability (Chae et al., 2016; Pérez-alea et al., 2016).

The mutations in the oncogenes and tumor suppressors result from a combination of genetic mutations or epigenetic alterations, referred to as epimutations (Coppedè et al., 2014). Tumor suppressor genes, in particular, tend to undergo epimutations in cancer (Shakeri et al., 2016). The most extensively studied epimutation is DNA methylation, but others such as histone acetylation and methylation are also well characterized. Many of the genes that are misregulated in cancer are the joint effect of DNA methylation and histone modifications (Blackledge et al., 2010; Rose & Klose, 2014; Rothbart & Strahl, 2014).

DNA methylation normally occurs in CpG dinucleotide rich regions termed CpG islands (CGI). These islands possess a CG content of at least 50% and are, at minimum 200bp in length (N. Jiang et al., 2014); they are usually found overlapping the promoter region of a significant proportion of genes (60 – 70%), of which most remain unmethylated in non-tumorigenic cells (Lao & Grady, 2011). Aberrant DNA methylation causing inactivation of tumor suppressor genes is frequently seen in colorectal, hepatocellular, prostate, breast, lung, and renal cancer (Weisenberger, 2014). DNA methylation is the conversion of cytosine to 5-methylcytosine through the addition of a methyl group to the 5' position of the pyrimidine ring of cytosine. This covalent modification is catalyzed by DNA methyltransferase (DNMT), of which there are multiple forms; DNMT -3a / -3b which mediate *de novo* methylation, while DNMT1 maintains pre-existing methylation patterns with the help of DNMT3b (Esteller, 2002; Gravina et al., 2010; Rhee et al., 2002). Methylation of DNA inhibits numerous proteins such as transcription factors from binding, and can aid in the formation of a heterochromatic state, thereby reducing gene transcription (Herman & Baylin, 2003; Nan et al., 1998; Thomson et al., 2010). In general DNA methylation is associated with a reduction in the transcription of the gene body in which it resides.

Methylation pattern maintenance occurs during DNA replication by copying the methylation pattern present on the parent strand, signifying that it is a reversible process. Recent studies have shown that DNA methylation relies on particular histone modifications and in turn, histone modifications alter presence and density of DNA methylation (Blackledge et al., 2010; Tamaru & Selker, 2001; Wozniak & Strahl, 2014).

Histone modifications affect chromatin structure and therefore also effect transcription. Numerous histone modifications occur; of these, the most frequently

investigated are histone methylation and histone acetylation. The histone modification, Histone 3 lysine 27 acetylation, is an example of a histone modification that is associated with active promoters (Cai et al., 2015). Acetylated histones receive their moieties through histone acetyltransferases (HATs) and cofactor acetyl-CoA. Acetyl groups can be removed from the ε-amino groups of lysine residues by histone deacetylases (HDACs), causing a more compact chromatin state and interfering with transcription (Heerboth et al., 2014; B. Shi, 2013). Most histone methylation occurs similar to histone acetylation as lysine methyltransferases (KMTs) add a methyl group to the histone tail from S-adenosylmethionine (Drake et al., 2014). Two commonly studied KMTs are DOT1L (Angelica & Fong, 2008) and EZH2 (Agarwal et al., 2016). In many cases histone methylation is viewed as a repressive mark (Zhao et al., 2016) while others, such as mono and tri methylation of histone 3 lysine 4 are considered active marks (Cai et al., 2016; Rawat et al., 2016). Histone methylation is dynamic, being added and removed from histones by multiple enzymes. There are over 20 lysine demethylases (KDMs) that are capable of removing the methyl group from particular histone residues (Zahnow et al., 2016).

The effects of DNMTs on DNA methylation and HDACs on histone deacetylation can be interrupted or reversed upon cellular division through the use of DNA methyltransferase inhibitors (DNMTi) and histone deacetylase inhibitors (HDACi) (Heerboth et al., 2014; Lyko & Brown, 2005; Treppendahl, et al., 2014). DNMTi are a class of drugs that either directly block DNA methyltransferase activity by blocking the active site of DNMT1 or by acting as a target themselves, and locking hold to DNMT1, depleting all that is available. Such is the case with 5-azacytidine (5-aza). By either means, inactivation or depletion, DNMTi effectively reduces the amount of methylation

with each successive division. The decrease in methylated DNA results in an increase in previously silenced genes. There is conflicting evidence as to whether or not DNMTi have much of an effect on non-cancerous cells. Some studies show that DNMTi caused an increased incidence of tumors in murine models (though these doses were higher than used in clinical settings), whereas more recent studies suggest that due to a lower rate of metabolism and division, normal cells do not experience the same effect as seen in treated malignant cells (J. C. et al. Cheng, 2004; Stresemann & Lyko, 2008). Interestingly, some studies have suggested that treatment with DNMTi not only results in demethylation of hypermethylated regions, but can also restore methylation to hypomethylated regions as well (Heller et al., 2008). Currently, 5-Azacytidine is used in the treatment of myelodysplastic syndromes such as leukemia (Raj & Mufti, 2006), . Others have proposed the use of DNA methylation biomarkers for the use in the detection of colorectal cancers with highly methylated CpG islands (Laird & PE Lange, 2013). The effectiveness of such a method would suggest that these tumors would be sensitive to treatment with 5-Aza as some studies have already illustrated (Zouridis, 2012). Further investigations have performed genome-wide methylation screening before and after 5-Aza treatment in order to discover potential methylation markers to better determine which patients may benefit more from treatment with the DNMTi (Decock et al., 2012), while yet others have searched for these types of markers based on single tumor suppressor genes such as RECK (G. Shi etal., 2016).

HDACi are a much more complex group of enzymes than DNMTi, that restore the histone acetylation balance. These agents have been of extreme interest due to the fact that they cause the acetylation of both histones and non-histone proteins, such as p53 (Treppendahl et al., 2014). Though histone deacetylation is common even among

non-cancer cells, HDACi only affect 2-10% of expressed genes (Heerboth et al., 2014). Some HDACi such as Vorinostat are commonly used in the clinical setting due to their intrinsic nature to prevent deacetylation and their ability to resensitize radiation and chemoresistant cells (Coppedè et al., 2014; Heerboth et al., 2014; Lyko & Brown, 2005; B. Shi, 2013; Treppendahl et al., 2014).

New treatment strategies for cancers are now including the application of epigenetic drugs including DNA de-methylating agents, histone acetyltransferase inhibitors, and histone methyltransferase inhibitors (Khanahmadi et al., 2015; Miranda et al., 2009; Takeshima et al., 2014; Vasilatos et al., 2013). Just as with the gene mutations, each individual tumor is likely to have a particular "epigenetic profile" associated with genes that show epimutations. Thus, a patient's response to an epigenetic drug will depend not only on the genes that are silenced by epigenetic mutations, but also on the specific epigenetic profile of those genes. As there has been a move towards genotyping tumors (Kumarakulasinghe et al., 2015), it may prove highly informative to determine the epigenetic profiles for specific silenced genes in the tumor types, as these profiles may prove predictive for drug therapy. This epigenetic profile would consist of the DNA methylation and histone modifications up and downstream of the gene's start of transcription, and could be obtained using a variety of high throughput approaches currently available.

The main goal of this study was to determine if currently available epigenetic bioinformatics data are predictive for gene activation by the DNA methyltransferase inhibitor 5-Azacytidine (5-Aza). In a more limited study, we also tested the effects of histone deacetylase inhibitor SAHA. The major hypothesis is that activation of a gene by 5-Aza will have sites of DNA methylation in sequences surrounding its start of

transcription, as determined by the bioinformatics databases and, gene activation by the inhibitor SAHA will be associated with a lack of histone acetylation marks. To accomplish our goals the following specific aims were completed:

1. Examined bioinformatics data to identify re-activated genes common in two different cancer cell types, HCT-116 and HepG2, treated with the DNA methyltransferase inhibitor, 5-Azacytidine and correlated these changes to the "epigenetic profiles" of the genes derived from current bioinformatics data. (Chapter II)
2. Determined the selective reactivation of a select group of the common genes in HCT-116 and HepG2 with 5-Aza and correlated these changes with their epigenetic profiles. (Chapter III)
3. Predicted the reactivation of a novel gene, WNT5A promoters A and B, after treatment with 5-Aza and SAHA, based on epigenetic profiles. (Chapter IV)

The epigenetic profiles were defined within a 1000bp region, centered at the transcription start site, with 500bp upstream and 500bp downstream. Within these regions a profile was defined as the number and locations of methylation sites and the type of histone modifications. ENCODE (the Encyclopedia of DNA Elements) was the main source of epigenetic information. This database allows access to epigenetic data (DNA methylation and histone modifications) for many different cell lines, and is searchable to the gene and then sequence levels. The data have been gathered by institutes worldwide and is verified multiple times before release to the public.

CHAPTER II

CONSTRUCTION OF EPIGENETIC PROFILES

The goal of this study was to determine if it is possible to predict the re-expression of genes upon treatment with the DNA methyltransferase inhibitor 5-Aza, based on “epigenetic profiles” derived from bioinformatics databases. To meet that goal it was first necessary to select microarray studies that had used 5-Aza and then define the epigenetic profiles of a subset of the genes, followed by statistical analysis of these profiles to build a model of gene reactivation after 5-Aza treatment.

ArrayExpress (www.ebi.ac.uk/arrayexpress) and Gene Expression Omnibus, GEO (http://www.ncbi.nlm.nih.gov/geo/), databases serve as easily accessible sources for high-throughput experimental data including microarray based DNA, mRNA, and proteomic experiments (Barrett et al., 2013; Kolesnikov et al., 2015). The three main goals of GEO include data organization, easy submission of data while simultaneously ensuring quality, and user friendly query and analysis. Submissions to GEO can take two forms, either a GEO Series record, which is the original submission file, or GEO Datasets, which are a collection of biologically relevant and statistically comparable GEO Series files which have been collected and analyzed by GEO. Only GEO Datasets are available for further analysis using the database’s advanced tools for gene expression profiles and clustering. This database is easily queried and provides quick search results based on selection of key words such as a specific cell line and treatment type (Edgar et al., 2002). While GEO uses only data submitted to it directly, ArrayExpress

uses both direct submissions and data from GEO, therefore containing a greater likelihood of containing a combination of a specific cell line and treatment. With the use of statistical programs such as R, it is possible to cross analyze and organize the data from two or more given microarrays, resulting in any shared information between the two. While this reorganization is straightforward it does not result in useable information in regards to fold-change. In order to ensure that the levels of fold-change are comparable to one another and that their fold change is due to more than chance alone, it is necessary to apply a technique such as the significance analysis of microarray (SAM) procedure (Tusher et al., 2001). The SAM statistic is a variant of the t-statistic that includes a constant inflation factor in the denominator to overcome large statistics attributable to small fold changes and extremely low error estimates that arise when performing thousands of tests on cohorts of small sample size. SAM also uses a permutation strategy to produce an empirical false discovery rate (FDR), and thus is a model-free approach. This FDR is a score assigned by the model, where higher scores suggest that the fold-change is due more to chance than experimental conditions.

In 2003, the National Human Genome Research Institute, launched an initiative to form an international collaboration of various research groups that would work on building a collection and annotation of the functional elements of the human genome called the Encyclopedia of DNA Elements, or ENCODE. A series of multiple levels of data production and verification ensure high quality data (Myers et al., 2011). Individual experiments are performed at multiple institutions across the world and sent to the data center where it is validated, stored, and made accessible to the public. Each experiment searches for a single functional element, including DNA methylation and histone

modifications. These experiments are represented as tracks in the public database and are specific to cell type.

It is possible to visualize the ENCODE cell specific tracks at any given locus by either typing the various gene ID/symbol or chromosome number and region. There are multiple DNA methylation tracks, comprised of the 450K microarray track and reduced representation bisulfite sequencing (RRBS) track. Most cell lines have multiple replicates which are experimental repeats for the purposes of validation and accuracy. The 450K microarray is a wonderful tool to use when investigating a large number of samples due to speed and overall coverage. Due to technical limitations of designing 450K microarrays, they are only able to assay about 1.8% of CpGs and typical only cover those found in genes or CpG islands (Butcher & Beck, 2015). Reduced representation bisulfite sequencing has the ability to cover a larger percentage of the CpGs, being a sequencing method rather than a microarray. With a sufficient number of reads for both the 450K microarray and RRBS, the two tend to agree at extreme values, very high levels of methylation, and very low levels. However, in the upper and lower 20% values, the two are inconsistent with one another (Pan et al., 2012). It is because of these reasons that it can be beneficial to use both the 450K and RRBS tracks when investigating DNA methylation in ENCODE.

Additional epigenetic tracks on ENCODE include histone modifications. Some cell lines have been investigated in more depth than others and therefore have a larger selection of histone tracks available. Three of the available tracks for many cell lines are for the histone marks: H3K4me1, H3K4me3, and H3K27ac. All three of these marks are associated with active gene promoters (Aday et al., 2011; Creyghton et al., 2010;

Pekowska et al., 2011). By considering this information in addition to DNA methylation, it is possible to get a clearer view of the epigenetic profile of the locus.

Analyzing regions spanning 10,000 to 1,000bp around the TSS is fairly common when looking for epigenetic marks (Ucar et al., 2011; Wang et al., 2009). In a recent study it was found that a 1000bp window was large enough to identify correlations between histone modifications and DNA methylation (Linghu et al., 2013). As a gene's promoter is within only a few hundred base pairs upstream of the TSS, attempts should be to try and capture a region not directly affected by the typical adenine and thymine rich nature of the promoter to better investigate the DNA methylation without bias. By using a 1000bp window centered at the TSS, this type of situation can be avoided in many cases. Further breaking the 1000bp region into 50bp bins will allow for a more thorough inquiry into specific regions. When this window is used in conjunction with the epigenetic track from ENCODE it is possible to construct an epigenetic profile of the TSS of each gene under study.

By constructing epigenetic profiles of genes we hypothesize that it will be possible to predict their re-expression upon treatment with the DNA methyltransferase inhibitor, 5-Azacytidine, using a model generated through analysis of 5-Aza treated microarray data.

Materials and Methods

5-Azacytidine treated microarrays for HCT-116 and HepG2

Microarrays in the ArrayExpress database were searched using the conditions of cell line and 5-Aza treatment. Two microarrays were obtained, HCT-116 (Data available in ArrayExpress database (Li et al., 2014), under accession number E-GEOD-57341)

and HepG2 (Data available in ArrayExpress database (Dannenberg & Edenberg, 2006), under accession number E-GEOD-5230). The microarray for HCT-116 treated cells with 0.5μM 5-Aza for 72 hours, whereas the HepG2 microarray treated with 2.5μM for 96 hours.

Determination of microarray shared genes and fold change

The Significance Analysis of Microarray (SAM) procedure was used to test for differential expression and estimate fold change between experimental conditions. SAM was implemented in R version 3.1 using the package 'samr'.

Construction of "epigenetic profiles"

"Epigenetic profiles," defined as 1000bp regions centered at the transcription start site and consisting of DNA methylation and histone mark scores and abundance, were constructed in two different manners; initially by hand for a group of 32 selected genes and later by using a mixture of GALAXY, ENCODE, and Excel.

Hand construction of the epigenetic profiles was performed using ENCODE to explore the methylation and histone modification tracks of 32 genes selected from the microarray studies based on differential re-expression of any given gene between cell lines. The genes that were selected illustrated both extremes (low fold-change and high fold-change) and others that fell in between in an attempt to prevent any form of bias. Each track was examined in windows of 50 base pairs each starting at the transcription start site and extending 500 base pairs in each direction. Methylation data was collected for HCT-116 and HepG2 by analyzing two RRBS tracks for HCT-116 and two RRBS tracks for HepG2; these numbers were averaged together for the tracks of each cell line and recorded. Additionally, two 450K tracks for HCT-116 and one 450K track for HepG2 were collected and recorded. When more than one track was available the scores for

450K were averaged in the same manner as RRBS. When analyzing histone modifications, we selected those which were shared between the two cell lines: H3K4me1, H3K4me3, and H3K27ac tracks (all which are associated with active promoters). Two tracks were used for analysis of HCT-116 H3K4me3 (Peaks 1 & 2), while 3 were used to analyze the same modification in HepG2 (Peaks 1, 2, & ENCODE/Broad). In the event that a peak was detected that stretched outside of the 50 base pair window, the score was carried over into all 50bp windows that it was located in, so long as it accounted for at least 25bp of that window. No scores were given for anything further than 500bp up/downstream. If more than one peak was present in a single track set, the two scores were averaged if they were the same length, so long as both accounted for at least 25bp of the given window and they were the same length, if not the shorter band's score was taken since it is considered more accurate by ENCODEs standards. For all histone tracks and the 450K track, numbers ranged from 0 – 1000; the RRBS track ranged from 0 – 100. These numbers indicated the RRBS track indicated the percentage of reads that showed methylation at a given CpG, whereas the 450K track scored based on the microarrays beta values, multiplied by 1000. Histone tracks were scored based on the signal value, indicating the average enrichment for the region. When available, multiple tracks of a given epigenetic mark are used as a form of biological replicates.

Computed construction of the epigenetic profiles was performed using a combination of GALAXY, ENCODE, and Excel on any genes in the microarrays where either cell line was upregulated after treatment with 5-Aza. Using the gene symbol for each gene, information was pulled on GALAXY from ENCODE containing the gene ID, chromosome number, transcription start site, and strandedness. This information was

imported into Excel and cross referenced to expression for each cell line. The compiled information was then uploaded into GALAXY and flanking regions were obtained for 50bp in a single direction until 500bp regions upstream and downstream were accounted for. Hg19 Tracks for DNA methylation and histone modifications were imported from ENCODE into GALAXY. The tracks used in the computed bioinformatics were the same as those that were used in the hand construction with the exception that no RRBS tracks were used due to importing issues. All rules followed previously for scoring were maintained. The 50bp flanking regions for each gene were crossed with the track information from ENCODE in GALAXY. The results were imported into Excel for easy data sorting and a single master file containing the information of each gene, location, and epigenetic mark.

Predicted Protein Interactions

String version 10 was used to determine protein-protein interactions of 32 selected genes using summary network, actions display. (www.string-db.org)

Principal Component Analysis

Interactions between 10 components of the hand constructed method and separately, nine in the computed method were investigated through PCA. PCA used R, statistical analysis software to determine where the largest amount of variation in the dataset existed while also reducing dimensionality. Data for the analysis was pulled from the 32 selected genes and standardized so that standard deviation was 1. PCA was executed to analyze the relationships between these components.

Elastic Net Regression

Each figure (scores, location, and number) was correlated with fold change for HelpG2 and HCT116 using spearman and pearson statistical tests prior to Elastic Net

Regression, a penalized likelihood regression method that selects and estimates regression coefficients based on the correlation structure of the predictors and performs well when predictors are highly correlated. If a correlation of >0.1 or <-0.1 was found then Elastic Net regression was run using R version 3.1 with the package, 'glmnet 2.0-5'.

Results

We hypothesized that "epigenetic profiles" derived from scores of epigenetic marks obtained from bioinformatics databases can be used to generate a model that predict changes in expression due to 5-Azacytidine (5-Aza).

To test this hypothesis, we used gene expression data from microarray studies of 5-Aza treated cells and combine those results with epigenetic marks derived from bioinformatics databases, in particular ENCODE. We searched the online database of microarray studies on ArrayExpress. Our search was refined by looking for the cell lines, HCT-116 and HepG2. These cell lines were selected due to the large number of tracks available on ENCODE for these two cell lines. Two microarrays with similar concentrations and durations were found, E-GEOD-57341 and E-GEOD-5230. The microarray, E-GEOD-57341, was of HCT-116 cells that had been treated with 0.5μM 5-Aza for three days, whereas E-GEOD-5230 represented HepG2 cells treated with 2.5μM 5-Aza for 4 days.

The microarrays were cross-analyzed to determine which genes were shared between both, along with the relative fold-change of the genes after treatment. In all, 28,247 probes were found to be shared between the two microarrays. Of those probes there were 17,381 unique genes and 734 loci or genes that did not match a known gene name or kgAlias in the ENCODE database.

To generate the hand constructed epigenetic profile (Fig. 1), we selected 32 genes with varying levels of fold-change from a list of the microarrays in which both cell lines showed upregulation for a given gene. These genes were then analyzed for interaction on the protein-DNA and protein-protein levels, using String ver. 10, finding that only two interacted (Fig. 4). This interaction was on a protein-protein level and upon review it was determined that the effect of 5-Aza on the 32 genes did not involve interaction among these proteins. We did not perform this same analysis on the entirety of both microarrays due to limitations of the online database. All genes were then analyzed for DNA methylation and histone modifications using ENCODE hg19, using as many tracks as were available and shared between both HCT-116 and HepG2 cell lines. The tracks found in common included reduced representation bisulfite sequencing (RRBS) and 450K for methylation and H3K4me1, H3K4me3, and H3K27ac to investigate histone modifications. These represent active histone marks; no repressive histone mark tracks were available. Any replicate tracks available for the epigenetic modifications were included in the study. Epigenetic profiles were constructed by analyzing 500bp upstream and 500bp downstream of the TSS, each being broken apart into 50bp windows, and recording the value and location for each epigenetic mark. Upon completion of the hand constructed epigenetic profiles, PCA and elastic net regression statistical tests were performed. An example of tracks for the highly activated gene MYL9 and PPDPF, found in both HepG2 and HCT-116 is shown in Fig. 3.

Principal component analysis of the hand constructed epigenetic profiles for the selected 32 genes (Table 1) illustrate the correlations in our dataset. Using approximately a 0.4 correlation cut-off value, principal component 1 displays a positive correlation between the total number of histone modifications and H3k4me1/me3 and

H3K27ac histone marks, accounting for the largest amount of variation in the dataset. Principal component 2 shows that the second largest variation is seen in genes possessing a higher proportion of methylated CpGs than non-methylated, scored by RRBS, located slightly upstream of the TSS. The third principal component finds that fold-change is positively correlated with high values of methylation scores, whereas these two features hold a negative correlation with H3K27ac. In principal component 4 a negative correlation in higher levels of methylation was found between the RRBS track and 450K methylation scores, located further downstream. The remaining principal components possessed a value of less than 1 for the square root of eigenvalues, indicating that they could be due simply to chance. For that reason, they were not considered during analysis or generation of a model. As a whole this analysis found that most variation is seen in histone mark scores and abundance, whereas methylation found through the RRBS track is located slightly upstream of the TSS and further downstream. Additionally, PCA revealed that expression (as indicated by fold-change) seems to be most closely correlated with higher levels of methylation scores and a lack of H3K27ac.

Following PCA, Spearman and Pearson statistical tests were run on the hand constructed epigenetic profiles to verify a correlation of >0.1 or <-0.1 between fold change and features. These features consisted of any combination of either histone modifications or DNA methylation. Verifying that there was a correlation, elastic net regression of the hand constructed profiles was performed in R. The data, while nonsignificant, showed a trend at approximately 15 features. These features consisted of any combination of methylation or histone modifications. This indicates types of

methylation scores (RRBS or 450K), number of methylated CpGs, and individual histone scores.

To increase the power to detect significant features, the analysis of the microarray data was extended to the entire dataset. The entire list of 17,381 gene symbols previously filtered for the generation of the manual list genes was used to construct the computed epigenetic profiles (Fig. 2). The gene ID's for these genes were pulled from ENCODE using GALAXY. This resulted in a list of 43,461 transcripts of the genes. The list was imported into Excel and cross referenced with gene symbol, TSS, strandedness, and fold-change for both cell lines. The number of transcripts was reduced to 11,480 by filtering out duplications and those for which GALAXY and ENCODE did not pull sufficient data. The new list was imported back into GALAXY in order to analyze 50bp windows of DNA methylation and histone modification that had been imported from ENCODE. All tracks used in the process were identical to those that were used in the hand construction, with the exception of that we did not include the RRBS tracks as we were unable to obtain the desired information from them. The resulting files were transferred to Excel and arranged into a single master file.

We ran principal component analysis on the computed epigenetic profile list of the 11,480 transcripts and their features (Table 2). During our analysis we applied a threshold of approximately 0.40 for the cut-off value when determining correlation. We found that PC1 is a measure of the total number of modified histones and that it strongly correlates in a positive manner with the scores of H3K4me1 and H3K27ac. Principal component 2 shows a positive correlation of 450K scores and assayed CpGs, located very near the TSS. PC3 was primarily a measurement of fold-change and correlated positively with an increase in methylation downstream of the transcription start site.

From this point correlations become biologically irrelevant as they possess square root eigenvalues that indicate that the correlations are likely due to chance alone.

Spearman and Pearson statistical tests were run on the computed epigenetic profiles. Both tests resulted in a lack of correlation of >0.1 or <-0.1 between expression and features. We then attempted to eliminate everything except the extreme values and found no correlation. In another attempt to find a correlation, we meaned and summed the entire region, still finding none, therefore, elastic net regression analysis was not performed.

The results of the computed epigenetic profile's principal component analysis are likely found to have little correlation due to the fact that it included multiple transcripts of each gene symbol as it was not feasible to identify which transcript for each was probed by the microarrays for so many thousands of genes. This differs from the hand constructed profile where it was easier to identify the transcript used, and where many of them had a single isoform. Additionally, the fact that the RRBS track was not able to be included in the computed epigenetic profiles likely contributed to the lack of correlation and significance.

Discussion

We had hoped to obtain microarray data from ArrayExpress for HCT-116 and HepG2 with treatment conditions that we more closely matched one another. This was not possible. This means that the data may be skewed toward a higher level of relative fold-change in HepG2 for any given gene or transcript. This is likely due to both the concentration of 5-Aza, which was 5 times more in HepG2, and the additional 24 hours of treatment compared to HCT-116. While this difference may have shifted the data

slightly, it still allowed us to investigate the features present in genes that were reactivated. For future studies, it would be beneficial to perform a microarray experiment prior to bioinformatics analysis in order to have consistency between treatments.

In our study we performed both a manual hand construction of the epigenetic profile and a computed epigenetic profile. Of these two, the computed epigenetic profile was able to collect a far greater amount of data in less time with less chance of recording the scores improperly, but also lost a large number of genes. These genes were lost due to difficulties in cross platform analysis. Using a custom microarray with clearly defined genes and specific transcripts, or alternatively, RNA-seq, could help to prevent the loss of information. Furthermore, GALAXY seemed to have difficulties pulling the pertinent information from the RRBS track. It may be possible in future studies to create a custom track of the RRBS reads, refining the information which it provides in order to obtain what is needed.

It appears that the hand construction of the epigenetic profiles was better because of the ability to record scores such as those of the RRBS track. Furthermore, we were able to select genes that we saw did not affect the transcription of one another, by running a protein-DNA and protein-protein interaction analysis. Using the manual method also allowed the opportunity to investigate the specific isoform that was analyzed through the microarrays and used a single transcript for each gene symbol as compared to the computed method, in which many cases used three or more transcripts per gene symbol. The issue with multiple transcripts per gene is that it creates multiple epigenetic profiles which differ in methylation and histone location and scores, and creating an inaccurate representation of correlations to fold-change.

PCA analysis of the hand constructed epigenetic profiles of the 32 selected genes and the computed epigenetic profiles both illustrated that the component with the largest variation was that of the total number of histones, meaning that the genes differed greatly in both their overall histones and which histone marks made these scores. Differences between the microarrays are likely due to RRBS scores not being included in the computed epigenetic profile PCA because of lack of availability. We suspect that the lack of this track greatly impacted the analysis. One of the first and most direct ways that lack of RRBS track score affected the data was by reducing the number of assayed CpGs. This is due to the fact that RRBS tends to more successfully analyze CpGs near the TSS. Indirectly, it can be seen that in the manually constructed principal component analysis, RRBS forms multiple correlations with histone marks. More importantly RRBS has a strong positive correlation with fold change, seen in PC3.

The PCA results showed variation in the level of methylation just prior to the transcription start site; this makes sense as it is commonly seen in genes with lower levels of expression. Previous studies have found that methylation in the promoter region can affect transcription factor binding sites thereby reducing gene transcription (Medvedeva et al., 2014); this likely accounts for a portion of the genes that were expressed at lower levels prior to treatment with 5-Aza. Other studies have found that the 1000bp window that we investigated contained as little as 1-2% to as much as 10%-13% CpG content (Couldrey et al., 2014); providing a varying density of CpGs with the potential of methylation, contributing to further variability that was picked up by the PCA analysis. The fact that H3K27ac was found to be negatively correlated with DNA methylation has been studied before find that tissue specific enhancer regions (such as those marked with H3K27ac) are typically unmethylated (Blattler et al., 2014; J. Xu et al.,

2009). This is likely to ensure expression of the particular gene. The correlation seen in the hand constructed PCA between fold-change and DNA methylation would be expected because of the activity that 5-Aza possesses. By preventing activity of DNMT1, 5-Aza prevents DNA methylation from being replicated after cell division. This reduction in DNA methylation leads to higher levels of gene transcription which we calculated as fold-change.

The elastic net regression of the manually generated epigenetic profiles did not contain enough samples in order to possess any statistical significance, however, a trend was discernable. This trend can be seen as the dip in P-value around 15 features (Fig. 5). These features represent any combination of epigenetic marks whether they be DNA methylation, histone modifications, or both in addition to their numbers. It stands to reason that there must be DNA methylation present in order for any DNA methyltransferase inhibitor to elicit an effect. Moreover, extremely high levels of DNA methylation within a particular gene would likely reduce the efficacy of the drug. The histone marks, which we analyzed are all considered activating marks are found, primarily near transcription start sites. It is therefore likely that these 15 features consisted of DNA methylation and sparse activating histone marks located just upstream of the TSS and stretching midway downstream. An example of this is MYL9 and PPDPF (Fig. 3). MYL9 possessed a relative fold-change of 28.59 in HepG2 and only 1.13 in HCT-116. HepG2 follows the trend found in the elastic net regression with only 19 sites of methylation and 3 histone marks. In HCT-116 it has no methylation and only two histone marks. The gene PPDPF had a fold-change of 1.75 in HepG2 and 1.11 in HCT-116. PPDPF had a total of 35 features in HepG2 while it contained only two in HCT-116.

Taking similarities and differences into account from both principal component analyses along with the trend seen in the elastic net regression, we were able to formulate a general model that represents how responsive a gene is to treatment with the DNA methyltransferase inhibitor 5-Azacytidine. *The gene must possess moderate levels (10-20 sites) of DNA methylation between 250bp upstream and 250bp downstream the start of transcription. Greater distance than those proposed will reduce efficacy of the drug. The methylated sites should be spread out throughout the region rather than clustered, for the most robust effect. It must also contain one to two activating histone marks, preferably either H3K4me3.*

Our methodology differs from that of others in several ways, namely, we used bioinformatics means to obtain both the expression profiles after drug treatment and then focused on the promoter regions of the genes that were upregulated. Other studies have selected groups of tumor suppressors to focus on, analyzing only their DNA methylation, not histone modifications and used little to no bioinformatics data (Follo et al., 2009; Voso et al., 2011). One study like this selected 32 promoter-associated CpG islands, finding them to possess a wide range of methylation in NCI-60 cell lines and using that information to predict drug response (Shen et al., 2007), while another was able to use DNA methylation profiles of differentially methylated regions of patients to predict response to a DNMTi, decitabine (Meldi et al., 2015).

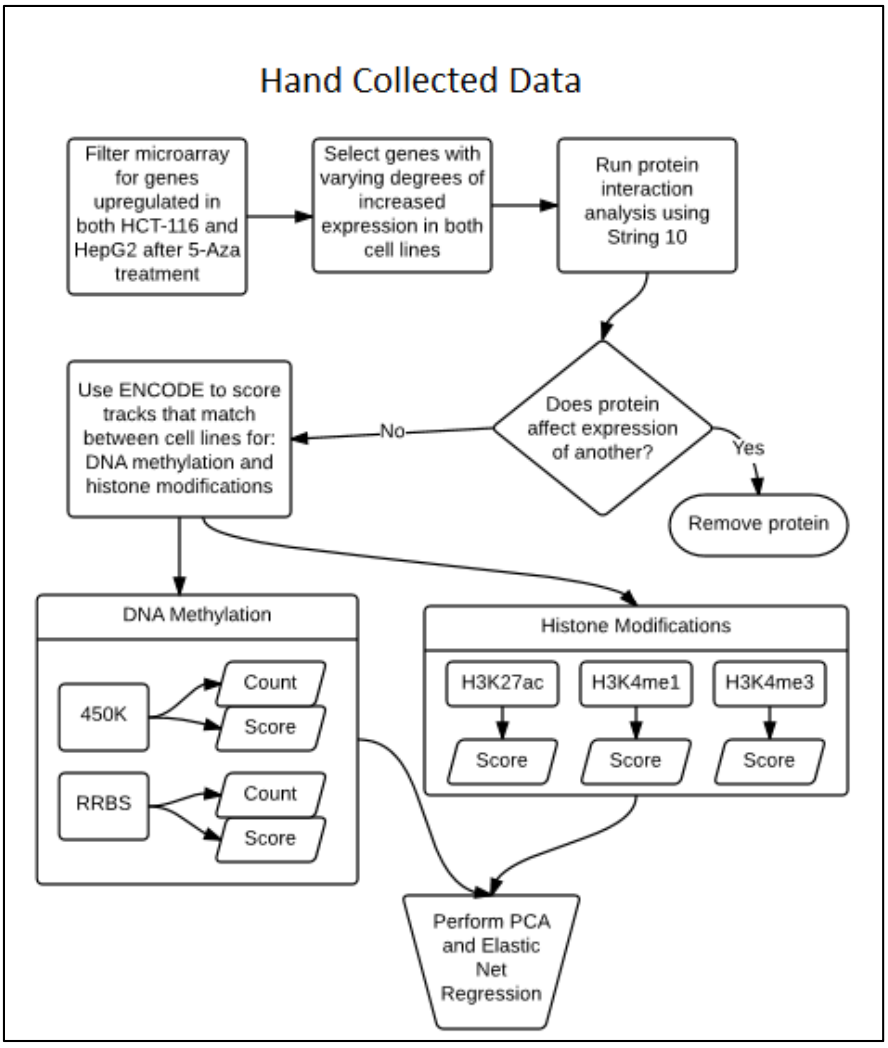


Figure 1. Hand Collected Flow Chart. Flowchart detailing the generation of the epigenetic profiles and statistical analysis for the hand selected genes from the HCT-116 and HepG2 5-Aza microarrays.

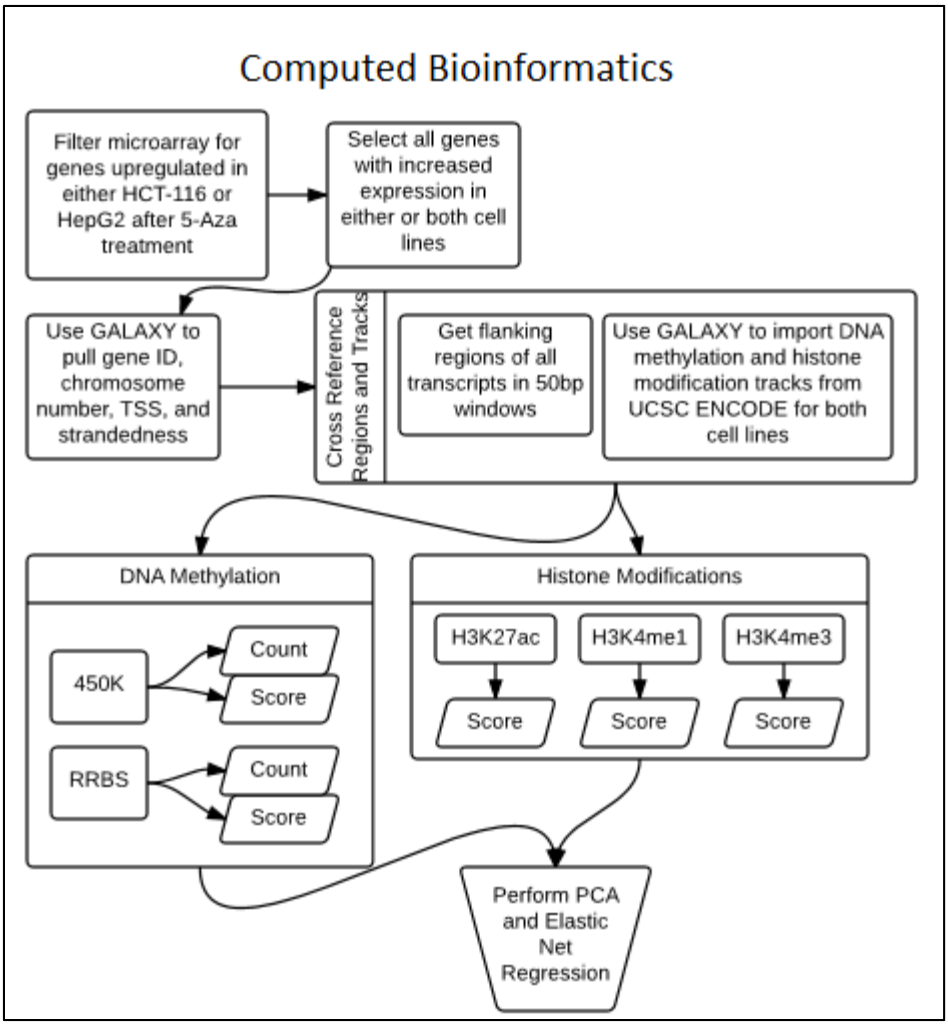


Figure 2. Computed Bioinformatics Data Flow Chart. Flowchart detailing the generation of the epigenetic profiles and statistical analysis for the HepG2 and HCT-116 microarrays.

Figure 3. ENCODE Snapshot of MYL9 and PPDPF. Snapshot of ENCODE showing 500bp flanking regions of the TSS of MYL9 (top) and PPDPF (bottom) methylation and histone modifications. DNA methylation in RRBS tracks is shown in red (100% methylated), yellow (50% methylated), or green (0% methylated). Representation of DNA methylation as found via 450K array is shown in either orange (methylated), purple (partially methylated), or blue (un methylated). Histone modification tracks are represented by dark bars with increased saturation that correlates with relative enrichment of the specified modification in that region. The red bar in the center is helps to identify location, with the center being at the transcription start site and a 250bp region blocked off on each side (where then thin red line ends); these regions are divided in half representing (2) 125bp regions.

Table 1. PCA of Hand Collected Data. Rotation scores of Principal Component Analysis of the hand collected data. Each PC column represents a single component. The columns 'Fold-Change' through 'Total Number of Histone Mods' illustrate the degree to which that feature is correlated with the others within the principal component. The last column of each principal component gives the square root of the eigenvalue, all values greater than one are considered due to more than chance alone. Blue cells highlight notable correlations within each principal component.

Principal Component Analysis of Hand Constructed Epigenetic Profiles

	PC1	PC2	PC3	PC4
FC	-0.09812491	0.185193116	0.6314743	0.257282221
Position In Relation To TSS	0.03980503	-0.339381031	-0.11278935	0.592713316
Distance From TSS	-0.00695304	0.314793766	-0.17110112	0.537234782
Number of Assayed CpGs	0.06132613	-0.618390433	0.11707592	0.021598222
RRBS Score Avg	-0.08297258	-0.424495013	0.28241413	0.337259781
450K Score Avg	-0.14234881	-0.269684318	0.35264368	-0.413215636
H3K4me1 Score	0.44825513	0.243908371	0.39616604	0.038108547
H3K4me3 Score	0.48907857	-0.040722519	0.1234678	-0.030282176
H3K27ac Score	0.37165298	-0.234310708	-0.40504704	-0.080905746
Total Number of Histone Mods	0.61609556	-0.006738708	0.07187605	0.001660682
Sqrt Of Eigenvalues	**1.570583**	**1.2145957**	**1.1638066**	**1.0405994**

	PC5	PC6	PC7	PC8	PC9	PC10
FC	-0.2414143	0.13983796	0.06265859	0.5084772	-0.386956136	0.032792409
Position In Relation To TSS	-0.5304724	0.36671769	0.0168857	-0.30057853	0.110570835	0.021546303
Distance From TSS	0.637095596	0.34401699	-0.23726479	0.03743419	-0.032017889	-0.00045185
Number of Assayed CpGs	0.140135798	-0.07050159	-0.55094884	0.44265778	0.274570832	0.013074267
RRBS Score Avg	0.407737892	-0.41687696	0.47041587	-0.21492549	-0.10886145	-0.00679564
450K score Avg	0.245024098	0.68236251	0.06727115	-0.27968773	-0.056589209	-0.01714881
H3K4me1 Score	0.036542941	0.01626512	0.13690133	-0.06706138	0.626351588	0.404994818
H3K4me3 Score	-0.00120342	-0.14602582	-0.42642402	-0.37627808	-0.551711414	0.306348508
H3K27ac Score	0.080410879	0.23772868	0.44412374	0.42459226	-0.223195787	0.38215447
Total Number of Histone Mods	0.026449626	0.07167503	0.11094039	0.0545624	-0.007310424	-0.7707378
Sqrt Of Eigenvalues	**0.946506**	**0.9003174**	**0.8486067**	**0.8269851**	**0.6643631**	**0.2624353**

Table 2. PCA of Computed Data. Rotation scores of Principal Component Analysis of the computed bioinformatics data. Each PC column represents a single component. The columns 'Fold-Change' through 'Total Number of Histone Mods' illustrate the degree to which that feature is correlated with the others within the principal component. The last column of each principal component gives the square root of the eigenvalue, all values greater than one are considered due to more than chance alone. Blue cells highlight notable correlations within each principal component

Principal Component Analysis of Computed Epigenetic Profiles

	PC1	PC2	PC3	PC4
Fold Change	0.131375553	0.14748867	-0.7879879	0.09776099
Position In Relation To TSS	0.026070224	0.07388047	0.2404684	0.91011328
Distance From TSS	0.005886808	-0.30238467	-0.3594781	0.37915005
Number of Assayed CpGs	0.176029433	0.63569519	0.2358044	-0.01317289
450K Score Avg	0.100462306	0.65987767	-0.2170623	0.04730746
H3K4me1 Score	0.500229216	-0.10412497	-0.1858967	-0.03045763
H3K4me3 Score	0.257485789	-0.04817444	0.1610774	0.08986901
H3K27ac Score	0.517416978	-0.12154507	0.1353418	-0.08279438
Total Number of Histone Mods	0.597242486	-0.1177683	0.102396	-0.01055019
Sqrt Of Eigenvalues	1.4334911	1.1611203	1.0212125	1.0002422

	PC5	PC6	PC7	PC8	PC9
Fold Change	-0.30892473	0.322235612	-0.35365087	-0.06340866	-0.0463204
Position In Relation To TSS	-0.32348852	-0.000148667	0.02833776	0.04352255	0.01971559
Distance From TSS	0.712587484	-0.319780348	-0.12477454	-0.081815	0.05698113
Number of Assayed CpGs	0.202725707	-0.108250694	-0.40292148	-0.51706655	0.16312135
450K Score Avg	0.227650898	-0.126174253	0.42075515	0.49789252	-0.11152593
H3K4me1 Score	-0.18390222	-0.175912817	0.56214626	-0.38257798	0.41988543
H3K4me3 Score	0.402867175	0.838200015	0.13536099	0.01386692	0.11816006
H3K27ac Score	-0.05537825	-0.153729702	-0.428748	0.56314274	0.40580784
Total Number of Histone Mods	0.005327858	-0.095441904	-0.0319823	-0.09191682	-0.77474298
Sqrt Of Eigenvalues	0.9794813	0.9631365	0.8340889	0.7741162	0.6095637

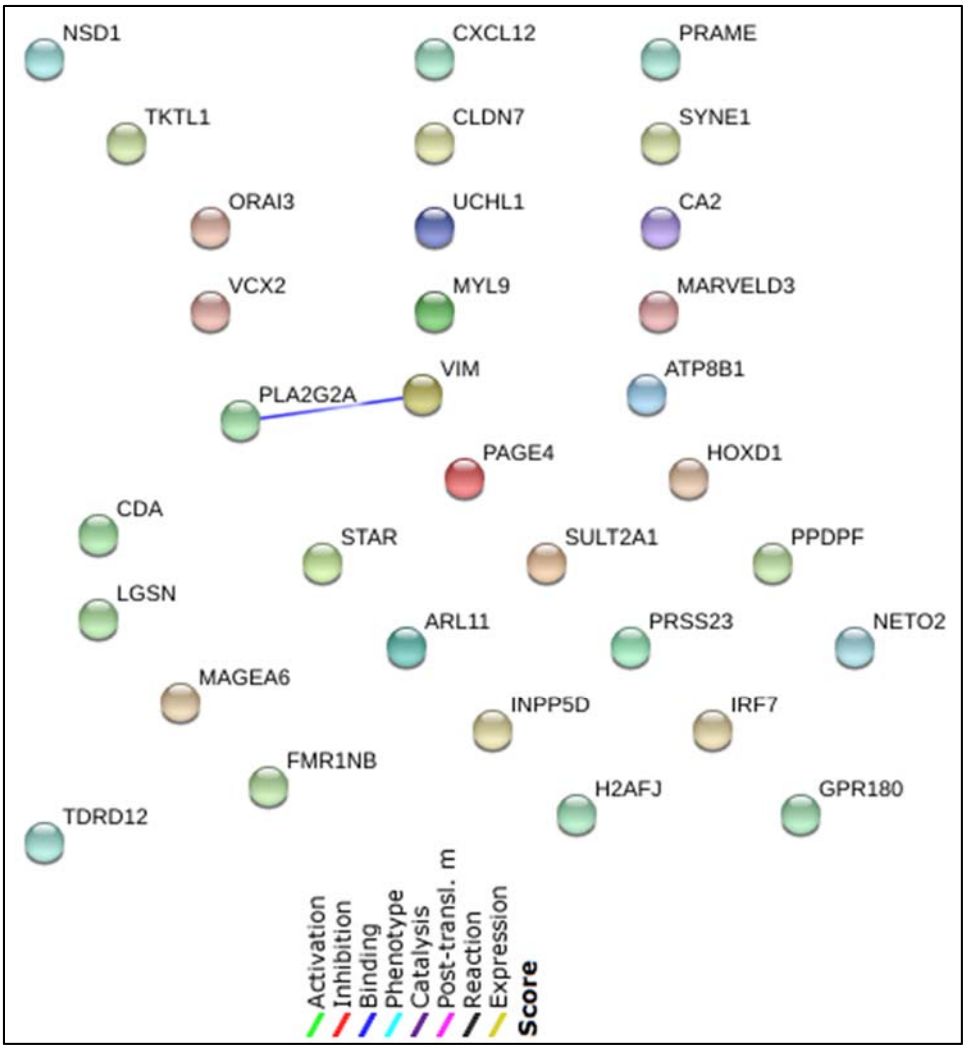


Figure 4. Protein Interactions. Analysis of protein interactions for the 32 selected proteins via String ver. 10. Proteins are spread for the purposes of visualization.

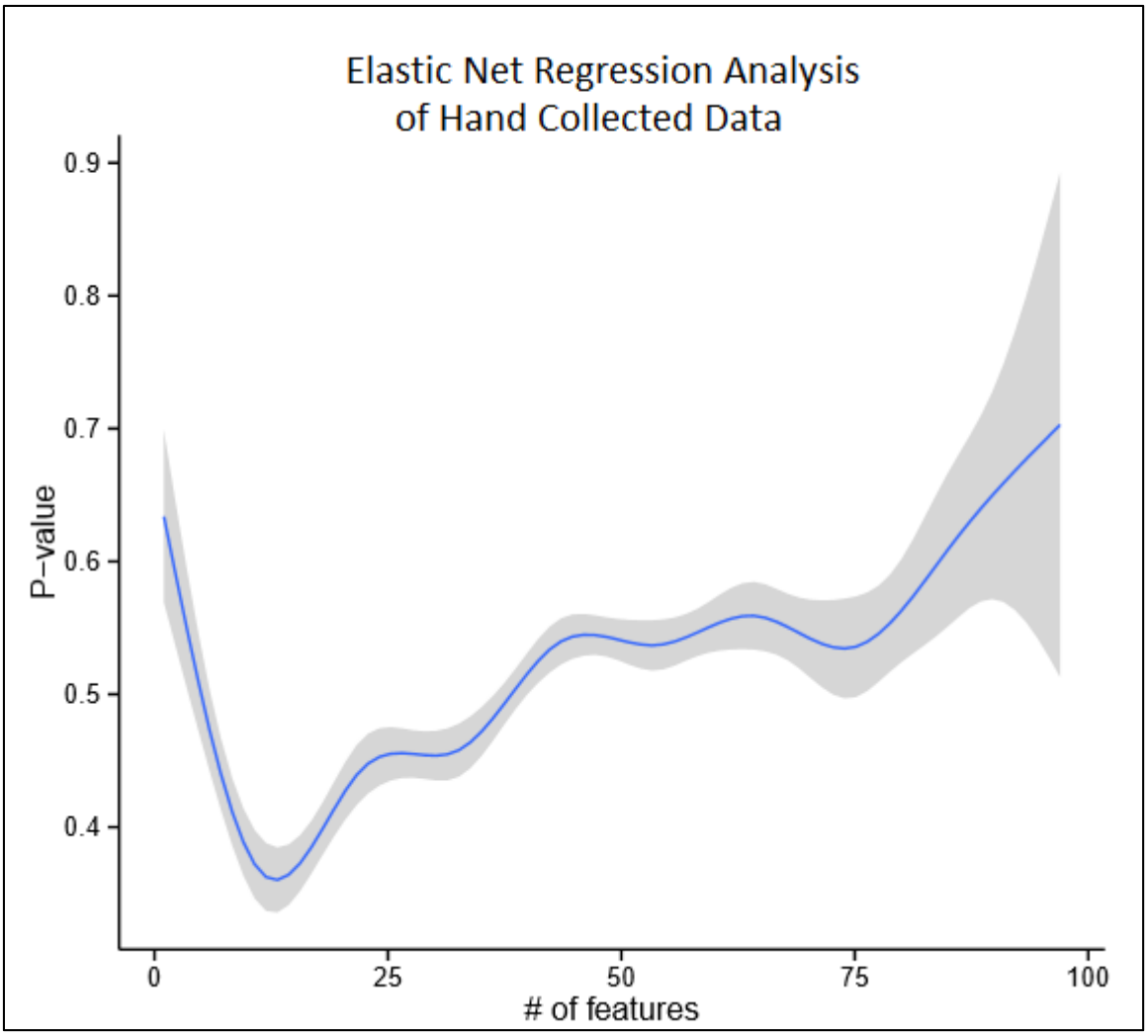


Figure 5. Elastic Net Regression Analysis of Hand Collected Data. Elastic Net Regression Analysis of hand collected 32 genes from the HepG2 and HCT-116 microarrays and bioinformatics analysis. Shown in blue is the smoothed curve of the data. The gray shaded region illustrates the 99% confidence interval.

CHAPTER III

VERIFICATION OF MICROARRAY DATA

During replication DNA methyltransferase 1 (DNMT1) copies the DNA methylation pattern from the parent strand onto the daughter strand (Robertson & Jones, 2000). Thus, the established methylation pattern is maintained throughout a cell lineage. In contrast, DNMT3a and DNMT3b are responsible for *de novo* DNA methylation (Tsai et al., 2012). The DNA methyltransferase inhibitor, 5-Azacytidine (5-Aza), interferes with DNMT1, preventing the re-establishment of DNA methylation (Robert et al., 2003). Therefore, treatment with 5-Aza will result in reduced DNA methylation. The reduction in DNA methylation is dependent partially upon the number of cell divisions and consequently the amount of time that a cell is exposed to the drug. In the 5-Aza microarray studies of HCT-116 (E-GEOD-57341) and HepG2 (E-GEOD-5230) available on ArrayExpress, two different periods of time were used for the duration of 5-Aza exposure; 72 hours and 96 hours respectively (Dannenberg & Edenberg, 2006; Li et al., 2014). This difference in exposure time could cause a greater relative fold-change in HepG2 gene expression. Furthermore, just as length of treatment with 5-Aza can cause difference in DNA methylation levels, so to can the concentration (He et al., 2015; Huichalaf et al., 2014). In these microarray studies treatment concentrations varied by 5 fold; HCT-116 was treated with 0.5 μM 5-Aza whereas HepG2 was treated with 2.5 μM. Verifying the response of the genes to identical treatment conditions of 1 μM 5-Aza for 72 days, a common treatment for cell lines using the drug, would provide

validation of the gene expression data in both cell lines under identical conditions and to see if the trend of re-expression was maintained.

The goal of the present study was to determine the response of four genes under identical treatment conditions, selected based on their unique epigenetic profiles and differential response to 5-Aza in the microarrays, both in regards to differing between cell lines, and the genes also having very different levels of fold-change from one another. Also, the microarray data represents relative fold-change comparing treated to untreated. Examination of the C_T values of the qPCR data will allow us to determine if the genes are actually in an inactive state and activated by 5-Aza or rather increased in expression by 5-Aza. This has implications regarding the interpretation of the 5-Aza microarray data. To meet this goal, we employed quantitative real-time PCR to investigate the relative fold-change of the four genes in comparison to β-actin between treated and untreated HepG2 and HCT-116 cells.

Materials and Methods

Cell Culture and Reagents

HCT-116 (obtained from ATCC, Cat. No. CCL-247) was cultured in McCoy's 5A medium, containing 10% FBS, 100 units/mL penicillin, and 100µg/mL streptomycin. HepG2 (obtained from ATCC, Cat. No. HB-8065) was cultured in Eagle's Minimum Essential Medium, containing 10% FBS. Both cell lines were cultured at 37°C in a humidified atmosphere with 5% CO_2/95% air.

5-Aza

5-Aza was purchased from Acros (Cat. No. 226620500). Stock solutions of 5-Aza (220mM) were prepared in 1:1 Acetic Acid:H_2O and stored at -20°C. Stock

solutions were further diluted (22mM) in 1:1 Acetic Acid:H_2O immediately prior to treatments in cell culture experiments.

Treatment and Collection of Cells

HCT-116 and HepG2 were plated in 100mm plates at $7x10^5$ cells. The next day the cells were treated with and without 1μM 5-Aza. Treatment continued for the next 72 hours with the medium and 5-Aza being replaced every 24 hours. At the end of treatment the medium was removed and the cells washed with 5mL of cold 1x PBS. Another 5mL of 1x PBS was added and the cells were scraped off the plates and transferred to 15mL Falcon tubes. The cells were pelleted by centrifugation in a clinical centrifuge at setting #4. The supernatant was removed and the pellets quick frozen in liquid nitrogen and stored at -80°C.

RNA Isolation and Evaporation

Total RNA was isolated from cell pellets using the Promega SV total RNA Isolation System according to manufacturer's protocol. Absorbance at 260/280 nm was measured for RNA samples using the Synergy 2 multi-mode microplate reader (BioTek) and ng/μL was calculated. Samples with low concentrations were evaporated using a Savant SpeedVac Concentrator (SVC100H) and Savant Refrigerated Condensation Trap until concentrations were sufficient for cDNA synthesis. Isolated RNA was stored at -80°C until use.

Conversion to cDNA

mRNA was converted to cDNA in 20μL reactions using Maxima's First Strand cDNA synthesis kit and following manufacturer's instructions for use of 2μg mRNA. The product was then diluted 1:1 using nuclease-free H_2O.

qPCR of Specific Transcripts

Quantitative real-time PCR (qPCR) was carried out in 96-well reaction plates in 10μL reactions using TaqMan® Gene Expression Master Mix from Applied Biosystems. Reactions were run on Applied Biosystems' StepOnePlus Real time PCR system with the following profile: 95°C for 15 seconds and 60°C for 60 seconds for 40 cycles. Relative expression of mRNA was calculated by the ΔΔCt method. This was done by obtaining the ΔCt value for each sample by subtracting the average Ct value of Actin mRNA from the average Ct value of the target gene's mRNA. The ΔΔCt value was then obtained by subtracting the treated Ct value from the untreated Ct value. The mRNA's fold-change was reported using the equation $2^{\wedge(\Delta\Delta Ct)}$ where ΔΔCt = ((target average experimental Ct value – Actin average experimental value) – (target average control Ct value – Actin average control value)). All primers used in qPCR were obtained from Applied Biosystems: Actin (Cat. No. Hs00357333_g1), PPDPF (Cat. No. Hs00225594_m1), H2AFJ (Cat. No. Hs1674946_s1), MARVELD3 (Cat. No. Hs0036935_m1), and MYL9 (Cat. No. Hs00697086_m1).

Results

In order to verify that the genes from the microarray studies trended in the same direction as our own treatment conditions and to illustrate their validity we treated HCT-116 and HepG2 cells with 1μM 5-Aza for three days. These conditions were used in order to fall between the two experimental conditions; doubling the concentration seen in the HCT-116 microarray yet meeting the same duration, while reducing the concentration and duration used in the other.

We hypothesized that treatment of HCT-116 and HepG2 with 5-Aza would change the transcript levels of PPDPF, H2AFJ, MARVELD3, and MYL9. Additionally, we suspected that these changes in transcript levels would be due to changes in the patterns of epigenetic profiles near the TSS of the genes.

Using qPCR we examined the relative fold-change of these four selected targets of the microarrays. These targets came from the list of 32 genes used in the hand constructed epigenetic profiles and were chosen based on differential levels of re-expression. MYL9 in HepG2 possessed the greatest fold-change relative to actin and the largest difference between cell lines. The gene with the lowest level of reactivation of the four genes was PPDPF, with minor difference between HCT-116 and HepG2. MARVELD3 showed similar levels of up-regulation in both cell lines, whereas in H2AFJ HepG2 was more upregulated than HCT-116 (Fig 6).

We were also able to determine the trend of upregulation using the qPCR (Table 3). Two of the four genes, MARVELD3, and MYL9 matched the fold-change trends seen in the microarray in bothHCT-116 and HepG2. H2AFJ matched the trends only in HepG2.

Analysis of the average C_T values of the same genes with and without 5-Aza treatment revealed that three of the four genes had average C_T values of less than 30. These values suggest that three of them (PPDPF, MARVELD3, and MYL9) are already transcribed in HCT-116 and HepG2. The level of expression prior to 5-Aza treatment does not seem to impact the sensitivity of the gene to 5-Aza treatment. For example, while MYL9 had a C_T value in the twenties, suggesting that it was expressed prior to treatment with 5-Aza, it responded more so than did H2AFJ which had an average C_T value of more than 30, and was therefore considered to be expressed only at very low

levels. We observed that while some genes that were “on” seemed to respond better to treatment than those that were “off” this trend was not consistent, as seen between H2AFJ and PPDPF with an average C_T value in the twenties but least amount of response in all four genes.

When viewing the epigenetic profiles for these genes clear distinctions can be made in regards to the patterns of DNA methylation and histone modifications. In all four genes, there was an alteration of the density and level of DNA methylation compared to the other genes. Between cell lines, the greatest difference in DNA methylation and histone modifications can be seen in both MYL9 and PPDPF (Fig. 3). In MARVELD3 there is minimal DNA methylation in either cell line within our 1000bp window, yet differences between histone modifications exist.

Discussion

The goal of this study was to validate the data from the microarrays. The findings from the qPCR analysis were consistent with both our hypotheses that treatment with 5-Aza would increase transcript levels and that the increase is based on epigenetic profiles. They also indicate that upregulation was not dependent upon the level of transcription prior to drug treatment. These findings support our model, in that the genes which possessed lower levels of DNA methylation near the TSS will show less change when treated with 5-Aza. Genes with a greater number of methylated CpG’s near the TSS will show a greater change when treated with 5-Aza.

Table 3 is a breakdown of the methylation patterns and histone modifications in the 4 genes being analyzed and Table 4 shows the fold-change of these genes after 5-Aza treatment, analyzed in the microarrays and our qPCR, these changes can also be

seen in Fig. 6. The role of histone modification in relation to the 5-Aza response is more complex and less easy to predict. However, it is known that more active histone marks (along with active transcription) preclude DNA methylation and the presence of particular histone marks enhance DNA methylation. We were able to find histone tracks for only active histone marks. It is likely however, that high levels of active histone marks will be associated with less activation by 5-Aza. Of these patterns, the MYL9 gene in HepG2 fits our model, with 15 sites of methylation and the histone modifications H3K4me1, falling in the 250bp regions upstream and downstream of the TSS with H3K4me3 also being found in the 250bp upstream region. An additional 12 sites of methylation were found outside of the model's 250bp flanking regions and the continuation of H3K4me1 across the entire 1000bp region along with the presence of all 3 histone modifications in the furthest end of the downstream region added to the epigenetic profile of MYL9. While there were more epigenetic modifications within the profile of MYL9 than the 10 to 15 features suggested by the model, this gene had the greatest response in either cell line, with HepG2 possessing a fold-change of 15.52 in our experiment and 28.59 seen in the microarray.

While this change was seen after 5-Aza treatment in HepG2, it was not seen in HCT-116 for MYL9, which only experienced a fold-change of 1.33 in our experiment and 1.13 in the microarray. These levels indicate that there was no response to treatment with 5-Aza. HCT-116 did not fit our model, in that it possessed an overall lower level of methylation with only 2 sites of methylation in the -250 - -500bp region and the H3K4me3 mark downstream of the TSS.

Another gene that came close to fitting our model was H2AFJ in HepG2. In this cell line H2AFJ contained 21 sites of methylation and the histone marks H3K4me3 and

H3K27ac within the first 250bp flanking each side of the transcription start site. Further upstream and downstream, within our 1000bp window were an additional 3 sites of DNA methylation. This gene slightly exceeded the number of features recommended by our model, yet experienced a fold-change of 1.8 and 2.08 in our experiment and the microarray respectively. In H2AFJ, the sites of methylation are primarily found downstream, pack relatively closely together. This likely contributed to the lower level of expression than seen in MYL9 (15.52 and 28.59), which had DNA methylation spread loosely on both sides of its transcription start site.

In HCT-116, H2AFJ did not respond in either our qPCR (0.48) or the microarray (1.16). Rather, there is a decrease in gene transcription in our qPCR after treatment with 5-Aza. The lack of response measured in the microarray matches our prediction based on the model, as H2AFJ in HCT-116 exceeded the recommended number of features, possessing 28 locations of DNA methylation within the first 250bp, upstream and downstream along with H3K4me3 and H3K27ac, with the majority (23 sites) of the methylation being located upstream of the TSS. A total of 3 other locations were found to have DNA methylation in both the -/+500bp regions with continued H3K4me3 and H3K27ac marks.

PPDPF failed to meet the criteria of our model in both HCT-116 and HepG2 and possessed a fold-change of only 0.37 and 1.11 for our qPCR and the microarray in HCT-116 and 0.67 in our experiment and 1.75 from the microarrays for HepG2. In HCT-116 there was no DNA methylation and only 2 histone marks, H3K4me3 and H3K27ac. HepG2 possessed DNA methylation downstream only, with 8 being found in the first 250bp and an additional 21 from 251bp – 500bp. It also contained the histone mark H3K4me3 throughout the entirety of the 1000bp window. The levels of fold-change

measured in PPDPF via the microarray, illustrate a lack of response when there is little to no methylation, while our own results suggest that there was a repression of transcription upon 5-Aza treatment.

MARVLED3 is another example of a gene that does not fit the model, as it contains no DNA methylation in either HCT-116 or HepG2. In HepG2 H3K4me3 is found throughout the 1000bp region, while in HCT-116 it is seen only from TSS to 500bp downstream. It is not surprising that our qPCR found MARVELD3 to have no change as indicated by the 1.08 and 1.23 levels of fold-change in HCT116 and HepG2 respectively. However, the microarray found different responses after treatment after 5-Aza, with fold-change levels 4.21 in HCT-116 and 2.36 in HepG2, indicating that the gene had responded and done so more in HCT-116 than in HepG2.

Overall, we found that 2 of the 4 verified genes in HepG2 fit our model's requirements for reactivation, while none fit the requirements in HCT-116. Those genes that did not fit the requirements met our prediction that there would be little response in a gene that did not fit the requirements based on the model. Additionally, we were able to verify through qPCR that 3 of the 4 genes, MYL9, H2AFJ, and PPDPF, matched the fold-change trends seen in the microarray. One gene did not match the trend of expression between our experiment and the data gathered from the microarray, MARVELD3, which possessed a fold-change of 1.08 in HCT-116 and 1.23 in HepG2 as seen in our qPCR. In the microarrays MARVELD3 showed a fold-change of 4.21 and 2.36 for HCT-116 and HepG2 respectively.

An investigation of the individual C_T values of MARVELD3 illustrates that there was no reading that skewed the data, as even in the most extreme case cycle times differed by less than 0.7. The inversion of response between HCT-116 and HepG2 seen

in H2AFJ could likely be due to the fact that the C_T values were near the thirties, where there is a greater margin of error. When low levels of transcript are present during the initial rounds of the polymerase chain reaction, lack of replication of even a single strand during the first reaction can result in a reduction of the transcripts at the 30^{th} cycle by more than a billion copies. Alternatively, there may have been an indirect effect on this gene or a gene enhancer outside of our considered region.

Another issue that must be considered is indirect effects of the 5-Aza. It is possible that the change in expression of some of these genes, particularly MARVELD3 with no methylated CpGs, is due to secondary effects. That is, another gene is being activated and affecting MARVELD3, either as an enhancer or transcription factor. Similar results with another gene have been seen in other studies and confirmed that indirect mechanisms were affecting the upregulation of genes after 5-Aza treatment (Angst et al., 2011). It is also possibly affecting the mRNA stability. This is important as our basic assumption for the entire project rests on the idea that most changes in gene expression are due to 5-Aza treatment and the direct effects on a particular gene. Of the four genes we selected, it is most likely that MYL9 is directly affected by 5-Aza. This conclusion is based on the finding that its epigenetic profile of methylation in HepG2 most closely matches our model and it's total of 27 sites of DNA methylation are spread throughout the 1000bp region. Had this been an indirect effect, it is likely that we would have seen a response in HCT-116 similar to that of HepG2. This was not observed.

As an alternative to the microarray, RNA-seq would have been a more effective method as it would have been able to more accurately determine the level of gene expression before and after treatment. Our method, while useful in determining reactivation or upregulation of a gene, does so based on a comparison between the

level of fluorescence of a given transcript and that of the transcript for β-Actin. Using RNA-seq would provide a more accurate quantification of lowly expressed genes, while simultaneously giving an exact level of gene transcription. The use of RNA-seq would allow for identification of “on” / ”off” genes while potentially avoiding the issues seen with MARVLED3.

Table 3. Count of DNA Methylation and Histone Modifications of Four Verified Genes. A count of the DNA methylation and list of the histone modifications present in each 250bp region of the 1000bp window of investigation for the four verified genes (PPDPF, H2AFJ, MYL9, and MARVELD3).

Count of DNA Methylation and Histone Modifications of Four Verified Genes

GENE SYMBOL	Cell Line	-500		-250		+250		+500	
		Histones	DNAm	Histones	DNAm	Histones	DNAm	Histones	DNAm
PPDPF	HCT-116	H3K4me3 H3K27ac	0	H3K27ac	0	H3K4me3 H3K27ac	0	H3K27ac	0
	HepG2	H3K4me3	0	H3K4me3	0	H3K4me3	8	H3K4me3	21
H2AFJ	HCT-116	H3K4me3 H3K27ac	2	H3K4me3 H3K27ac	5	H3K4me3 H3K27ac	23	H3K4me3 H3K27ac	1
	HepG2	H3K4me3 H3K27ac	2	H3K4me3 H3K27ac	5	H3K4me3 H3K27ac	16	H3K4me3 H3K27ac	1
MYL9	HCT-116	N/A	2	N/A	0	H3K4me3	0	H3K4me3	0
	HepG2	H3K4me1	2	H3K4me1	6	H3K4me1 H3K4me3	9	H3K4me1 H3K4me3 H3K27ac	10
MARVELD3	HCT-116	N/A	0	N/A	0	H3K4me3	0	H3K4me3	0
	HepG2	H3K4me3	0	H3K4me3	0	H3K4me3	0	H3K4me3	0

Table 4. qPCR VS Microarray Results for 4 Verified Genes. Relative fold change as determined by qPCR in the four verified genes (PPDPF, H2AFJ, MARVELD3, and MYL9) compared to fold change seen in the microarrays. Difference (experimental/microarrays) is the difference between the two cell lines for a given gene.

qPCR VS Microarray Results for 4 Verified Genes

Gene Name	Cell Line	Fold Change (Experimental)	HCT116 : HepG2	HepG2 : HCT116	Fold Change (Micro-arrays)	HCT116 : HepG2	HepG2 : HCT116
PPDPF	HCT116	0.37	0.55		1.11	0.63	
	HepG2	0.67		1.81	1.75		1.58
H2AFJ	HCT116	0.48	0.27		1.16	0.56	
	HepG2	1.8		3.75	2.08		1.8
MARVELD3	HCT116	1.08	0.88		4.21	1.78	
	HepG2	1.23		1.14	2.36		0.56
MYL9	HCT116	1.33	0.09		1.13	0.04	
	HepG2	15.52		11.67	28.59		25.3

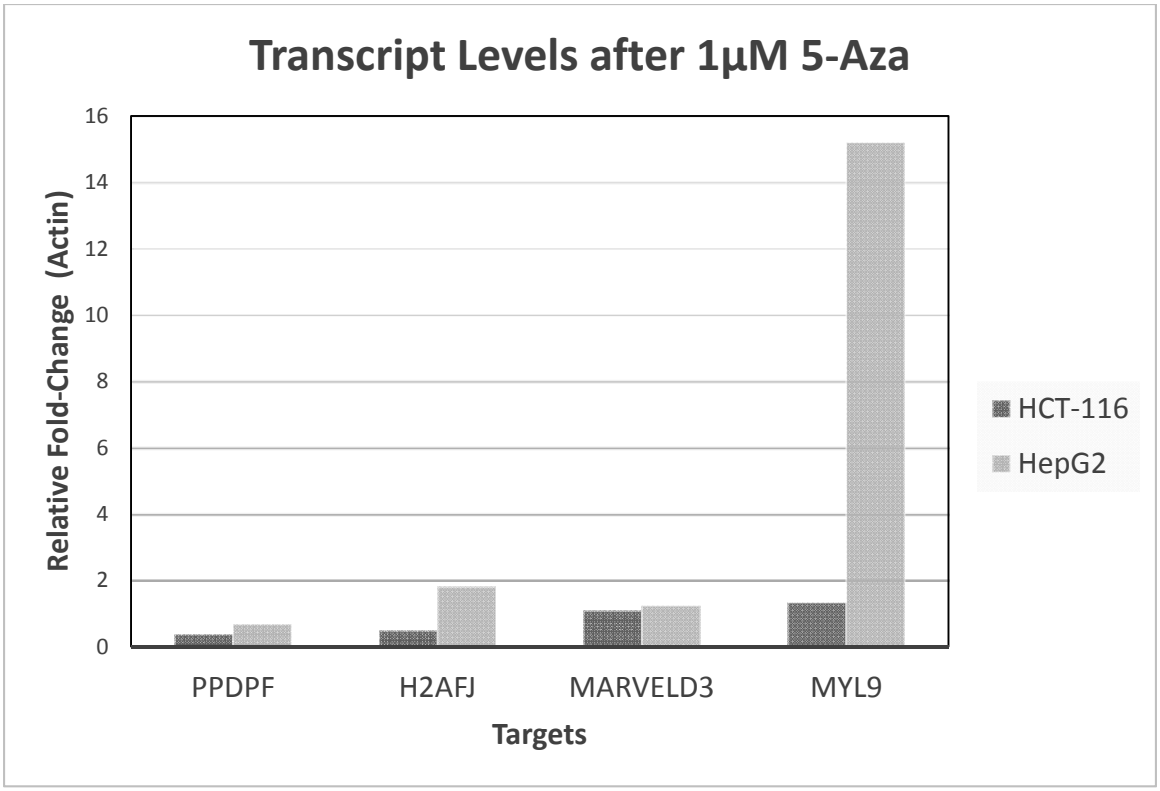


Figure 6. Graph of Relative Fold-Change in Verified Genes. Reactivation of verified genes: PPDPF, H2AFJ, MARVELD3, and MYL9 after treatment with 5-Azacytidine in HCT-116 and HepG2.

CHAPTER IV

PREDICTING THE RE-EXPRESSION OF NOVEL GENES USING EPIGENETIC PROFILES

The major goal of this study was to determine if the model that we derived in Chapter II can be applied to the re-expression of a novel gene by treatment of 5-Aza. That is, will the epigenetic profile of the novel gene be predictive of its expression. For this experiment we used the gene WNT5A.

WNT5A is a secreted protein ligand that is misregulated in multiple forms of cancer and plays a role in migration, adhesion, differentiation, proliferation, and apoptosis(Kikuchi et al., 2012). In some cancers, WNT5A is upregulated and recognized as an oncogene, whereas in others it is downregulated and implicated as a tumor suppressor (Hung et al., 2014; Ying et al., 2008). Further studies have found that this gene possesses two isoforms (Fig. 7). There is limited evidence that these isoforms are functionally distinct (Bauer, er al.). Importantly, for this study the isoforms are derived from unique mRNA's, transcribed from unique promoters. The TSS of these two promoters lie 6,244 base pairs away from one another and therefore represent adequate distance that does not overlap in the use of our epigenetic profiling. In this lab we had previously investigated the methylation status and transcription of WNT5A in multiple cell lines. WNT5A isoform A and B promoters were found to be inactive in HCT-116 cells. (Vaidya et al., 2016, and unpublished). A second goal of this study was to determine if WNT5A could be further re-activated using a combination of the DNA methyltransferase inhibitor, 5-Azacytidine, and the histone deacetylase inhibitor, SAHA also known as Vorinostat or suberoylanilide.Hydroxamic. This drug is a histone deacetylase inhibitor

that inhibits all class I and class II HDACs (Marks, 2007) that has been approved by the FDA for use in cutaneous T-cell lymphoma (W. S. Xu et al., 2007). SAHA binds to the catalytic region of these HDACs, preventing the removal of the acetyl group from histones (Yoon & Eom, 2016). It has been shown in previous studies that combinations of HDACi and DNMTi can have a synergistic effect (Chen et al., 2012; Leclercq et al., 2011).

To achieve our goals we used ENCODE to examine the pattern of DNA methylation and histone modifications surrounding the WNT5A promoter A and promoter B. Then HCT-116 cells were treated with combinations of the DNA demethylating agent, 5-Aza and histone deacetylase inhibitor, SAHA. The derived pattern of DNA methylation and histone modifications were then compared to the expression pattern.

Materials and Methods

Cell Culture and Reagents

HCT-116 (obtained from ATCC, Cat. No. CCL-247) was cultured in McCoy's 5A medium, containing 10% FBS, 100 units/mL penicillin, and 100µg/mL streptomycin. HepG2 (obtained from ATCC, Cat. No. HB-8065) was cultured in Mem-C medium, containing 10% FBS. Both cell lines were cultured at 37°C in a humidified atmosphere with 5% CO_2/95% air.

Treatment with 5-Aza and SAHA

5-Aza was purchased from Acros (Cat. No. 226620500). Stock solutions of 5-Aza (220mMol/L) were prepared in 1:1 Acetic Acid:H_2O and stored at -20°C. Stock solutions were further diluted (22mMol/L) in 1:1 Acetic Acid:H_2O immediately prior to treatments in cell culture experiments. Treatment with 5-Aza (either 1µM or 5µM) took

place over 72 hours with medium and 5-Aza being replaced every 24 hours. SAHA was obtained from TSZ Chem (Cat. No. RV005) and stored at -20°C. Solutions of SAHA (10mMol/L) were prepared in DMSO immediately prior to use in cell culture experiments. Treatment with SAHA (either 1μM or 5μM) took place over 24 hours. Cells were collected in 5mL PBS after treatment, supernatant was removed, cell pellets were frozen and stored at –80°C.

RNA Isolation and Evaporation

Total RNA was isolated from both cell lines using Promega's SV total RNA isolation system according to manufacturer's protocol. Absorbance at 260/280 nm was measured for RNA samples using the Synergy 2 multi-mode microplate reader (BioTek) and ng/μL was calculated. Samples with low concentrations were evaporated using a Savant SpeedVac Concentrator (SVC100H) and Savant Refrigerated Condensation Trap until concentrations were sufficient for cDNA synthesis. Isolated RNA was stored at -80°C until use.

Conversion to cDNA

mRNA was converted to cDNA in 20μL reactions using Maxima's First Strand cDNA synthesis kit and following manufacturer's instructions for use of 2μg mRNA. The product was then diluted 1:1 using nuclease-free H_2O.

Quantification of mRNA, qPCR

Quantitative real-time PCR was carried out in 96-well reaction plates in 10μL reactions using TaqMan® Gene Expression Master Mix from Applied Biosystems. Reactions were run on Applied Biosystems' StepOnePlus Real time PCR system with the following profile: 95°C for 20 seconds followed by 40 cycles of 95°C for 3 seconds and 60°C for 30 seconds. Relative expression of mRNA was calculated by the ΔΔCt

method. This was done by obtaining the ΔCt value for each sample by subtracting the average Ct value of Actin mRNA from the average Ct value of the target gene's mRNA. The ΔΔCt value was then obtained by subtracting the treated Ct value from the untreated Ct value. The mRNA's fold-change was reported using the equation $2^{\wedge(\Delta\Delta Ct)}$. All primers used in qPCR were obtained from Applied Biosystems: Actin (Cat. No. Hs00357333_g1), WNT5A isoform A (Cat. No. AID1VPA) and WNT5A isoform B (Cat. No. AIFATVI).

Construction of "epigenetic profiles"

"Epigenetic profiles," defined as 1000bp regions centered at the transcription start site and consisting of DNA methylation and histone mark scores and abundance, were constructed in two different manners; initially by hand for a group of 32 selected genes and later by using a mixture of GALAXY, ENCODE, and Excel.

Hand construction of the epigenetic profiles was performed using ENCODE to explore the methylation and histone modification tracks of 32 genes selected from the microarray studies based on differential re-expression of any given gene between cell lines. Each track was examined in windows of 50 base pairs each starting at the transcription start site and extending 500 base pairs in each direction. Methylation data was collected for HCT-116 and HepG2 by analyzing two RRBS tracks for HCT-116 and two RRBS tracks for HepG2; these numbers were averaged together for the tracks of each cell line and recorded. Additionally, two 450K tracks for HCT-116 and one 450K track for HepG2 were collected and recorded. When more than one track was available the scores for 450K were averaged in the same manner as RRBS. When analyzing histone modifications, we selected those which were shared between the two cell lines: H3K4me1, H3K4me3, and H3K27ac tracks (all which are associated with active

promoters). Two tracks were used for analysis of HCT-116 H3K4me3 (Peaks 1 & 2), while 3 were used to analyze the same modification in HepG2 (Peaks 1, 2, & ENCODE/Broad). In the event that a peak was detected that stretched outside of the 50 base pair window, the score was carried over into all 50bp windows that it was located in, so long as it accounted for at least 25bp of that window. No scores were given for anything further than 500bp up/downstream. If more than one peak was present in a single track set, the two scores were averaged if they were the same length, so long as both accounted for at least 25bp of the given window and they were the same length, if not the shorter band's score was taken since it is considered more accurate by ENCODEs standards. For all histone tracks and the 450K track, numbers ranged from 0 – 1000; the RRBS track ranged from 0 – 100. These numbers indicated the RRBS track indicated the percentage of reads that showed methylation at a given CpG, whereas the 450K track scored based on the microarrays beta values, multiplied by 1000. Histone tracks were scored based on the signal value, indicating the average enrichment for the region. When available, multiple tracks of a given epigenetic mark are used as a form of biological replicates.

Results

We examined the bioinformatics information available on ENCODE to construct the epigenetic profiles of WNT5A isoform A and B promoters. As before, we used a 1000bp window of 500bp upstream and downstream of the TSS of WNT5A promoters (Fig. 8). We found that HCT-116, WNT5A promoter A, contained 12 sites of DNA methylation. Eight of these sites are located from +500 to +251bp, 2 from +250 to the TSS and 2 more from the TSS to +250bp downstream. These scores come from a combination of both the 450K and RRBS tracks. Generally, these sites were found

downstream of TSS, located in Exon1 (Fig11A), while only two of them were found upstream. Four sites of methylation occurred spread out over 250bp upstream and downstream of the TSS. The eight remaining sites were found further downstream. The same sites were analyzed in HepG2 for promoter A, yet only two sites were found by the 450K microarray to have partial methylation scores; all other sites in the region were found to have no methylation. In both HCT-116 and HepG2, WNT5A promoter B, contained 4 identical sites of DNA methylation (Fig. 8, Fig. 9B). These sites were located in the TSS to +250bp region as identified by 450K.

Both promoter A and promoter B lacked any histone modifications for HCT-116 and HepG2 within our window. Expansion of our window revealed a single H3K4me1 histone mark over 4000bp upstream of the TSS for promoter B and over 2000bp downstream of the TSS for promoter A, separating the two. This suggests that the two promoters are heavily regulated by DNA methylation and transcription factors.

Bisulfite sequence analysis had been completed in the lab on a portion of promoter B including 4 CpGs. All four of these sites were found to be methylated (Fig. 9B). These sites fall within the 1000bp window of our WNT5A promoter B epigenetic profile and were not found to be methylated in the ENCODE tracks (Fig. 8). Based on these findings it would be predicted that the high-throughput sequencing methods utilized to generate the ENCODE data do not pick up every site of methylation.

From our analysis of the epigenetic profiles of the two WNT5A promoters in HCT-116 and HepG2 we expected HCT-116 promoter A to have the greatest response to 5-Aza. These expectations come from the fact that it more closely resembles our model generated after PCA and elastic net regression due to its 12 sites of DNA methylation

within the entire 1000bp region, 4 of which fall within the 250bp flanking regions of the transcription start site. To investigate this, we treated HCT-116 and HepG2 cells with 1μM 5-Aza for 72 hours and performed qPCR using custom, promoter specific probes. We found that WNT5A promoter A in HCT-116 became re-expressed to a higher degree than that HepG2 or isoform B in either HCT-116 or HepG2. Furthermore, WNT5A promoter B experienced the same level of relative fold-change in both HCT-116 and HepG2, while promoter A in HepG2 was found to be re-expressed more than promoter B, but less than that of promoter A in HCT-116.

The epigenetic profiles of the WNT5A promoters A and B regions that were derived from ENCODE did not include any histone modifications, active or repressive. Taking this finding into account and based on our model, we would predict that drugs affecting histone acetylation or methylation would have an effect on promoter A and B expression.

To test this prediction we treated HCT-116 with1μM or 5μM 5-Aza and/or SAHA for 72 hours with 5-Aza and applying SAHA for the last 24 hours. We found that WNT5A promoter A was reactivated with all treatments. Used alone, 1μM 5-Aza increased expression 3.68 fold, and 5μM would cause a fold-change of 6.17. Using SAHA alone increased the fold-change to 5.35 at 1μM and to 8.15 at 5μM. When used in combination the effect was even greater, showing a fold change of 26.24 (1μM 5-Aza, 1μM SAHA), 79.03 (1μM 5-Aza, 5μM SAHA), and 143.74 (5μM 5-Aza, 1μM SAHA). Promoter B was found to be less responsive to treatment, showing a fold change of 1.43 with 1μM 5-Aza, and 0.39 when treated with 5μM 5-Aza. SAHA caused a fold-change of 0.64 at 1μM and 1.86 at 5μM. When used in combination, the fold-changes of 38.68 (1μM 5-Aza, 1μM SAHA), 36.95 (1μM 5-Aza, 5μM SAHA), and 43.15 (5μM 5-Aza, 1μM

SAHA) were measured. Again, promoter A was more responsive to treatment than promoter B.

Discussion

The epigenetic profiles of HCT-116 and HepG2, as determined by the ENCODE analysis was found to be identical for WNT5A promoter B. Each cell line showed only 4 sites of methylation at the same location and both lacked any form of histone modification. The prediction, based on our model was that these cell lines would display a low level of response when exposed to 5-Aza treatment. The results show that promoter B is unresponsive to 5-Aza in both HCT-116 and HepG2. The epigenetic profiles of promoter A for HCT-116 and HepG2 were distinctly different with HCT-116 having 12 methylated CpGs and HepG2 having only 2. This means that HCT-116 would be more responsive to treatment with 5-Aza because of the impact methylation has on transcription. Our results support this prediction, as promoter A in HCT-116 is activated nearly 6-fold by 5-Aza. However, in spite of only the two methylation sites, which are located in the +500 interval, promoter A is activated 3-fold in HepG2. Although the activation is less then in HCT-116, it is responsive to 5-Aza, unlike promoter B in HCT-116. Overall, these results, partially support the model that more DNA methylation is related to the responsiveness to 5-Aza.

Looking at epigenetic profiles in small windows such as our 1000bp region with 50bp bins can provide unique insights that might otherwise not be noticeable. Promoter B shows identical methylated CpG sites in both HepG2 and HCT-116. These 4 methylated sites and lack of histone modifications fall well below the model's number but allow us to look closer at the fold changes between cell lines with matching epigenetic

profiles. When treated with 1μM 5-Aza both HCT-116 and HepG2 promoter B was unaffected and showed no change. With identical epigenetic profiles and matching treatment conditions, the identical response of promoter B in both cell lines matches expectations.

Unpublished data from our lab illustrates the limits of bioinformatics data. We found 4 additional sites of methylation in promoter B that had not been assayed by the RRBS or 450K tracks of HCT-116 used in ENCODE. It is likely then, that many more methylation sites were missed in WNT5A promoter A and promoter B along with other genes of both of these cell lines. While there may be other sites of methylation, we have based our model on information that is obtainable through bioinformatics and databases such as ENCODE and therefore, feel that the model is still applicable.

In comparison to the four genes we verified (PPDPF, H2AFJ, MARVELD3, and MYL9), MYL9 showed the highest level of fold-change in both cell lines; 15.51 in HepG2 and 1.33 in HCT-116. In comparison, WNT5A promoter A showed a fold-change of 6.16 in HCT-116 and 3.9 in HepG2. Based on our model, it would be predicted that the epigenetic profiles of MYL9 and WNT5A Promoter A would include methylated sites and that MYL9 would be more highly methylated in HepG2 and WNT5A would be more highly methylated in HCT-116. In fact, MYL9 was found to have 27 methylation sites in the epigenetic profile for HepG2 and only 2 for HCT-116. In contrast, WNT5A promoter A was found to have 12 methylation sites in HCT-116 and only 2 in HepG2. Thus, the epigenetic profiles of the MYL9 gene and novel gene WNT5A promoter A fit the prediction. Other factors likely contribute to the unique epigenetic profiles of each gene.

We predicted that SAHA, a histone deacetylase inhibitor would have an effect on promoters A and B expression. This prediction is based on the findings that no histone

acetylation was found surrounding the WNT5A promoters A or B through the H3K27ac track on ENCODE and therefore may be undergoing histone deacetylation. We saw a response to treatment with SAHA alone and in combination with 5-Aza at promoter A, whereas promoter B displayed a fold-increase only when SAHA was used in combination with 5-Aza. Treatment with both 1μM and 5μM SAHA, alone, increased expression from promoter A in the range of that seen with 5-Aza treatment. When SAHA was combined with 5-Aza, expression was increased even higher to approximately 10 to 20 times higher than either compound alone. These finding suggest that, indeed, histone modifications plays some role in the epigenetic state of promoter A in HCT-116 cells.

In contrast to promoter A, promoter B is less responsive to SAHA; only in combination with 5-Aza was there a substantial increase in promoter activity (38-42 fold). This finding illustrates that both promoters are responsive to SAHA. This suggests that normally this region is being deacetylated by HDACs. One approach to clarify this issue would be to perform ChIP analysis on both promoters A and B for a H3K27ac.

In summary, we used the two distinct promoters of WNT5A as “novel genes” to investigate if the pattern of re-activation of these promoters by the epigenetic drugs 5-Aza was found to be higher in cell lines with greater methylation as a means to test our model from Chapter II. WNT5A promoter A and another selected gene MYL9 were found to be somewhat consistent with the model. Surprisingly, promoter B was unaffected by 5-Aza but this lack of response was identical in both HCT-116 and HepG2, which have identical epigenetic profiles. SAHA was found to have a strong positive effect on expression, especially in combination with 5-Aza. This is what was expected, as the epigenetic profiles were void of histone marks.

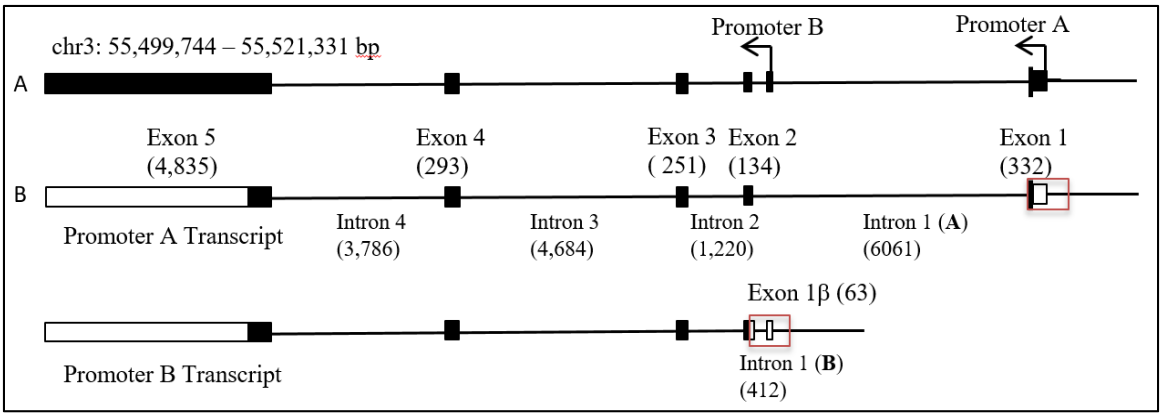


Figure 7. WNT5A Alternative Isoforms. WNT5A isoforms. (A) Genomic structure of the WNT5A gene. Arrow indicates transcription start sites of promoter A and Promoter B. (B) Primary transcripts for Promoter A (top) and Promoter B (bottom). Red boxes indicate 1000bp region of investigation centered at the TSS. Exons represented by filled boxes and introns, illustrated by lines to the left of the TSS display their specific size in parentheses.

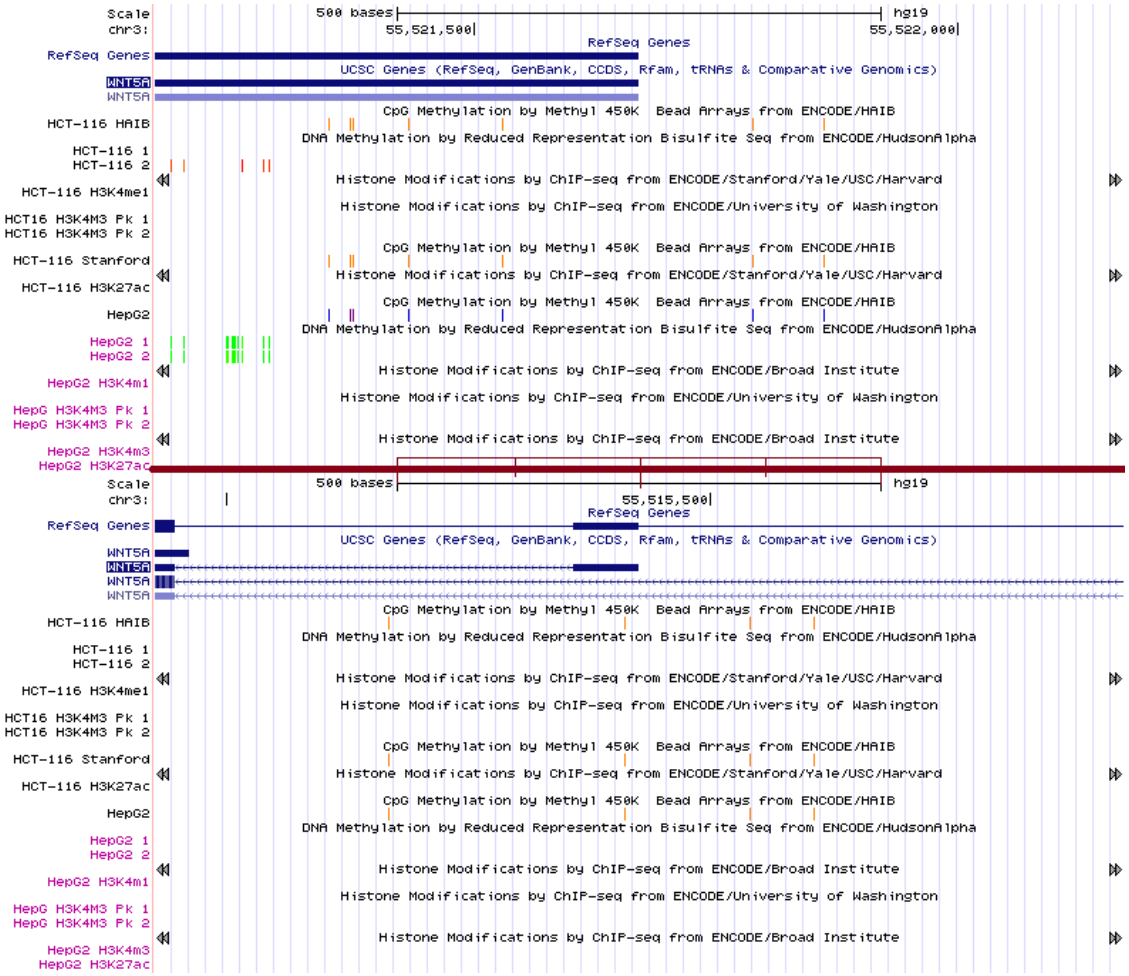


Figure 8. ENCODE Snapshot of *WNT5A* Promoter A and Promoter B. ENCODE snapshot showing 500bp flanking regions of the WNT5A promoter A transcription start site (top) and WNT5A promoter B transcription start site (bottom) methylation and histone modifications. DNA methylation in RRBS tracks is shown in either red (100% methylated), yellow (50% methylated), or green (0% methylated). Representation of DNA methylation as found via 450K array is shown in either orange (methylated), purple (partially methylated), or blue (un methylated). The red bar in the center is helps to identify location, with the center being at the transcription start site and a 250bp region blocked off on each side (where then thin red line ends); these regions are divided in half representing (2) 125bp regions. Gene is on the negative strand.

Figure 9. Visualization of the WNT5A Methylation Sites in the Sequence. The sequence of WNT5A at Promoter A (A) and Promoter B (B) 500bp upstream and downstream from the transcription start site indicated by blue arrow. Exons are illustrated in upper case, introns in lower case. All CpGs within the sequence have been boxed in. Shaded CpGs represent sites where ENCODE found methylation at ≥50% (RRBS) or at least partially methylated (450K) in HCT-116 and HepG2. The blue, boxed in area represents the first exon of each promoter.

Table 5. Count of DNA Methylation and Histone Modifications of WNT5A Promoters A and B. A count of the DNA methylation and list of the histone modifications present in each 250bp region of the 1000bp window of investigation for the two novel genes (WNT5A Promoter A and WNT5A Promoter B).

Count of DNA Methylation and Histone Modifications of WNT5A Promoters A and B

GENE SYMBOL	Cell Line	-500		-250		+250		+500	
		Histones	DNAm	Histones	DNAm	Histones	DNAm	Histones	DNAm
WNT5A Prom A	HCT-116	N/A	0	N/A	2	N/A	2	N/A	8
	HepG2	N/A	0	N/A	0	N/A	0	N/A	2
WNT5A Prom B	HCT-116	N/A	0	N/A	3	N/A	1	N/A	0
	HepG2	N/A	0	N/A	3	N/A	1	N/A	0

Table 6. qPCR VS Microarray Results for WNT5A Promoters A and B. Relative fold change as determined by qPCR in the two "novel genes" (WNT5A promoter A and promoter B. Difference (experimental) is the difference between the two cell lines for a given gene.

qPCR VS Microarray Results for WNT5A Promoters A and B

Gene Name	Cell Line	Fold Change (Experimental)	HCT116 : HepG2	HepG2 : HCT116
WNT5A	HCT116	6.16	1.58	
Prom A	HepG2	3.9		0.63
WNT5A	HCT116	1.09	0.11	
Prom B	HepG2	0.98		0.9

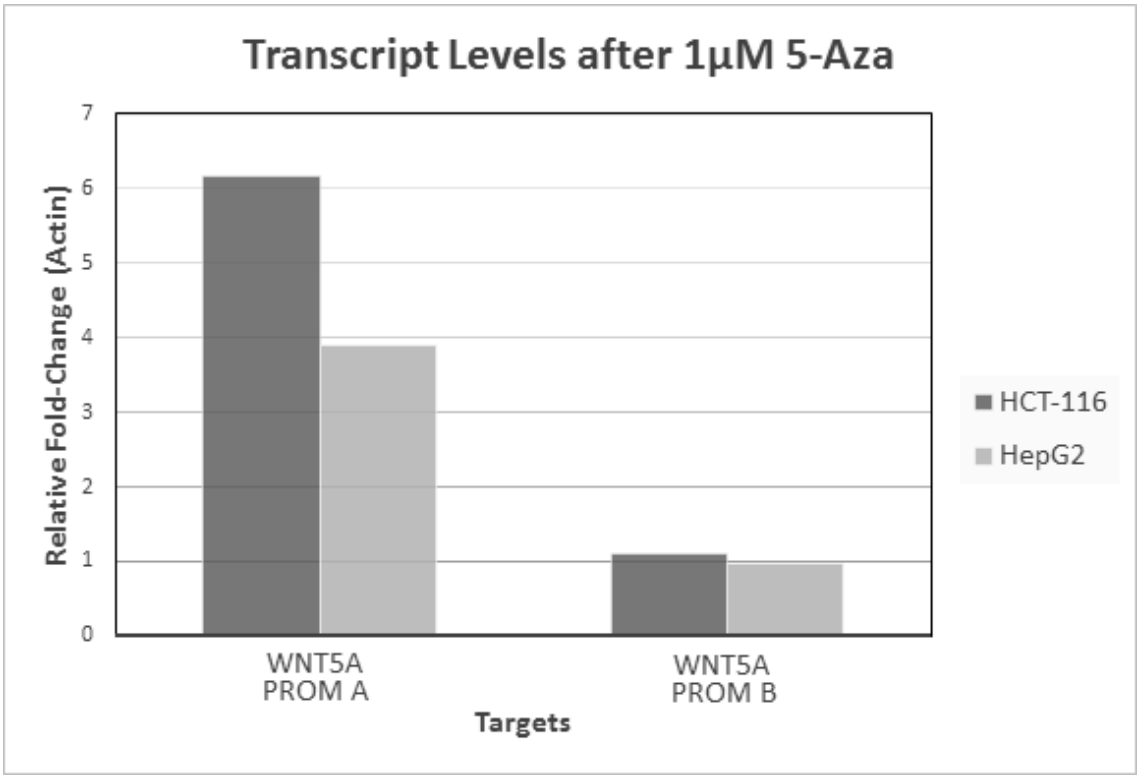
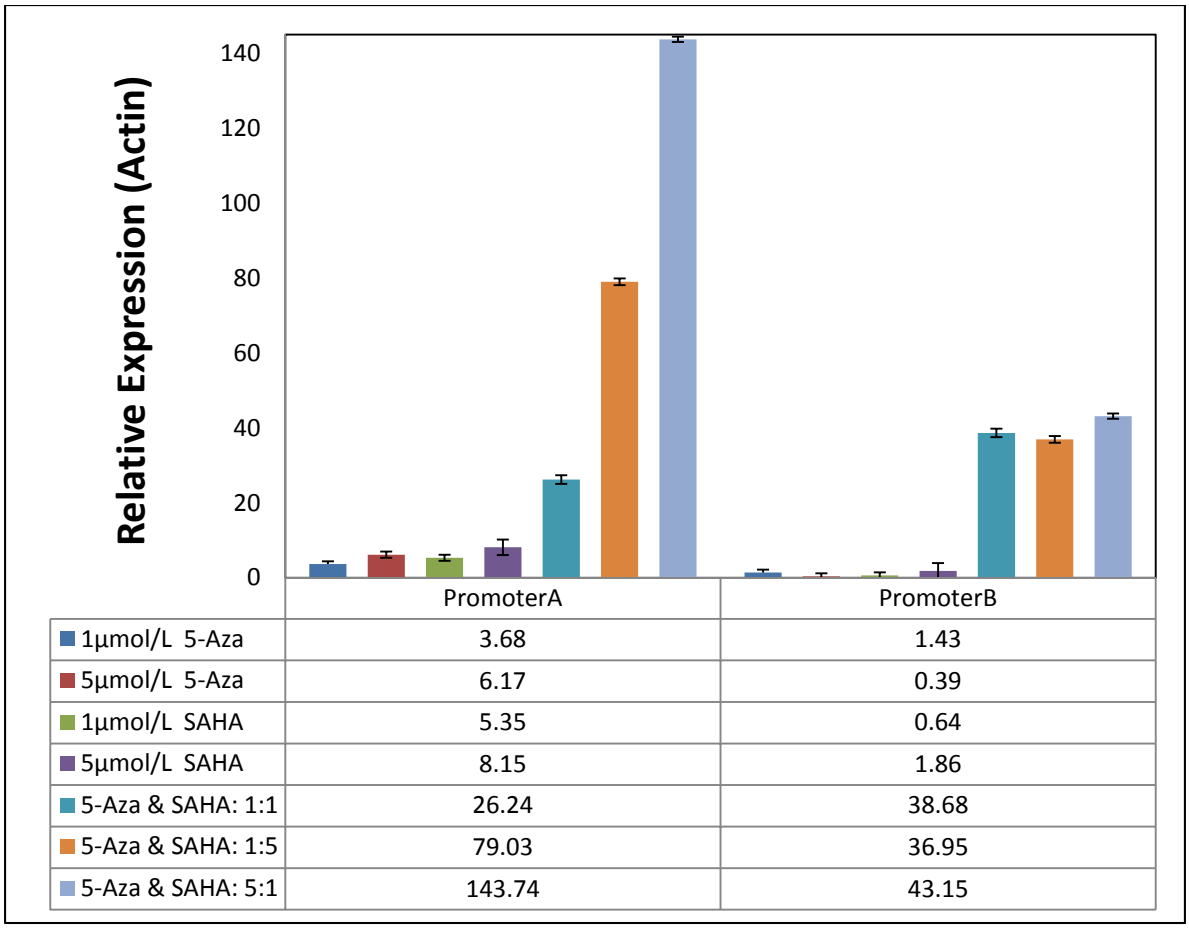


Figure 10. Graph of Relative Fold-Change in WNT5A Promoters A and B. Reactivation of novel genes WNT5A promoter A and promoter B after treatment with 5-Azacytidine in HCT-116 and HepG2.



	PromoterA	PromoterB
1µmol/L 5-Aza	3.68	1.43
5µmol/L 5-Aza	6.17	0.39
1µmol/L SAHA	5.35	0.64
5µmol/L SAHA	8.15	1.86
5-Aza & SAHA: 1:1	26.24	38.68
5-Aza & SAHA: 1:5	79.03	36.95
5-Aza & SAHA: 5:1	143.74	43.15

Figure 11. Graph of Relative Fold-Change in WNT5A Promoters A and B Using Combinatorial Treatment. Relative fold-change of WNT5A promoter A and B transcript level expression after treatment with 5-Azacytidine (5-Aza) and SAHA. HCT-116 cells were treated with the indicated amounts of 5-Aza and SAHA as described in Materials and Methods. The levels of promoter A and promoter B transcripts determined by qPCR and relative levels determined based on actin transcripts. Error bars represent standard deviation.

REFERENCES

Adams, J. M., & Cory, S. (2010). NIH Public Access. *Genetics*, *26*(9), 1324–1337. http://doi.org/10.1038/sj.onc.1210220

Aday, A. W., Zhu, L. J., Lakshmanan, A., Wang, J., & Lawson, N. D. (2011). Identification of cis regulatory features in the embryonic zebrafish genome through large-scale profiling of H3K4me1 and H3K4me3 binding sites. *Developmental Biology*, *357*(2), 450–462. http://doi.org/10.1016/j.ydbio.2011.03.007

Agarwal, P., Alzrigat, M., Párraga, A. A., Enroth, S., Singh, U., Ungerstedt, J., … Jernberg-wiklund, H. (2016). Genome-wide profiling of histone H3 lysine 27 and lysine 4 trimethylation in multiple myeloma reveals the importance of Polycomb gene targeting and highlights EZH2 as a potential therapeutic target, *7*(6). http://doi.org/10.18632/oncotarget.6843

Angelica, M. D., & Fong, Y. (2008). NIH Public Access. *J Am Chem Soc*, *141*(4), 520–529. http://doi.org/10.1016/j.surg.2006.10.010.Use

Angst, E., Dawson, D. W., Nguyen, A., Park, J., Go, V. L. W., Reber, H. A., … Eibl, G. (2011). NIH Public Access, *39*(5), 675–679. http://doi.org/10.1097/MPA.0b013e3181c8b476.Epigenetic

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., … Soboleva, A. (2013). NCBI GEO: Archive for functional genomics data sets - Update. *Nucleic Acids Research*, *41*(D1), 991–995. http://doi.org/10.1093/nar/gks1193

Bauer, M., Bénard, J., Gaasterland, T., Willert, K., & Cappellen, D. (2013). WNT5A Encodes Two Isoforms with Distinct Functions in Cancers. *PLoS ONE*, *8*(11), 1–14. http://doi.org/10.1371/journal.pone.0080526

Berx, G., & van Roy, F. (2009). Involvement of members of the cadherin superfamily in cancer. *Cold Spring Harbor Perspectives in Biology*, *1*(6). http://doi.org/10.1101/cshperspect.a003129

Blackledge, N. P., Zhou, J. C., Tolstorukov, M. Y., Farcas, A. M., Park, P. J., & Klose, R. J. (2010). CpG Islands Recruit a Histone H3 Lysine 36 Demethylase. *Molecular Cell*, *38*, 179–190. http://doi.org/10.1016/j.molcel.2010.04.009

Blattler, A., Yao, L., Witt, H., Guo, Y., Nicolet, C. M., Berman, B. P., & Farnham, P. J. (2014). Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes. *Genome Biology*, *15*(9), 469. http://doi.org/10.1186/s13059-014-0469-0

Butcher, L. M., & Beck, S. (2015). Probe Lasso: A novel method to rope in differentially methylated regions with 450K DNA methylation data. *Methods*, *72*(C), 21–28. http://doi.org/10.1016/j.ymeth.2014.10.036

Cai, M., Gao, F., Zhang, P., An, W., Shi, J., Wang, K., & Lu, W. (2015). Analysis of a transgenic Oct4 enhancer reveals high fidelity long-range chromosomal interactions. *Scientific Reports*, *5*, 14558. http://doi.org/10.1038/srep14558

Cai, M., Kim, S., Wang, K., Farnham, P. J., Coetzee, G. A., & Lu, W. (2016). 4C-seq revealed long-range interactions of a functional enhancer at the 8q24 prostate cancer risk locus. *Scientific Reports*, *6*(February), 22462. http://doi.org/10.1038/srep22462

Chae, Y. K., Anker, J. F., Carneiro, B. A., Platanias, C., & Giles, F. J. (2016). Genomic landscape of DNA repair genes in cancer. *Oncotarget*.

Chen, M., Liao, W. S., Lu, Z., Bornmann, W. G., Hennessey, V., Washington, M. N., … Jr, R. C. B. (2012). NIH Public Access, *117*(19), 4424–4438. http://doi.org/10.1002/cncr.26073.Decitabine

Cheng, J. C. et al. (2004). preferential response of cancer cells to zebularine. *Cancer Cell*, *6*(2), 151–158.

Cheng, N., Chytil, A., Shyr, Y., Joly, A., & Moses, H. L. (2008). TGF-β signaling deficient fibroblasts enhance Hepatocyte Growth Factor signaling in mammary carcinoma cells to promote scattering and invasion. *Mol Cancer Res.*, *6*(10), 1521–1533. http://doi.org/10.1158/1541-7786.MCR-07-2203.TGF-

Cho, H., Mariotto, A. B., Schwartz, L. M., Luo, J., & Woloshin, S. (2014). When do changes in cancer survival mean progress? The insight from population incidence and mortality. *Journal of the National Cancer Institute - Monographs*, *2014*(49), 187–197. http://doi.org/10.1093/jncimonographs/lgu014

Coppedè, F., Lopomo, A., Spisni, R., & Migliore, L. (2014). Genetic and epigenetic biomarkers for diagnosis, prognosis and treatment of colorectal cancer. *World Journal of Gastroenterology : WJG*, *20*(4), 943–56. http://doi.org/10.3748/wjg.v20.i4.943

Couldrey, C., Brauning, R., Bracegirdle, J., Maclean, P., Henderson, H. V., & McEwan, J. C. (2014). Genome-wide DNA methylation patterns and transcription analysis in sheep muscle. *PLoS ONE*, *9*(7), 1–7. http://doi.org/10.1371/journal.pone.0101853

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., … Jaenisch, R. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(50), 21931–21936. http://doi.org/10.1073/pnas.1016071107

Dannenberg, L. O., & Edenberg, H. J. (2006). Epigenetics of gene expression in human hepatoma cells: expression profiling the response to inhibition of DNA methylation and histone deacetylation. *BMC Genomics*, *7*, 181. http://doi.org/10.1186/1471-2164-7-181

Decock, A., Ongenaert, M., Hoebeeck, J., De Preter, K., Van Peer, G., Van Criekinge, W., … Vandesompele, J. (2012). Genome-wide promoter methylation analysis in neuroblastoma identifies prognostic methylation biomarkers. *Genome Biology*, *13*(10), R95. http://doi.org/10.1186/gb-2012-13-10-r95

Drake, K. M., Watson, V. G., Kisielewski, A., Glynn, R., & Napper, A. D. (2014). A sensitive luminescent assay for the histone methyltransferase NSD1 and other SAM-dependent enzymes. *Assay and Drug Development Technologies*, *12*(5), 258–271. http://doi.org/10.1089/adt.2014.583

Edgar, R., Domrachev, M., & Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research*, *30*(1), 207–10. http://doi.org/10.1093/nar/30.1.207

Esteller, M. (2002). CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. *Oncogene*, *21*(35), 5427–40. http://doi.org/10.1038/sj.onc.1205600

Ferraiuolo, M., Di Agostino, S., Blandino, G., & Strano, S. (2016). Oncogenic Intra-p53 Family Member Interactions in Human Cancers. *Frontiers in Oncology*, *6*(March), 1–10. http://doi.org/10.3389/fonc.2016.00077

Follo, M. Y., Finelli, C., Mongiorgi, S., Clissa, C., Bosi, C., Testoni, N., … Cocco, L. (2009). Reduction of phosphoinositide-phospholipase C beta1 methylation predicts the responsiveness to azacitidine in high-risk MDS. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(39), 16811–16816. http://doi.org/10.1073/pnas.0907109106

Gravina, G. L., Festuccia, C., Marampon, F., Popov, V. M., Pestell, R. G., Zani, B. M., & Tombolini, V. (2010). Biological rationale for the use of DNA methyltransferase inhibitors as new strategy for modulation of tumor response to chemotherapy and radiation. *Molecular Cancer*, *9*(1), 305. http://doi.org/10.1186/1476-4598-9-305

Hanahan, D., & Weinberg, R. a. (2011). Hallmarks of cancer: the next generation. *Cell*, *144*(5), 646–74. http://doi.org/10.1016/j.cell.2011.02.013

He, S., Barron, E., Ishikawa, K., Nazari Khanamiri, H., Spee, C., Zhou, P., … Hinton, D. R. (2015). Inhibition of DNA Methylation and Methyl-CpG-Binding Protein 2 Suppresses RPE Transdifferentiation: Relevance to Proliferative Vitreoretinopathy. *Investigative Opthalmology & Visual Science*, *56*(9), 5579. http://doi.org/10.1167/iovs.14-16258

Heerboth, S., Lapinska, K., Snyder, N., Leary, M., & Rollinson, S. (2014). Use of Epigenetic Drugs in Disease : An Overview. *Genetics & Epigenetics*, 9–19. http://doi.org/10.4137/GEG.S12270.RECEIVED

Heller, G., Schmidt, W. M., Ziegler, B., Holzer, S., Müllauer, L., Bilban, M., … Zöchbauer-Müller, S. (2008). Genome-wide transcriptional response to 5-aza-2'-deoxycytidine and trichostatin a in multiple myeloma cells. *Cancer Research*, *68*(1), 44–54. http://doi.org/10.1158/0008-5472.CAN-07-2531

Herman, J. G., & Baylin, S. B. (2003). Gene silencing in cancer in association with promoter hypermethylation. *The New England Journal of Medicine*, *349*, 2042–2054. http://doi.org/10.1056/NEJMra023075

Huichalaf, C., Micheloni, S., Ferri, G., Caccia, R., & Gabellini, D. (2014). DNA methylation analysis of the macrosatellite repeat associated with FSHD muscular dystrophy at single nucleotide level. *PLoS ONE*, *9*(12), 1–24. http://doi.org/10.1371/journal.pone.0115278

Hung, T.-H., Hsu, S.-C., Cheng, C.-Y., Choo, K.-B., Tseng, C.-P., Chen, T.-C., … Chong, K.-Y. (2014). Wnt5A regulates ABCB1 expression in multidrug-resistant cancer cells through activation of the non-canonical PKA/β-catenin pathway. *Oncotarget*, *5*(23), 12273–90. Retrieved from

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4322984&tool=pmcentrez&rendertype=abstract

Jacques, C., Lamoureux, F., Baud, M., Rodriguez, L., Quillard, T., Amiaud, J., … Rédini, F. (2016). Targeting the epigenetic readers in Ewing Sarcoma inhibits the oncogenic transcription factor EWS / Fli1. *Oncotarget*.

Jiang, D., Dumur, C. I., Massey, H. D., Ramakrishnan, V., Subler, M. A., & Windle, J. J. (2015). Comparison of effects of p53 null and gain-of-function mutations on salivary tumors in MMTV-Hras transgenic mice. *PLoS ONE*, *10*(2), 1–23. http://doi.org/10.1371/journal.pone.0118029

Jiang, N., Wang, L., Chen, J., Wang, L., Leach, L., & Luo, Z. (2014). Conserved and Divergent Patterns of DNA Methylation in Higher Vertebrates. *Genome Biology and Evolution*, *2998*, 2998–3014. http://doi.org/10.1093/gbe/evu238

Khanahmadi, M., Farhud, D. D., & Malmir, M. (2015). Genetic of Alzheimer ' s Disease : A Narrative Review Article. *Iran J Public Health*, *44*(7), 892–901.

Kikuchi, A., Yamamoto, H., Sato, A., & Matsumoto, S. (2012). Wnt5a: Its signalling, functions and implication in diseases. *Acta Physiologica*, *204*(1), 17–33. http://doi.org/10.1111/j.1748-1716.2011.02294.x

Kolesnikov, N., Hastings, E., Keays, M., Melnichuk, O., Tang, Y. A., Williams, E., … Brazma, A. (2015). ArrayExpress update-simplifying data submissions. *Nucleic Acids Research*, *43*(D1), D1113–D1116. http://doi.org/10.1093/nar/gku1057

Kumarakulasinghe, N. B., van Zanwijk, N., & Soo, R. A. (2015). Molecular targeted therapy in the treatment of advanced stage non-small cell lung cancer (NSCLC). *Respirology (Carlton, Vic.)*, *20*(3), 370–8. http://doi.org/10.1111/resp.12490

Laird, P. W., & PE Lange, C. (2013). Clinical applications of DNA methylation biomarkers in colorectal cancer. *Epigenomics*, *5*(2), 105–108. http://doi.org/10.2217/epi.13.4

Lao, V. V., & Grady, W. M. (2011). Epigenetics and Colorectal Cancer. *Nature Reviews Gastroenterology and Hepatology*, *8*, 686–700.

Leclercq, S., Gueugnon, F., Boutin, B., Guillot, F., Blanquart, C., Rogel, A., … Gr??goire, M. (2011). A 5-aza-2???-deoxycytidine/valproate combination induces cytotoxic T-cell response against mesothelioma. *European Respiratory Journal*, *38*(5), 1105–1116. http://doi.org/10.1183/09031936.00081310

Li, H., Chiappinelli, K. B., Guzzetta, A. a, Easwaran, H., Yen, R.-W. C., Vatapalli, R., … Ahuja, N. (2014). Immune regulation by low doses of the DNA methyltransferase inhibitor 5-azacitidine in common human epithelial cancers. *Oncotarget*, *5*(3), 587–98. http://doi.org/24583822

Liang, Y., Liu, Y., Hou, B., Zhang, W., Liu, M., Sun, Y.-E., … Gu, X. (2016). CREB-regulated transcription coactivator 1 enhances CREB-dependent gene expression in spinal cord to maintain the bone cancer pain in mice. *Molecular Pain*, *12*(0), 1–11. http://doi.org/10.1177/1744806916641679

Linghu, C., Zheng, H., Zhang, L., & Zhang, J. (2013). Discovering common combinatorial histone modification patterns in the human genome. *Gene*, *518*(1), 171–178. http://doi.org/10.1016/j.gene.2012.11.038

Lyko, F., & Brown, R. (2005). DNA methyltransferase inhibitors and the development of epigenetic cancer therapies. *Journal of the National Cancer Institute*, *97*(20), 1498–506. http://doi.org/10.1093/jnci/dji311

Marks, P. a. (2007). Discovery and development of SAHA as an anticancer agent. *Oncogene*, *26*(9), 1351–1356. http://doi.org/10.1038/sj.onc.1210204

Medvedeva, Y. A., Khamis, A. M., Kulakovskiy, I. V, Ba-Alawi, W., Bhuyan, M. S. I., Kawaji, H., … Bajic, V. B. (2014). Effects of cytosine methylation on transcription factor binding sites. *BMC Genomics*, *15*(1), 119. http://doi.org/10.1186/1471-2164-15-119

Meldi, K., Qin, T., Buchi, F., Droin, N., Sotzen, J., Micol, J. B., … Figueroa, M. E. (2015). Specific molecular signatures predict decitabine response in chronic myelomonocytic leukemia. *Journal of Clinical Investigation*, *125*(5), 1857–1872. http://doi.org/10.1172/JCI78752

Miranda, T. B., Cortez, C. C., Yoo, C. B., Liang, G., Abe, M., Kelly, T. K., … Jones, P. A. (2009). NIH Public Access. *Molecular Cancer Therapeutics*, *8*(6), 1579–1588. http://doi.org/10.1158/1535-7163.MCT-09-0013.DZNep

Miyazono, K., Ehata, S., & Koinuma, D. (2012). Tumor-promoting functions of transforming growth factor-β in progression of cancer. *Upsala Journal of Medical Sciences*, *117*(2), 143–52. http://doi.org/10.3109/03009734.2011.638729

Myers, R. M., Stamatoyannopoulos, J., Snyder, M., Dunham, I., Hardison, R. C., Bernstein, B. E., … Good, P. J. (2011). A user's guide to the Encyclopedia of DNA elements (ENCODE). *PLoS Biology*, *9*(4). http://doi.org/10.1371/journal.pbio.1001046

Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N., & Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, *393*(6683), 386–389. http://doi.org/10.1038/30764

Pan, H., Chen, L., Dogra, S., Teh, A. L., Tan, J. H., Lim, Y. I., … Holbrook, J. D. (2012). Measuring the methylome in clinical samples: Improved processing of the infinium human methylation450 beadchip array. *Epigenetics*, *7*(10), 1–15. http://doi.org/10.4161/epi.22102

Pekowska, A., Benoukraf, T., Zacarias-Cabeza, J., Belhocine, M., Koch, F., Holota, H., … Spicuglia, S. (2011). H3K4 tri-methylation provides an epigenetic signature of active enhancers. *The EMBO Journal*, *30*(20), 4198–4210. http://doi.org/10.1038/emboj.2011.295

Pérez-alea, M., Vivancos, A., Caratú, G., Matito, J., Ferrer, B., Hernandez-losa, J., … Garcia-patos, V. (2016). Genetic profile of GNAQ -mutated blue melanocytic neoplasms reveals mutations in genes linked to genomic instability and the PI3K pathway. *Oncotarget*.

Peyser, N. D., Du, Y., Li, H., Lui, V., Xiao, X., Chan, T. A., & Grandis, J. R. (2015). Loss-of-function PTPRD mutations lead to increased STAT3 activation and sensitivity to STAT3 inhibition in head and neck cancer. *PLoS ONE*, *10*(8), 1–13. http://doi.org/10.1371/journal.pone.0135750

Raj, K., & Mufti, G. J. (2006). Azacytidine (Vidaza(R)) in the treatment of myelodysplastic syndromes. *Therapeutics and Clinical Risk Management*, *2*(4), 377–388. http://doi.org/10.2147/tcrm.2006.2.4.377

Rawat, M., Bhosale, M. A., & Karmodiya, K. (2016). Plasmodium falciparum epigenome: A distinct dynamic epigenetic regulation of gene expression. *Genomics Data*, *7*(28908), 79–81. http://doi.org/10.1016/j.gdata.2015.11.026

Rhee, I., Bachman, K. E., Park, B. H., & Jair, K. (2002). DNMT1 and DNMT3b cooperate to silence genes in human cancer cells, *416*(April).

Robert, M. F., Morin, S., Beaulieu, N., Gauthier, F., Chute, I. C., Barsalou, A., & MacLeod, A. R. (2003). DNMT1 is required to maintain CpG methylation and aberrant gene silencing in human cancer cells. *Nat Genet*, *33*(1), 61–65. http://doi.org/10.1038/ng1068

Robertson, K. D., & Jones, P. A. (2000). DNA methylation: past, present and future directions. *Carcinogenesis*, *21*(3), 461–467. http://doi.org/10.1093/carcin/21.3.461

Rose, N. R., & Klose, R. J. (2014). Biochimica et Biophysica Acta Understanding the relationship between DNA methylation and histone lysine methylation ☆. *BBA - Gene Regulatory Mechanisms*. http://doi.org/10.1016/j.bbagrm.2014.02.007

Rothbart, S. B., & Strahl, B. D. (2014). Interpreting the language of histone and DNA modifications. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, *1839*(8), 627–643. http://doi.org/10.1016/j.bbagrm.2014.03.001

Ryerson, A. B., Eheman, C. R., Altekruse, S. F., Ward, J. W., Jemal, A., Sherman, R. L., … Kohler, B. A. (2016). Annual Report to the Nation on the Status of Cancer, 1975-2012, featuring

the increasing incidence of liver cancer. *Cancer*, n/a–n/a. http://doi.org/10.1002/cncr.29936

Shakeri, H., Gharesouran, J., Fakhrjou, A., & Esfahani, A. (2016). Original article : DNA Methylation assessment as a prognostic factor in invasive breast cancer using methylation specific multiplex ligation dependent probe amplification. *EXCLI*, *15*(Idc), 11–20.

Shay, J. W., & Wright, W. E. (2000). Hayflick, his limit, and cellular ageing. *Nature Reviews. Molecular Cell Biology*, *1*(October), 72–76. http://doi.org/10.1038/35036093

Shen, L., Kondo, Y., Ahmed, S., Boumber, Y., Konishi, K., Guo, Y., … Issa, J. P. J. (2007). Drug sensitivity prediction by CpG island methylation profile in the NCI-60 cancer cell line panel. *Cancer Research*, *67*(23), 11335–11343. http://doi.org/10.1158/0008-5472.CAN-07-1502

Shi, B. (2013). The development and potential clinical utility of biomarkers for HDAC inhibitors. *Drug Discoveries & Therapeutics*, *7*(4), 129–136. http://doi.org/10.5582/ddt.2013.v7.4.129

Shi, G., Yoshida, Y., Yuki, K., Nishimura, T., & Kawata, Y. (2016). Pattern of RECK CpG methylation as a potential marker for predicting breast cancer prognosis and drug-sensitivity.

Siegel, R., Miller, K., & Jemal, A. (2015). Cancer statistics , 2015 . *CA Cancer J Clin*, *65*(1), 29. http://doi.org/10.3322/caac.21254.

Siegel, R., Miller, K., & Jemal, A. (2016). Cancer Statistics, 2016. *CA Cancer J Clin*, *66*(1), 7–30. http://doi.org/10.3322/caac.21332.

Stresemann, C., & Lyko, F. (2008). Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine. *International Journal of Cancer. Journal International Du Cancer*, *123*(1), 8–13. http://doi.org/10.1002/ijc.23607

Takeshima, H., Wakabayashi, M., Hattori, N., Yamashita, S., & Ushijima, T. (2014). Identification of coexistence of DNA methylation and H3K27me3 specifically in cancer cells as a promising target for epigenetic therapy. *Carcinogenesis*, *36*(2), 192–201. http://doi.org/10.1093/carcin/bgu238

Tamaru, H., & Selker, E. U. (2001). A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. *Nature*, *414*, 277–283. http://doi.org/10.1038/35104508

Thomson, J. P., Skene, P. J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., … Bird, A. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature*, *464*, 1082–1086. http://doi.org/10.1038/nature08924

Treppendahl, M. B., Kristensen, L. S., & Grønbæk, K. (2014). Review series Predicting response to epigenetic therapy, *124*(1), 47–55. http://doi.org/10.1172/JCI69737.It

Tsai, C. C., Su, P. F., Huang, Y. F., Yew, T. L., & Hung, S. C. (2012). Oct4 and Nanog Directly Regulate Dnmt1 to Maintain Self-Renewal and Undifferentiated State in Mesenchymal Stem Cells. *Molecular Cell*, *47*(2), 169–182. http://doi.org/10.1016/j.molcel.2012.06.020

Tusher, V. G., Tibshirani, R., & Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(9), 5116–21. http://doi.org/10.1073/pnas.091062498

Ucar, D., Hu, Q., & Tan, K. (2011). Combinatorial chromatin modification patterns in the human genome revealed by subspace clustering. *Nucleic Acids Research*, *39*(10), 4063–4075. http://doi.org/10.1093/nar/gkr016

Vaidya, H., Rumph, C., & Katula, K. S. (2016). Inactivation of the WNT5A Alternative Promoter B Is Associated with DNA Methylation and Histone Modification in Osteosarcoma Cell Lines U2OS and SaOS-2. *Plos One*, *11*(3), e0151392. http://doi.org/10.1371/journal.pone.0151392

Vasilatos, S. N., Katz, T. A., Oesterreich, S., Wan, Y., Davidson, N. E., & Huang, Y. (2013). Crosstalk between lysine-specific demethylase 1 (LSD1) and histone deacetylases mediates antineoplastic efficacy of HDAC inhibitors in human breast cancer cells. *Carcinogenesis*, *34*(6), 1196–1207. http://doi.org/10.1093/carcin/bgt033

Voso, M. T., Fabiani, E., Piciocchi, a, Matteucci, C., Brandimarte, L., Finelli, C., … Leone, G. (2011). Role of BCL2L10 methylation and TET2 mutations in higher risk myelodysplastic syndromes treated with 5-azacytidine. *Leukemia*, *25*(12), 1910–3. http://doi.org/10.1038/leu.2011.170

Wang, Z., Zang, C., Rosenfeld, J. a, Schones, D. E., Cuddapah, S., Cui, K., … Michael, Q. (2009). NIH Public Access, *40*(7), 897–903. http://doi.org/10.1038/ng.154.Combinatorial

Weisenberger, D. J. (2014). Review series Characterizing DNA methylation alterations from The Cancer Genome Atlas, *124*(1). http://doi.org/10.1172/JCI69740.domains

WHO. (2014). Global status report on noncommunicable diseases 2014. *World Health*, 176. http://doi.org/ISBN 9789241564854

Wicki, A., & Christofori, G. (2008). The angiogenic switch in tumorigenesis. *Tumor Angiogenesis: Basic Mechanisms and Cancer Therapy*, *19*, 67–88. http://doi.org/10.1007/978-3-540-33177-3_4

Wozniak, G. G., & Strahl, B. D. (2014). Hitting the " mark " : Interpreting lysine methylation in the context of active transcription. *BBA - Gene Regulatory Mechanisms*, 2–10. http://doi.org/10.1016/j.bbagrm.2014.03.002

Xu, J., Watts, J. A., Pope, S. D., Gadue, P., Kamps, M., Plath, K., … Smale, S. T. (2009). Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells. *Genes and Development*, *23*(24), 2824–2838. http://doi.org/10.1101/gad.1861209

Xu, W. S., Parmigiani, R. B., & Marks, P. (2007). Histone deacetylase inhibitors: molecular mechanisms of action. *Oncogene*, *26*, 5541–5552. http://doi.org/10.1038/sj.onc.1210620

Ying, J., Li, H., Yu, J., Ka, M. N., Fan, F. P., Wong, S. C. C., … Tao, Q. (2008). WNT5A exhibits tumor-suppressive activity through antagonizing the Wnt/??-catenin signaling, and is frequently methylated in colorectal cancer. *Clinical Cancer Research*, *14*(1), 55–61. http://doi.org/10.1158/1078-0432.CCR-07-1644

Yoon, S., & Eom, G. H. (2016). HDAC and HDAC Inhibitor : From Cancer to Cardiovascular Diseases, 1–11. http://doi.org/10.4068/cmj.2016.52.1.1

Yu, Y., Zhang, M., Zhang, X., Cai, Q., Zhu, Z., Jiang, W., & Xu, C. (2014). Transactivation of epidermal growth factor receptor through platelet-activating factor/receptor in ovarian cancer cells. *Journal of Experimental & Clinical Cancer Research : CR*, *33*(1), 85. http://doi.org/10.1186/s13046-014-0085-6

Zahnow, C. A., Topper, M., Stone, M., Murray-Stewart, T., Li, H., Baylin, S. B., & Casero, R. A. (2016). Inhibitors of DNA Methylation, Histone Deacetylation, and Histone Demethylation, *130*, 55–111. http://doi.org/10.1016/bs.acr.2016.01.007

Zhao, Q.-Y., Lei, P.-J., Zhang, X., Zheng, J.-Y., Wang, H.-Y., Zhao, J., … Wu, M. (2016). Global histone modification profiling reveals the epigenomic dynamics during malignant transformation in a four-stage breast cancer model. *Clinical Epigenetics*, *8*(1), 34. http://doi.org/10.1186/s13148-016-0201-x

Zouridis, H. H. (2012). Methylation subtypes and large-scale epigenetic alterations in gastric cancer. *Science Translational Medicine*, *4*(156), 156ra140–156ra140.