INVESTIGATING THE IMPACT OF INTRAGENIC DNA METHYLATION ON GENE EXPRESSION, AND THE CLINICAL IMPLICATIONS ON TUMOR CELLS AND ASSOCIATED STROMA

Michael McGuire

Follow this and additional works at: https://digitalcommons.library.tmc.edu/utgsbs_dissertations

Part of the Bioinformatics Commons, Medicine and Health Sciences Commons, Molecular Biology Commons, and the Molecular Genetics Commons

Recommended Citation
McGuire, Michael, "INVESTIGATING THE IMPACT OF INTRAGENIC DNA METHYLATION ON GENE EXPRESSION, AND THE CLINICAL IMPLICATIONS ON TUMOR CELLS AND ASSOCIATED STROMA" (2018). The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access). 852.
https://digitalcommons.library.tmc.edu/utgsbs_dissertations/852

This Dissertation (PhD) is brought to you for free and open access by the The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences at DigitalCommons@TMC. It has been accepted for inclusion in The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@TMC. For more information, please contact digitalcommons@library.tmc.edu.
INVESTIGATING THE IMPACT OF INTRAGENIC DNA METHYLATION ON GENE EXPRESSION, AND THE CLINICAL IMPLICATIONS ON TUMOR CELLS AND ASSOCIATED STROMA

by

Michael McGuire, B.S.

APPROVED:

Anil Sood, M.D.
Advisory Professor

Sharon Dent, Ph.D.

Keith Baggerly, Ph.D.

Menashe Bar-Eli, Ph.D.

Mien-Chie Hung, Ph.D.

APPROVED:

Dean, The University of Texas
MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences
INVESTIGATING THE IMPACT OF INTRAGENIC DNA METHYLATION ON GENE EXPRESSION, AND THE CLINICAL IMPLICATIONS ON TUMOR CELLS AND ASSOCIATED STROMA

A

DISSERTATION

Presented to the Faculty of

The University of Texas

MD Anderson Cancer Center UTHealth

Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

Michael McGuire, B.S.

Houston, Texas

May 19th, 2018
Dedication

I dedicate this dissertation to my family: My mother, Colleen, my sisters, Jenna and Laura, and my grandmothers, Janet McGuire and Linda Hefner. Without them, I never would have been able to get into graduate school, much less carry out this dissertation. I owe them greatly for all of their love and support.

My friends, of which are too numerous to name, who have always been there to share in my happiness, and lift my spirits when I was down.

All of the scientists who have laid the foundation for our knowledge in biology, and who currently work toward a greater understanding of the natural world.
Acknowledgements

First and foremost, I would like to acknowledge the immense opportunity provided to me by my mentor, Dr. Anil Sood. I have learned an immense amount from Dr. Sood, both within science, and professionally as well. His encyclopedic intelligence, and tireless work ethic, both served as an inspiration to me, and a goal for me to achieve in my life. I would also like to thank my committee members greatly for their guidance during the past 5 years. Dr. Keith Baggerly, Dr. Sharon Dent, Dr. Gary Gallick, and Dr. Mien-Chie Hung for steering me down the proper path. Additionally, I would like to extend my gratitude to Dr. Menashe Bar-Eli, who graciously took Dr. Gallick’s spot on my committee after Dr. Gallick became ill. Thank you as well to Dr. KK Wong, who has supported me in multiple ways, including replacing Dr. Hung for my defense. I also would like to thank Dr. Mong-Hong Lee and Dr. Eugenie Kleinerman, who allowed me to serve in their labs during my time as a rotation student. I learned quite a lot from them during that short time, and I am very grateful for these opportunities. I also have a very special thanks for Drs. Tony Gutschner and Monika Haemmerle, who were incredibly instrumental in assisting me with my project and with how to design experiments. Tony not only gave me the idea to make a Cas9-Tet1 fusion construct, but also helped me carry out the necessary experiments. I would also like to thank all of the lab members I have met during the Sood lab and who have given
me constructive criticism both within and outside of science. I have spent the last 5 years working in the Sood lab, and have had the opportunity to meet incredible people from diverse backgrounds. It has been extremely enriching. I would especially like to thank Dr. Sherry Wu, who is one of the best scientists I have come across, and whose support and advice were essential for finishing my Ph.D.
Abstract

INVESTIGATING THE IMPACT OF INTRAGENIC DNA METHYLATION ON GENE EXPRESSION, AND THE CLINICAL IMPLICATIONS ON TUMOR CELLS AND ASSOCIATED STROMA

Michael Hefner McGuire, B.S.

Advisory Professor: Anil Sood, M.D.

Investigations into the function of non-promoter DNA methylation have yielded new insights into epigenetic regulation of gene expression. Previous studies have highlighted the importance of distinguishing between DNA methylation in discrete functional regions; however, integrated non-promoter DNA methylation and gene expression analyses across a wide number of tumor types and corresponding normal tissues have not been performed. Through integrated analysis of gene expression and DNA methylation profiles, we uncovered an enrichment of DNA methylation sites within the gene body and 3'UTR in which DNA methylation is strongly positively correlated with gene expression. We examined 32 tumor types and identified 57 tumor suppressors and oncogenes out of 224 genes containing a correlation of > 0.5 between gene body methylation and gene expression in at least 1 tumor type. The lymphocyte-specific gene CARD11 exhibits robust association between gene body methylation and expression across 19 of 32 tumor types examined. It is significantly overexpressed in KIRC and LUAD, and has a z-score of 4 in KIRC, meaning that high expression of CARD11 in this tumor type was associated with lower patient overall survival. Contrary to its canonical function in lymphocyte NF-kB activation, CARD11 activates the mTOR pathway in KIRC and LUAD, resulting in suppressed autophagy,
and demethylation of a CpG island within the gene body of CARD11 decreases gene expression. In addition to methylation of the open reading frame portion of a gene, other regions of site-specific DNA methylation along the gene body remain to be explored. Upon segregating the gene body into discrete functional units (5'UTR, 1st exon, 3'UTR), it was noted that the 3'UTR contained an enrichment of probes positively correlated between DNA methylation and gene expression. In 5 of 10 tumor types examined, DNA methylation of the 3'UTR is associated with patient survival in a significant number of genes. Filtering for genes in which 3'UTR DNA methylation, relative to gene body DNA methylation, is more strongly correlated with gene expression yields a list of 156 genes, enriched for functions involving T cell activation. Activating T cells ex vivo caused the immune checkpoint gene HAVCR2, but not other genes examined, to show a substantial increase in 3'UTR DNA methylation, but not adjacent exonic/intronic, or promoter DNA methylation, upon upregulation of gene expression. Furthermore, this increase in HAVCR2 gene expression can be abrogated by treatment with demethylating agents. These findings implicate the 3'UTR as a functionally relevant DNA methylation site, particularly regarding T cell activity. Additionally, they reveal a novel mechanism by which HAVCR2 is upregulated in T cells, providing a new molecular target for immune checkpoint blockade.
## Table of Contents

Dedication .................................................................................................................................................. iii  
Acknowledgements ................................................................................................................................. iv  
Abstract ...................................................................................................................................................... vi  
Table of Contents ......................................................................................................................................... viii  
List of Figures ............................................................................................................................................... x  
List of Tables .............................................................................................................................................. xv  
Abbreviations ............................................................................................................................................. xvi  

### Section 1: Introduction

1.1 The molecular and cellular biology of DNA Methylation ................................................................. 1  
   1.1.1 Promoter Methylation .................................................................................................................... 8  
   1.1.2 Enhancer methylation .................................................................................................................. 9  
   1.1.3 Intergenic methylation ............................................................................................................... 12  
   1.1.4 Gene body methylation ............................................................................................................. 16  
1.2 DNA Methylation and disease .......................................................................................................... 20  
   1.2.1 Methylation in cancer .............................................................................................................. 22  
   1.2.2 Clinical strategies for targeting methylation ............................................................................... 26  
1.3 Non-coding RNA ............................................................................................................................... 28  
   1.3.1 Long non-coding RNA .............................................................................................................. 30  
   1.3.2 Natural antisense transcripts .................................................................................................... 31  
1.4 Immunology ......................................................................................................................................... 35  
   1.4.1 T cells .......................................................................................................................................... 38  
   1.4.2 Cancer Immunology .................................................................................................................. 40  
   1.4.3 Immune Checkpoint .................................................................................................................. 41  
1.5 Autophagy .......................................................................................................................................... 44  
   1.5.1 Autophagy and cancer .............................................................................................................. 47
1.6 Bioinformatics and TCGA ................................................................. 48

**Section 2: Materials and Methods** ................................................................. 49

Overarching hypothesis and specific aims ......................................................... 65

**Section 3: Identification of methylation sites correlated with gene expression, and investigating the function and regulation of CARD11 in cancer** .......... 66

**Section 4: Determining the relationship between 3’UTR methylation and gene expression** ........................................................................... 104

**Section 5: Discussion** ......................................................................................... 136

**Bibliography** ........................................................................................................ 144

**Vita** ....................................................................................................................... 190
List of Figures

Figure 1: Positional location of DNA methylation on cytosine nucleotide .... 3

Figure 2: Routes of active versus passive DNA demethylation ..................... 3

Figure 3: Mechanism of TET-mediated DNA demethylation ....................... 4

Figure 4: The various impacts of promoter methylation on gene expression .. 5

Figure 5: Types of histone modifications, and functional consequences ....... 6

Figure 6: Differential epigenetic modification according to cellular developmental state ................................................................. 7

Figure 7: The relationship between DNA methylation and histone modification .................................................................................................................. 8

Figure 8: The structure of the enhancer in transcriptional activation .......... 10

Figure 9: Common histone modifications in the enhancer region ............... 11

Figure 10: The types of transposons, and mechanisms of replication .......... 13

Figure 11: Methylation-sensitive regulation of insulator function ............... 15

Figure 12: The differential associations between histone modifications and gene body versus promoter DNA methylation ................................................. 17

Figure 13: CTCF-mediated regulation of CD45 splicing through differential gene body methylation and exon inclusion ................................................. 19

Figure 14: Tandem repeat expansion of CGG trinucleotide in Fragile X syndrome ................................................................................................................ 21

Figure 15: Overview of altered methylation profiles in cancer versus normal ......................................................................................................................... 23

Figure 16: Mechanism of action of the demethylating agent Azacytidine ...... 27
Figure 17: Types of non-coding RNAs

Figure 18: Overview of the various functions of long non-coding RNAs

Figure 19: Illustrations of the mechanisms of Natural Antisense Transcript function

Figure 20: Mechanism of action of Zeb2 Natural Antisense Transcript in regulating EMT through increased Zeb2 translation

Figure 21: Comparison of the innate versus adaptive immune system

Figure 22: Activation of T cells

Figure 23: The CTLA-4 pathway in T cell priming

Figure 24: The PD-1/PD-L1 pathway in T cell killing of tumor cells

Figure 25: Steps of the autophagic process

Figure 26: Heat map of cancer-related genes whose expression is positively correlated with gene body DNA methylation

Figure 27: Molecular mechanism of CARD11-mediated activation of downstream pathways

Figure 28: Normal vs tumor gene expression of CARD11 across 15 tissue types

Figure 29: Z-score of CARD11 expression across 33 tumor types

Figure 30: Kaplan-Meier curve of KIRC patient overall survival as a function of CARD11 gene expression

Figure 31: Knockdown of CARD11 in H1975 samples used for RPPA analysis
Figure 32: Heat map of change in protein amount and phosphorylation according to RPPA analysis .......................................................... 76

Figure 33: Cell panel of CARD11 protein expression in eight KIRC cell lines. 78

Figure 34: Cell cycle and proliferation assays after CARD11 knockdown ...... 80

Figure 35: Apoptosis and colony formation assays after CARD11 knockdown .................................................................................. 81

Figure 36: Colony formation assay after CARD11 overexpression.......... 83

Figure 37: Impact of overexpression of CARD11 on A498 colony formation.. 84

Figure 38: Lipidated LC3 protein levels after CARD11 knockdown ............. 85

Figure 39: The impact of bafilomycin treatment on lipidated LC3 in the context of CARD11 knockdown in UOK111 ............................................................. 86

Figure 40: The impact of bafilomycin treatment on lipidated LC3 in the context of CARD11 knockdown in H1975 ................................................................. 87

Figure 41: Live-cell imaging of GFP/RFP-LC3-expressing UOK111 cells after CARD11 knockdown .................................................................................... 89

Figure 42: Comparison of CARD11 gene body methylation between KIRC and normal kidney renal tissue.......................................................... 90

Figure 43: The impact of treating with decitabine on CARD11 expression ..... 91

Figure 44: CARD11 gene layout from ensembl .................................................. 92

Figure 45: Correlation between expression of CARD11 and CARD11-NAT in KIRC and LUAD ................................................................................... 93

Figure 46: Expression of CARD11-NAT across 12 tumor types using TCGA datasets ......................................................................................... 95
Figure 47: Location of \textit{CARD11} gene body CpG island, and expression of \textit{CARD11-NAT} as a function of gene body CpG island methylation .................. 98

Figure 48: The impact of \textit{CARD11-NAT} knockdown on gene body CpG island methylation ............................................................................................................ 99

Figure 49: Illustration of the Cas9-Tet1 fusion protein, and confirmation of protein expression by western blot............................................................................................................ 100

Figure 50: Expression of \textit{CARD11} and \textit{CARD11-NAT} after site-specific demethylation ........................................................................................................ 101

Figure 51: Expression of \textit{CARD11} after \textit{CARD11-NAT} overexpression in the context of site-specific demethylation ........................................................................................................ 102

Figure 52: Heat map of location of methylation probes with a strong negative correlation between methylation and gene expression .................. 105

Figure 53: Heat map of location of methylation probes with a strong positive correlation between methylation and gene expression .................. 106

Figure 54: Vector-based visualization of probes enriched in the 3'UTR that exhibit a positive correlation between methylation and gene expression .... 107

Figure 55: Dotplot of \textit{HOXC13} expression in normal versus cancer across 10 tissue types ........................................................................................................ 110

Figure 56: \textit{HOXC13} copy number alteration across 10 tissue types .......... 111

Figure 57: Heatmap of \textit{HOXC13} methylation patterns in relation to highest or lowest expressing BLCA and LUSC tumors, by location .................. 113

Figure 58: Differences in patient overall survival as a function of 3'UTR methylation for \textit{MYO1G} ........................................................................................................ 115
Figure 59: Representative plot of correlation between 3'UTR methylation and gene expression across multiple tumor types for ITK ........................................... 117
Figure 60: Node of interconnected genes according to NetWalker network analysis........................................................................................................................................ 120
Figure 61: Representative plot of correlation between 3'UTR methylation and gene expression across multiple tumor types for VAV1 ........................................... 121
Figure 62: Representative plot of correlation between 3'UTR methylation and gene expression across multiple tumor types for HAVCR2 ......................... 122
Figure 63: Association between T cell count and ITK expression in each individual TCGA sample ........................................................................................................ 123
Figure 64: Association between T cell count and VAV1 expression in each individual TCGA sample ........................................................................................................ 125
Figure 65: Association between T cell count and HAVCR2 expression in each individual TCGA sample ........................................................................................................ 126
Figure 66: Gene expression and 3'UTR methylation of Itk, Vav1, and Havcr2 in T cells after ex vivo activation ........................................................................................................ 128
Figure 67: Havcr2 methylation at various locations after ex vivo T cell activation ................................................................................................................................. 129
Figure 68: Schematic of the Havcr2 gene, with the location of the interrogated sites in each discrete region of the gene highlighted. ........................................ 130
Figure 69: Havcr2 gene expression after ex vivo activation of T cells and subsequent treatment with the demethylating agent decitabine. ................................. 131
Figure 70: Heat map of gene array showing changes in gene expression of naïve and exhausted T cells after Dnmt3a knockout........................................................................133

Figure 71: Quantitative expression data for Havcr2 expression in naïve or exhausted T cells after Dnmt3a knockout ..................................................................................134

Figure 72: Visualization of bisulfite sequencing array data that shows extent of methylation of Havcr2 gene ..................................................................................................................135
List of Tables

Table 1: NetWalker analysis of the top pathways affected by CARD11 knockdown in H1975 lung adenocarcinoma cells ........................................................................................................ 77

Table 2: The correlation between CARD11 sense and NAT across 19 tumor types .. 94

Table 3: The difference in CARD11 gene body methylation after segregating samples based on high versus low expression of CARD11-NAT ........................................................................ 97

Table 4: Partial list of correlation between 3’UTR methylation and gene expression by tissue type ........................................................................................................................................ 108

Table 5: Number of genes in which extent of 3’UTR methylation is significantly associated with patient overall survival, by tumor type ................................................................. 114

Table 6: List of genes that exhibit a correlation coefficient of > 0.5 between gene expression and 3’UTR methylation, but a correlation coefficient of < 0.5 between gene expression and gene body methylation, in at least two tissue types .................. 118

Table 7: NetWalker pathway analysis of genes from table 6 ........................................ 119
Abbreviations

DNMT: DNA methyltransferase

TET: ten-eleven translocation

UTR: Untranslated region

AID: Activation-induced cytidine deaminase

APOBEC: Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like

CTCF: CCCTC-binding factor

FMR1: Fragile X mental retardation 1

IGF2: Insulin-like Growth Factor 2

AML: Acute Myelogenous Leukemia

CMML: Chronic Myelomonocytic Leukemia

MDS: Myelodysplastic syndrome

PRC2: Polycomb Repressor Complex 2

RPPA: Reverse Phase Protein Array

TCGA: The Cancer Genome Atlas

ACC: Adrenocortical carcinoma

BLCA: Bladder urothelial carcinoma

BRCA: Breast invasive carcinoma

CESC: Cervical squamous cell carcinoma and endocervical adenocarcinoma

CHOL: Cholangiocarcinoma

COAD: Colon adenocarcinoma

DLBC: Lymphoid neoplasm diffuse large B-cell lymphoma

ESCA: Esophageal carcinoma
**GBM**: Glioblastoma multiforme

**HNSC**: Head and Neck squamous cell carcinoma

**KICH**: Kidney chromophobe

**KIRC**: Kidney renal cell carcinoma

**KIRP**: Kidney renal papillary cell carcinoma

**LAML**: Acute myeloid leukemia

**LGG**: Low grade glioma

**LIHC**: Liver hepatocellular carcinoma

**LUAD**: Lung adenocarcinoma

**LUSC**: Lung squamous cell carcinoma

**MESO**: Mesothelioma

**PAAD**: Pancreatic adenocarcinoma

**PCPG**: Pheochromocytoma and Paraganglioma

**PRAD**: Prostate adenocarcinoma

**READ**: Rectum adenocarcinoma

**SARC**: Sarcoma

**SKCM**: Skin cutaneous melanoma

**STAD**: Stomach adenocarcinoma

**TGCT**: Testicular germ cell tumor

**THCA**: Thyroid carcinoma

**THYM**: Thymoma

**UCEC**: Uterine corpus endometrial carcinoma

**UCS**: Uterine carcinosarcoma
UVM: Uveal melanoma

NAT: Natural Antisense Transcript

sgRNA: short guiding RNA
Section 1.1: The Molecular and Cellular Biology of DNA Methylation

DNA methylation is a modification that can be added directly to the DNA of a cell. It consists of a methyl group (a carbon with three hydrogens attached) that is most commonly added to a CpG dinucleotide (Fig1.), but can also be added to CpT, CpA, or CpC dinucleotides, depending on the cell type (17), and can also appear at position 6 of adenines (18). DNA methylation can be deposited or removed through passive methylation, active (de novo) methylation, or active (de novo) demethylation (19).

Passive DNA methylation occurs when identical patterns of DNA methylation are maintained from parent somatic cell to daughter somatic cell during cell division (20), primarily through the Dnmt1 enzyme. Dnmt1 has robust enzymatic activity toward hemi-methylated DNA (21), which occurs after DNA replication, when the parent strand contains methylated CpGs, and the newly synthesized daughter strand lacks DNA methylation (22) (Fig2.). The Dnmt1 enzyme adds a methyl group to the daughter strand at the same location as where the methylation exists on the parent strand (23), allowing for passing down of methylation patterns during DNA synthesis and cell division. However, cells require the ability to alter their DNA methylation patterns for a variety of reasons, including as they differentiate, or enter into different states of activation (24). Therefore, cells must de novo methylate or demethylate specific loci. De novo methylation can be carried out by the aforementioned Dnmt1
enzyme, but is more commonly governed by the activity of Dnmt3a and Dnmt3b (25), as well as the co-factor Dnmt3L (26). Together, these enzymes can alter DNA methylation patterns based on intracellular and extracellular stimuli, allowing for flexibility in cellular regulation of gene expression according to the context.
Figure 1: (Top): DNA methylation occurs by enzymatic activity, which adds a methyl group to position 5 of the cytosine carbon ring. Figure taken with permission from <http://www.ks.uiuc.edu/Research/methylation/>

Figure 2: (Bottom): DNA demethylation can occur passively, when DNA methylation enzymes are inhibited, as DNA replicates; DNA methylation can also be an active process dependent upon enzymatic activity. Figure taken with permission from <Wu, S. and Zhang, Y. Active DNA demethylation: many roads lead to Rome. Nature Reviews Molecular Cell Biology 2010(14)>
Conversely, dynamic changes to the epigenome also necessitates the capacity of the cell to *de novo* demethylate specific loci based on varying stimuli. This demethylation is most comprehensively understood to occur by the TET enzymes (27). As an example, TET1 induces the conversion of 5-methylcytosine (5-mC) into 5-hydroxymethylcytosine (28). This conversion can result in indirect demethylation after DNA replication due to Dnmt1 not recognizing 5-hmC as methylated CpG, thereby leading to an unmethylated daughter stand (19). On the other hand, 5-hmC can be deaminated by AID or APOBEC cytidine deaminases, resulting in 5-hydroxymethyluridine (5-hmU) which then is recognized by the Base Excision Repair pathway, removing the 5-hmU and replacing it with an unmethylated cytosine (29) (Fig3.). Through the Dnmt and TET enzymes, cells can both maintain and modulate their methylation profiles.

**Figure 3:** TET enzymes catalyze the conversion of 5-mC to 5-hmC, 5-fC, and 5-caC, which is then recognized by the base excision pathway that replaces the modified cytosine with an unmethylated cytosine. Taken with permission from <https://www.epigentek.com/catalog/dna-demethylation.php>
Differences in DNA methylation are a major contributor to the remarkable plasticity amongst cells that share identical genomes by providing a means of activating or silencing genes whose expression impacts the developmental or activation state of specific cell types (30). This plasticity is primarily due to the capacity of DNA methylation to affect the ability of site-specific proteins to bind to a target sequence (31) (Fig 4).

**Figure 4:** (a) Methylation of the promoter can attract methylation-specific binding proteins which prevent transcription factors from activating transcription. (b) Methylation of the enhancer can occlude this region from binding the necessary co-factors to initiate transcription. (c) Conversely, methylation of the enhancer can provide a binding site for transcription-activating factors. Taken with permission from Spruijt, C. and Vermeulen, M. DNA methylation: old dog, new tricks? Nature Structural & Molecular Biology 2014(8)
Proteins such as transcriptional repressors or activators may have increased or decreased binding affinity to a target sequence when a methyl group is present at this location (32). Furthermore, DNA methylation can affect the binding of adaptor proteins that can recruit co-factors responsible for modifying histones (33).

The modification of histones has profound effects on the structure and accessibility of DNA (34), thereby influencing manifold cellular processes (35) (Fig5.).

**Figure 5:** Histones are composed of 4 subunits, with tails that can be modified through methylation and acetylation at various locations, which impacts the structure of the histones and therefore access to regions of DNA. Taken with permission from <Li, S. et al. Host-virus interactions: from the perspectives of epigenetics. Rev Med Virol 2014(6)>
Mammalian development is heavily dependent on DNA methylation, as this process consists of a single cell giving rise to daughter cells with a large variety of different phenotypes (30). After fertilization, embryos undergo genome-wide demethylation, with levels of methylation remaining low until implantation, in which pluripotent cells begin to differentiate (36). A striking example of global changes to DNA methylation occur during hematopoiesis, where multi-potent progenitor cells give rise to numerous cell types that acquire more rigid specificity of function (37) \((\text{Fig 6.)}\).

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure6.png}
\caption{The different methylation states of each cell type and each tissue grants plasticity in morphology and function. These states evolve over time during cell differentiation. Taken with permission from <Suarez-Alvarez, B. et al. DNA methylation: a promising landscape for immune system-related diseases. Trends in Genetics 2012 (10)>}
\end{figure}
Given the broad range of regulatory power, fully elucidating the function and regulation of DNA methylation is integral to a complete understanding of cellular biology.

**Subsection 1.1.1: Promoter Methylation**

DNA methylation is widespread throughout the genome. However, methylation at the gene promoter region is the most comprehensively understood. Regulatory promoter DNA methylation typically occurs within CpG islands, which are regions of DNA enriched for CpG dinucleotides (38, 39). This results in powerful repression of transcription, primarily by recruiting repressor proteins or chromatin modifiers that enhance the binding of DNA to histones, resulting in formation of inaccessible regions called heterochromatin (40, 41) (Fig 7).

![Figure 7: Methylation of the DNA can prompt the binding of chromatin-modifying enzymes that remove histone acetylation/add histone methylation, resulting in formation of heterochromatin, and preventing access to the DNA. Taken with permission from Raabe, F.J. and Spengler, D. Epigenetic risk factors in PTSD and depression. Frontiers Psychiatry 2013(15)>
This allows for an “off switch” for gene expression within each individual cell by altering their accessibility. Methylated promoters typically contain the H3K27me3 histone modification, causing an adoption of tightly associated DNA/histone interactions, which are tightly wound together, sequestering the genes away from transcriptional activators (42). Therefore, when the CpG island of a gene promoter is methylated, this gene can be considered “turned off”; however, if the promoter is unmethylated, this does not guarantee that the gene will be expressed, but only that it can be readily expressed if the necessary transcription factors, co-factors, and transcriptional complexes are present (43).

1.1.2: Enhancer Methylation

Enhancers, as their name implies, enhance the transcription of a downstream gene by serving as a binding site for transcription factors and co-factors that can activate or boost gene expression (44) (Fig8.).
Figure 8: The enhancer is located upstream of the promoter, and during transcripational activation, is bound by activators which interact with transcription factors and RNA polymerase through bending of the DNA. Taken with permission from <http://www.mun.ca/biology/desmid/brian/BIOL2060/BIOL2060-23/CB23.html>
In a similar fashion to promoter sequences, the methylation of enhancer sequences can modulate the ability of said factors to bind and influence gene expression (45, 46), and DNA methylation of enhancers is inversely correlated with expression of enhancer-regulated genes (47). Also like promoter methylation, enhancer methylation alters the chromatin structure by promoting changes to the chromatin architecture via recruitment of histone modifiers. Methylated enhancers are highly associated with H3K27me3, a repressive marker, and are inversely correlated with H3K27ac, which promotes an open chromatin state, and access to DNA (48) (Fig9.).

**Figure 9:** Like promoters, enhancer DNA and histones can be modified, typically by methylation or acetylation. These modifications influence the activation state of the genes they regulate by impacting the ability of critical factors and co-factors necessary for transcriptional activation. Taken with permission from <Atlasi, Y. and Stunnenberg, H. The interplay of epigenetic marks during stem cell differentiation and development. Nature Reviews Genetics 2017(12)>
Conversely, hydroxymethylated CpGs in the enhancer region (a product of TET enzymatic activity on methylated CpGs as discussed above), are associated with H3K27ac and H3K4me1, two markers of highly active gene expression (49, 50). Taken together, DNA methylation at enhancers can impart powerful repression on downstream genes in a manner similar to promoter methylation.

**Subsection 1.1.3: Intergenic Methylation**

Intergenic DNA is any region of DNA that is not a part of a gene or a regulatory element, and makes up approximately 75%-90% of the genome (51). While these regions do not code for proteins, they can code for non-coding RNAs, many of which are essential for regulating cellular processes (52). Therefore, understanding the intergenic methylation landscape holds key insights into cell biology. Indeed, the regulation by intergenic methylation of important non-coding RNAs such as Xist (53) and Air (54) have been thoroughly demonstrated. New non-coding RNAs transcribed from the intergenic region are discovered each year, and with these discoveries comes a growing importance for ascertaining the effects of intergenic methylation.

Approximately 45% (55) of intergenic DNA consists of the remnants of genomic parasites such as retroviruses, transposons, and retrotansposons (56) (Fig 10.).
These transposable elements are oftentimes detrimental to the normal function of the cell, and their replication and transcription are actively opposed by cellular defense mechanisms (57, 58). The primary mechanism for suppressing transposable elements is DNA methylation, and the cell's effort to inactivate these parasites is estimated to account for 35% of all the DNA methylation found in the genome (59).

**Figure 10**: Multiple types of transposons exist, and are classified by how they replicate themselves. Some transposons are converted to RNA and reverse transcribed into DNA, whereas others do not go through an RNA intermediate. Taken with permission from <Bonchev, G. and Parisod, C. Transposable elements and microevolutionary changes in natural populations. *Molecular Ecology Resources* 2013(7)>
Repression of transposable elements by DNA methylation typically occurs by methylating the promoter region of the transposable element, thereby silencing its expression (60). Loss of DNA methylation enzymes and co-factors result in substantially increased transcription of these elements (58), which can cause catastrophic cellular consequences, such as meiotic recombination in mitotic cells (61).

DNA methylation can also take place at insulators, which serve to diminish transcriptional activity by disrupting the interaction between enhancers and transcriptional sites through looping of the DNA; additionally, insulators isolate heterochromatic regions, preventing them from spreading to euchromatic areas (62). The most well-studied cause for how methylation impacts these sites is through excluding CTCF binding (63). CTCF is instrumental in forming functionally isolated segments of DNA through chromatin looping (64). Binding of DNA by CTCF is methylation sensitive (65), and methylation at CTCF binding regions is highly variable in different tissues (66) (Fig11.).
Figure 11: CTCF induces isolated regions of DNA through chromatin looping. CTCF binding to its target sequence is methylation sensitive. If CTCF is bound to DNA in between an enhancer and its corresponding promoter, it can inhibit the transcriptional activation effects of the enhancer. Therefore, these binding sites are called “insulators”, because they insulate segments of DNA. Taken with permission from <van Montfoort, AP. Et al. Assisted reproduction treatment and epigenetic inheritance. Hum. Reprod. Update 2012(16)>
Intergenic methylation is therefore highly important for normal cellular function, and unraveling currently unknown functions has the potential to reveal novel insights into gene regulation and genomic structural assembly.

**Subsection 1.1.4: Gene Body Methylation**

Gene body methylation, also called intragenic methylation, is the term used to describe methylation on any part of the gene that is transcribed into mRNA. This includes exons and introns, as well as the 5’ and 3'UTR, and accounts for 40% of all methylated CpGs in the mammalian genome (67). Unlike promoter methylation, gene body methylation has a more complicated set of functions that can either increase or decrease gene expression, both directly and indirectly. Some of the first investigations involving this type of modification were conducted on the active X chromosome (68). The active X chromosome is differentially methylated from its inactive counterpart, and is characterized by methylation in the gene body of certain highly expressed genes (69). A large-scale study of cell line-specific genome-wide methylation levels observed high gene body methylation in highly expressed genes that was generalizable to the entire genome (70). Further analysis of gene body methylation uncovered a robust association with H3K36me3, a histone marker of actively transcribed genes (71) (**Fig12.**).
Figure 12: DNA methylation is low in the promoter region and enriched within the gene body of actively transcribed genes. Gene body methylation is strongly associated with H3K36me3. Taken with permission from <Moreselli, M. et al. In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. Elife 2015(2).
While there is evidence that DNA methylation may be downstream of activated transcription and H3K36me3 deposition (72), gene body methylation has been shown to be upstream of gene expression, and blocking re-methylation of a gene can abrogate its expression (73).

Much remains to be understood about the mechanisms by which gene body methylation conveys its effects. Indeed, multiple mechanisms have been uncovered, depending on the specific location of the methylation. Like promoter methylation, gene body methylation can suppress expression, but not of the gene itself; rather, it serves to prevent spurious transcription of downstream promoters. An analysis of differentially methylated regions within the gene body showed these locations to have similar architecture as promoters (74). By suppressing intragenic transcription initiation, methylation of the gene body results in greater transcription from the canonical transcription initiation site, leading to greater expression. This phenomenon is primarily mediated by Dnmt3b, and is dependent on H3K36me3 deposition by the SetD2 enzyme (75). Generally speaking, exons contain higher levels of methylation relative to introns (76, 77). Additionally, the intron/exon junction is enriched for methylation, pointing to a role in determining the boundaries between these regions of genes (78). Gene body methylation has also been shown to directly modulate mRNA splicing by recruiting the methylation-sensitive protein MeCP2 to an exon that is typically spliced out of the mRNA product (79). This causes a slowing of the RNA polymerase, which results in inclusion of the MeCP2-bound exon. Additionally, methylation of a weak exon in the CD45 gene prevents CTCF binding, which results in a splicing out of the exon when methylated (80) (Fig13.).
Figure 13: Methylation of a specific exon within the CD45 gene prevents CTCF from binding. Binding of CTCF to this exon slows the RNA polymerase, which facilitates inclusion of the exon into the mRNA transcript. Therefore, methylation of the exon induces its exclusion, which demonstrates that DNA methylation can influence RNA splicing. Taken with permission from Shukla, S. et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature 2011(4).
In certain contexts, dense gene body methylation can diminish transcriptional output, mainly by inducing chromatin structures that close the gene off to RNA polymerase binding(81). Given these diverse and often contradictory functions, the impact of gene body methylation on normal cellular processes and disease remains a topic requiring further exploration.

### Section 1.2: DNA Methylation in Disease

Given the numerous purposes of DNA methylation as described above, aberrant changes to DNA methylation can be both a cause and a consequence of disease (82). From the very beginning of a new life, DNA methylation is essential; embryonic differentiation requires proper DNA methylation, and loss of DNA methylation results in embryonic lethality (83). Mutations to methylation enzymes have been linked to diseases such as Immunodeficiency, Centromeric Instability and Facial Anomalies syndrome, of which approximately 70% of patients contain a $DNMT3B$ mutation (84). Furthermore, mutations in $DNMT1$ lead to a variety of neurodegenerative conditions, including Hereditary Sensory and Autonomic neuropathy (85, 86). One group of diseases indirectly linked to DNA methylation are those caused by expansion of repetitive sequences (82). For example, Fragile X Syndrome is characterized by numerous repeats of CGG trinucleotides in the $FMR1$ gene, which serves as a target for hypermethylation, and consequently reduces transcription of this gene (87) (Fig14.). Conversely, hypomethylation due to tandem repeats as seen in Facioscapulohumeral Muscular Dystrophy can also be deleterious (88).
In Fragile X syndrome, tandem repeats of the CGG trinucleotide sequence can result in hypermethylation of this region, which silences transcription of the gene, and produces deleterious effects. Taken with permission from <http://modencode.sciencemag.org/drosophila/autism/>
A major source of methylation-related disease is improper imprinting (82). Because imprinting allows for allele-specific gene expression, improper expression of both alleles, or abnormal paternal/maternal expression can disrupt cellular processes. One specific example is that of Prader-Willi Syndrome (PWS), in which a piece of chromosome 15 loses normal imprinting patterns. The consequence is lack of expression of paternal alleles in this region (89), which together cause mental impairment and behavioral issues. Another disease, Angelman Syndrome (AS), also arises from changes to imprinting on chromosome 15 in approximately 5% of cases, and has been isolated due to a single gene, *UBE3A*. This ubiquitin ligase is specifically expressed from the maternal allele in the brain, and loss of imprinting leads to neurological defects (90). Numerous other disorders result from similar changes to normal imprinting (91).

**Subsection 1.2.2: DNA Methylation in Cancer**

DNA methylation patterns experience global changes within cancer cells (92). These changes can be summarized into two groups: regions of hypomethylation and regions of hypermethylation (93). In general, cancer cell genomes are hypomethylated (93), with intergenic regions containing the most drastic differences (94), as well as centromeric and pericentromeric regions (95).
Figure 15: Cancer DNA methylation is characterized by hypermethylated promoters of tumor suppressors, hypomethylated promoters of oncogenes, global decrease in DNA methylation, and resulting genomic instability. Taken with permission from <Phillips, T. The role of methylation in gene expression. Nature Education 2008(5)
Perturbed methylation of these areas contributes significantly to the cancer hallmark of genomic instability (96). Hypomethylation produces genomic instability through multiple mechanisms, such as increased transcription of transposable elements (97), destabilization of centromeres (98, 99), and errors in chromosome segregation during mitosis (100). Genomic instability through changes in methylation can lead to higher rates of mutation, hastening tumorigenesis and/or worsening tumor progression (101). Hypomethylation can also directly influence gene expression, oftentimes by promoter demethylation and consequent overexpression of oncogenes (Fig15.). This is frequently observed in imprinted genes (94). As a specific example, certain growth-regulatory genes such as IGF2 are only expressed from the maternal or paternal allele (102) by allele-specific methylation; when normal imprinting is disrupted, both alleles are switched on, causing dysregulated cell growth. This phenomenon has been observed across multiple tumor types, including colorectal cancer (103), prostate cancer (104), and others (105, 106). Furthermore, loss of imprinting in IGF2 serves as a marker of colorectal cancer risk (107). Promoter demethylation of oncogenes is also frequently observed, and can worsen outcomes. ELMO3, a gene that promotes NSCLC migration and metastasis, contains less promoter methylation and higher expression in metastases than in primary tumors or untransformed cells (108). Numerous other oncogenes display a similar phenomenon (109). As our knowledge of site-specific methylation across the genome continues to grow, so too the list of oncogenic processes influenced by hypomethylation will grow as well.
While the vast majority of the cancer genome experiences hypomethylation, hypermethylation of specific regions substantially impacts the formation and progression of nearly every type of cancer (110). The function of tumor suppressor genes such as TP53 are frequently inactivated through mutation (111), but mutation is a random process, whereas DNA methylation allows cancers to eliminate tumor suppressors in a controlled manner. Since the p53 protein is the “guardian of the genome” (112), it comes as no surprise that the TP53 is frequently methylated in cancer cells relative to normal cells (113-115). Other genes essential for genomic integrity, such as the mismatch repair gene MLH1 are observed to be switched off by promoter methylation in cancer, compromising mismatch repair, and massively increasing mutation rates (116). In addition to impacting genomic stability, aberrant tumor suppressor methylation can immensely increase cell proliferation and growth by targeting cell cycle regulators. For example, the CDKN2B gene is a negative regulator of cell cycle, and is frequently methylated in ovarian cancer, leading to unrestrained tumor growth (117). Moreover, aberrant epigenetic modification has been implicated in immune evasion through downregulation of antigen presentation molecules, tumor-specific antigens, apoptosis-inducing receptors, and receptors employed by NK cells to kill tumor cells (118).

In summary, cancer DNA methylation experiences global changes, both hypomethylation and hypermethylation, that serve as a driver both for tumorigenesis and for tumor progression. Therefore, developing clinical strategies to modulate these changes in the epigenome is a promising avenue to developing new clinical tools to treat cancer.
Subsection 1.2.2: Clinical Strategies for Targeting DNA Methylation

Given the systemic changes in cancer DNA methylation as discussed above, gaining the ability to normalize or disrupt the epigenetic network established in cancer holds therapeutic potential. Therefore, drugs that modulate DNA methylation have become increasingly studied, both at the pre-clinical and clinical levels (119). Two broad categories of demethylating drugs exist: non-nucleoside analogs and nucleoside analogs (120). Non-nucleoside analogs typically convey their effects through direct actions to DNMT enzymes (119). For example, procanimide works as a competitive inhibitor for DNMT1 binding, preventing DNMT1 from methylating DNA (121). While these therapies produce lower toxicity relative to nucleoside analogs, they also achieve lower levels of demethylation (122). Nucleoside analogs are those drugs that are incorporated into nucleic acids (123). Two forms of this type of methylation inhibition have been employed in the clinic: azacytidine and 5’-deoxy-azacytidine (119). These work by incorporating themselves into actively replicating DNA, then trap DNMT enzymes and induce their degradation by the proteasome (124) (Fig16.).
Figure 16: Azacytidine works by binding to the DNMT enzymes, which lead to their degradation by the proteasome pathway. Upon DNA replication, lack of DNMT enzymes result in passive DNA demethylation. Taken with permission from <http://www.vidaza.net/mds/about-vidaza/moa/>
Therefore, these inhibitors require DNA replication to convey their effects, making them well-suited for cancer, which replicates its DNA more rapidly than normal cells (125). These drugs are currently being used in cancer patients, with blood-borne diseases being the most susceptible (126, 127). Currently, these two types of drugs are approved for use in AML, CMML, and MDS (128), with clinical trials also being conducted in solid tumors such as melanoma and renal cell carcinoma (120). Interestingly, recent reports have shown demethylating agents to boost T cell killing of tumor cells (129), and to have a synergistic effect with immune checkpoint inhibitors (130-132). However, not all patients respond to either demethylating agents alone, or in combination; therefore, a greater understanding of which patients should receive this therapy, and the specific mechanism of how this therapy conveys its effects, is needed.

Section 1.3: Non-coding RNAs

The central dogma of molecular biology is that DNA-based genes are transcribed into messenger RNAs, which are then translated into proteins that exert specific functions (133). However, a more comprehensive understanding of genomics has broadened our conception of what constitutes a gene, and what composes the final product of a gene (52). While protein coding genes make up only 2% of the genome, upwards of 70% of the genome is transcribed into RNA (134). RNAs that do not code for a protein have been discovered many years ago, particularly those involved in the ribosomes (135) and transfer RNAs (136). Due to technologies such as RNA-seq, our knowledge of RNAs that are transcribed but never code for a protein
has expanded exponentially (137). Non-coding RNAs have now been implicated in a wide range of cellular processes, and constitute an essential means of regulation(52). Numerous types of non-coding RNAs exist, from the aforementioned transfer and ribosomal RNAs, to piRNAs, snoRNAs, and snRNAs (138) (Fig17.); this discussion will focus on long non-coding RNAs (lncRNAs).

Furthermore, RNAs that arise from the opposite strand of genes, called natural antisense transcripts, will also be discussed separately, although many fall under the categories of lncRNAs and microRNAs (139).

![Figure 17: Non-coding RNAs can fall into several groups, including microRNAs, small RNAs, and medium to long RNAs that are involved in a wide variety of cellular processes. Taken with permission from <https://noncodingrnaexplorer.wordpress.com/non-coding-rnas/>](image)
Subsection 1.3.1: Long non-coding RNAs

Long non-coding RNAs are any RNAs that contain > 200 nucleotides (140). They can be transcribed from introns, from intergenic RNA, or can overlap with exons (141). Like messenger RNA, IncRNAs can contain exons and introns (142). Because of the inherent flexibility of RNA, IncRNAs can fold into structures that can function as an enzyme (143), or that can bind DNA, proteins, or other RNAs, and function as scaffolds (144).

**Figure 18:** The flexible nature of RNA allows for IncRNAs to function as enzymes, as scaffolds, or as sponges for other RNAs. Taken with permission from <Malik, B. and Feng, F. Long noncoding RNAs in prostate cancer: overview and clinical implications. *Asian Journal of Andrology* 2016(9)>
Given this plasticity, lncRNAs are involved in processes ranging from transcriptional regulation, to chromatin modification, to post-transcriptional regulation (143) (Fig 18). Arguably, the most well-studied lncRNA is that of XIST, which induces the silencing of the X chromosome in cells that contain two X chromosomes (145).

While the mechanism of XIST is complicated, one of the most essential features is its recruitment of PRC2, a chromatin modifier that methylates H3K27, causing formation of heterochromatin and gene silencing (146). In this case, one domain of the lncRNA is capable of recognizing specific regions of DNA, while another domain can bind to PRC2, resulting in localization of PRC2 to the region to be silenced, a common mechanism of lncRNA function. As mentioned, lncRNAs can also bind mRNAs; this aspect will be explained in the natural antisense transcripts section.

Subsection 1.3.2: Natural Antisense Transcripts

Natural antisense transcripts (NAT) are any RNA transcripts that are transcribed from the opposite strand of a gene (147). Previously written off as transcriptional noise (148, 149), new technologies have implicated these fascinating RNA molecules as more than a byproduct (150-153). NAT can regulate gene expression both in cis and in trans through a variety of mechanisms, including transcriptional interference (154), modulating DNA methylation (76), imprinting (155), influencing histone modification (156), and impacting sense mRNA stability (157) (Fig 19).
Figure 19: Natural Antisense transcripts can have a wide range of both *in cis* and *in trans* functions, such as blocking sense transcription through RNA polymerase collision, modulating RNA splicing, promotion or inhibition of RNA stability, and induction of chromatin modification. Taken with permission from <Lapidot, M. and Pilpel, Y. Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. *EMBO* 2006(11)
Due to the overlapping sites of transcription between sense and antisense strands, active transcription of the antisense can result in decreased expression of the sense mRNA (154). This is primarily due to the sheer size of the transcriptional complex physically occluding transcriptional initiation, as well as collision of the two transcriptional complexes as they transcribe toward one another (158, 159). The antisense transcript of \textit{HBA2} extends into the promoter CpG island, and knockout of the antisense RNA prevents the methylation of this promoter, thereby resulting in increased expression of \textit{HBA2} (160). The \textit{IGF2R} gene is well known to be imprinted, and this imprinting depends on expression of a NAT (161). The paternal copy of \textit{IGF2R} contains an unmethylated CpG island in an intron, which regulates expression of the NAT \textit{AIR}, which represses expression of \textit{IGF2R} by recruiting transcriptional repressors \textit{in cis} to this locus (162). The NAT \textit{ANRIL} is capable of binding to chromobox 7 of the Polycomb Repressor Complex, and can localize this histone modifying enzyme to discrete regions of the \textit{INK4/ARF} gene complex, inducing H3K27 methylation, which silences expression of these genes through heterochromatin formation (163).

Considering the diverse functions of NAT, mutation or aberrant expression of these RNAs can play a role in diseases, including cancer (164-166). For example the aforementioned targets of \textit{ANRIL, INK4/ARF}, are tumor suppressor genes (167). \textit{ANRIL} expression is altered in many cancers (168-170), and through a variety mechanisms both \textit{in cis} and \textit{in trans}, can disrupt genomic stability (171) and normal cellular proliferation (171). Another example is that of the \textit{ZEB2} NAT, which impacts cancer metastasis (172). The \textit{ZEB2} NAT promotes retention of the \textit{ZEB2} mRNA.
5'UTR, which is critical for efficient translation (172) (Fig20.). Therefore, determining the function of NATs, especially those who are antisense to oncogenes or tumor suppressors, are essential for a complete understanding of tumorigenesis and tumor progression. Furthermore, they provide novel therapeutic targets.

**Figure 20**: The NAT of Zeb2 is critical for retention of the Zeb2 5'UTR in the mRNA transcript. Retention of the 5'UTR promotes efficient translation, which thereby increases Zeb2 protein. This process is important for mediating EMT. Taken with permission from <Serviss, J. et al. An emerging role for long non-coding RNAs in cancer metastasis. Front. Genet. 2014(13)>
Section 1.4: Immunology

Human immunology is an extraordinarily complex system composed of scores of specific cell types that carry out specific roles in defending the body from invaders, both from outside the body and from within (173). Immune cells begin with a common progenitor cell, and differentiate into a spectacular array of cell types, a process known as hematopoiesis (174). This is governed through a complex series of cell-specific changes to gene expression, which is accomplished primarily through modulating the epigenetic profile (175). Investigating changes to the immune epigenome has major implications for autoimmune disease, treatment of lymphoma, immunotherapy, vaccine generation, and a host of other topics. Therefore, gaining a comprehensive understanding of how the epigenome is controlled, and the functional implications of these modifications, is essential for developing new drugs and advancing medicine.
The immune system can be broken down into two major components: innate and adaptive immunity (176) (Fig 21.).

Figure 21: The immune system can be broken down into two major components: the innate immune system is activated rapidly upon foreign invasion, and consists of cells such as granulocytes, macrophages, and dendritic cells that both eliminate the invader, and prime the immune system for long-term combat. The adaptive immune system is triggered by the innate immune system, and can work over the span of days to weeks. The adaptive immune system specifically hones in on targets after being primed by antigen from the innate immune system. Taken with permission from <https://cias.rit.edu/faculty-staff/101/faculty/336>
The innate immune system serves as the first line of defense against pathogens, activating within the span of minutes (177). Innate immunity is a general phenomenon that does not determine its targets based on specific epitopes; rather, it can rid regions of pathogens or toxins through non-specific recognition of foreign bodies (178).

Common bacterial epitopes such as lipopolysaccharides and common viral molecules like double-stranded RNA serve as triggers that induce immune cells to unleash a torrent of cytokines and other factors that recruit more immune cells to the site of interest (179). The most numerous cell types include neutrophils, macrophages, and dendritic cells, which perform the actions listed above, and serve as a link to the adaptive immune system (180).

Dendritic cells and macrophages, along with other cell types, are able to activate the adaptive immune response by presenting molecules obtained from foreign bodies on their surfaces (181). The two most prominent cell types of adaptive immunity are that of B cell and T cells (181). B cells are responsible for secreting antibodies that can bind with exquisite precision to specific antigens on a systemic scale (182). When the B cell receptor comes in contact with an antigen presented by dendritic cells, macrophages, etc., that matches, it will trigger the production of large quantities of secreted antibodies, which upon binding to their target, can render the target non-functional, as well as induce destruction by other cell types in the area (182). Antibodies work with such efficacy and specificity, that dozens of monoclonal antibodies toward various targets in both cancer and other diseases, have proven as
beneficial in the clinic (183). For example, a monoclonal antibody against the B cell receptor CD20, called rituximab, has been approved for multiple types of blood cancers (184). However, despite massive successes, monoclonal antibodies have major limitations, such as the ability only to bind surface molecules, and the capacity of cancer to gain mutations that invalidate the antibody targeting (185). Therefore, the greatest immunological response also requires activation of T cells, which are able to target intracellular or extracellular proteins.

Subsection 1.4.1: T cells

T cells, similar to B cells, are activated upon stimulation by antigen presented by antigen-presenting cells that binds to their T cell receptor (186). This triggers clonal expansion of the activated T cell, primarily through intracellular signaling from the T cell receptor to downstream targets such as NF-kB (187) and mTOR (188). T cells come in two overarching types: CD4 and CD8, which carry out different functions in mounting an immune response (189). CD4 cells are known as “T helper cells”, which support other immune cells through the secretion of cytokines such as IFNγ, IL-2, and IL-4 (190). Some CD4 T cells serve as negative regulators of immune response as a means of reigning in immune activation after a threat has been dealt with. These are known as regulatory T cells, and are a major mechanism by which cancer evades the immune system (191). CD8 T cells, also known as cytotoxic T cells, directly interact with and destroy any cells that express antigen that binds with their T cell receptor (192) (Fig22.). Therefore, CD8 T cells offer great potential in seeking out and killing tumor cells at an individual cell basis. Understanding how CD8 T cell killing is
mediated and regulated is therefore of utmost importance in treating diseases, whether it be cancer, autoimmune disease, or infection.

Figure 22: T cells are activated by recognizing antigen that is presented on the MHC by antigen-presenting cells such as macrophages and dendritic cells. Upon stimulation, T cells undergo clonal expansion, and can specifically seek out and kill cells that express the antigen that their T cell receptor recognizes. Taken with permission from <http://bit.ly/2GXvLDR>
Subsection 1.4.2: Cancer Immunology

The immune response to cancer becomes gradually compromised by specific mechanisms employed by the tumor to evade immune recognition (193). This is accomplished through direct and indirect interaction with immune cells (193). However, for every collection of cancer cells that become a full-fledged tumor, many potential cancer cells are recognized and removed by the immune system, whereas a compromised immune system creates a more favorable environment for tumor development and growth (194). Due to mutation, cancer cells produce proteins that can be discerned from “self”, allowing for an immune response (195). How the tumor is able to circumvent this system is the topic of the following discussion.

Cancer has two primary strategies for evading the immune system: the first is to prevent cancer-detecting and –killing immune cells from entering the tumor microenvironment, and the second is to alter, suppress, or invalidate the function of the immune cells that are present in the microenvironment (196). One example of how tumors can alter cellular phenotype is that of macrophages, which undergo a phenomenon called polarization that causes them to adopt an anti-tumor (M1) or pro-tumor (M2) cellular state (197). Tumors push macrophages to the M2 phenotype by secreting factors such as CCL2, CSF1, and IL-4, as well as by producing hypoxic conditions that induce these changes (198). M2 macrophages then assist the tumor in immunosuppression, growth, angiogenesis, and metastasis (199). Tumors suppress dendritic cells by blocking upstream dendritic cell differentiation, and inhibit antigen presentation through secretion of factors such as IL-10 and VEGF (200). The most potent anti-tumor role of dendritic cells is the priming of T cells, which as
previously mentioned, can identify and kill tumor cells on an individual cell basis. Therefore, a frequent strategy of tumor cells is to prevent T cells from infiltrating into the microenvironment (201). The central mechanism is by downregulating the surface proteins on endothelium that cause T cells to identify the site of interest (202), and by depositing stromal components that block T cell infiltration (203). This is a highly effective strategy, as low T cell infiltration is a robust predictor of tumor growth and spread (204). Additionally, tumor cells will actively recruit T regulatory cells and myeloid-derived suppressor cells, which are immunosuppressive cells that inhibit cytotoxic T cells (205). If T cells are able to navigate to the cancer microenvironment, tumor cells are able to exploit immune pathways that rein in the immune response in order to evade being killed.

Subsection 1.4.3: Immune Checkpoint

Just like every cellular process, the immune system must have a mechanism for activation, and a mechanism for repression once the threat has been dealt with. In the case of cytotoxic T cells, this mechanism comes in the form of the immune checkpoint pathways. Over the past two decades, components of immune checkpoint pathways have been and continue to be uncovered (206). To date, the two most clinically important pathways are the CTLA-4 and the PD-1/PD-L1 pathways. The CTLA-4 pathway was first discovered by Dr. James Allison as a negative regulator of
T cell priming and activation when they are being stimulated by antigen-presenting cells (207). (Fig 23.)

**Figure 23:** During T cell activation that occurs when a T cell engages antigen specific to its receptor via antigen-presenting cells, co-activation of T cell CD28 by B7 is also required. CTLA-4 can bind to B7 and prevent the necessary co-activation signals required for T cell activation.
Based on this understanding, it was hypothesized that blocking CTLA-4 could induce greater T cell activity in cancer patients, leading to increased tumor cell killing by cytotoxic T cells. This hypothesis proved correct, providing stark, sustained survival benefits and stable or eliminated disease in a fraction of patients, with the most efficacious tumor type being advanced stage melanoma, a disease that had few treatment options (208). The second immune checkpoint that has served as a clinical target is the PD-1/PD-L1 pathway. While CTLA-4 functions between antigen-presenting cells and T cells, PD-1/PD-L1 works at the interaction between a T cell and its potential target (209). When PD-L1 is expressed by the target cell, it functions as an “off switch” for the T cell, preventing it from killing the target. PD-L1 is overexpressed in numerous cancers, which has allowed for PD-1/PD-L1 inhibitors to achieve remarkable success in treating a diverse set of tumor types (210) (Fig 24).

**Figure 24:** PD-1/PD-L1 binding between tumor and T cell nullifies activation of the T cell receptor. Blocking this interaction with an antibody can reactivate T cells.
However, some patients respond extremely well, while others do not respond at all, necessitating a great need for biomarkers (210). Furthermore, a large variety of drugs, including CTLA-4, are now being used in combination with PD-1/PD-L1 inhibitors to boost efficacy (211). One promising discovery in this regard has been that of the TIM-3 protein, which is a powerful regulator of T cell exhaustion, and has been implicated in resistance to PD-1/PD-L1 therapy (212). Therefore, greater understanding the mechanisms that govern expression and function of these immune checkpoints is essential for rational discovery of biomarkers, as well as efficacious combinations that fully harness the potential of immune checkpoint blockade.

Section 1.5: Autophagy

Autophagy is a cellular process that grants cells the ability to “eat themselves” in certain contexts by degrading and recycling components of the cytoplasm such as organelles (213). There are multiple contexts in which this process is important for normal cell function and cell survival. For example, autophagy is employed to sequester and degrade protein aggregates that can accumulate in cells (214). Loss of autophagy in this context results in cell death, and in line with this function, autophagy has been linked to Alzheimer’s disease (215). Conversely, autophagy is typically activated in low-nutrient conditions. This allows for the cell to persist without external nutrients by recycling macromolecules in non-essential organelles (216) (Fig 25.).
Figure 25: Autophagy is a process by which intracellular components can be degraded and recycled. These components are engulfed by a lipid membrane that forms the autophagosome. This then fuses with the lysosome, which contains degradative properties that can break down the cellular components. These macromolecules are then used based on the cell's need.
The central regulator of autophagy activation is mTOR (217). This signaling node is activated in the context of plentiful nutrients, as it serves to drive cell growth and proliferation. Therefore, when nutrients are low, the mTOR pathway is inactivated, which can then trigger induction of autophagy (218). However, mTOR pathway inactivation is not the only step necessary to activate this process. Atg1, a serine/threonine kinase, must be activated, which then phosphorylates downstream mediators of autophagy such as Atg13 and Atg17. Furthermore, Beclin-1 must be dissociated from its negative regulator for autophagy to proceed. This illustrates the tight regulation of this process (217).

The process of autophagy can be roughly divided into two stages: early stage and late stage (219). The early stage consists of the initial steps, until formation of the autophagosome. When autophagy is induced, the cytoplasmic target is engulfed in an autophagosome, which initially consists of a double membrane containing proteins such as ATG5 and ATG12. In the late stage, the autophagosome fuses with the lysosome, creating the autophagolysosome. The lysosome contains low pH and numerous degradative enzymes that facilitate the breakdown of components contained within the autophagolysosome. The degraded molecules are then exported to the cytoplasm by a process still not well-understood (220).

Given the nature of autophagy, as well as the functions outlined above, too much or too little can be detrimental to the cell, and can eventually lead to cell death (221). Like apoptosis, autophagy can function as a cellular mechanism to induce death in certain contexts, including in developmental programming or stress-inducing
extrinsic factors (222). Therefore, the ability to modulate autophagy in cancer cells offers a promising avenue for induced cell death.

Subsection 1.5.1: Autophagy and cancer

Autophagy plays a complex role in various types of cancer, and depending on the context, can function as either oncogenic or tumor suppressive (223). For example, since autophagy is involved in mitigating oxidative stress, reduced autophagy can increase this condition in tumor cells, which may promote tumorigenesis through greater genomic instability (224). This is borne out both in the lab and in the clinical setting. The critical autophagy gene \( \text{BECN1} \) is frequently found deleted in prostate cancer (225). Mice heterozygous for this gene have been shown to spontaneously develop lung and liver cancer (226). Conversely, autophagy can be highly beneficial to tumor cells, particularly when considering the high-stress intracellular and extracellular environment that results from cancer metabolism (227). One clear example of this is the metabolically overtaxed cancer cells with RAS mutations. These tumor cells have been shown to contain high levels of autophagy, which has been hypothesized to be a coping mechanism that allows cells to survive in an otherwise hostile situation (228). Indeed, genetic ablation of autophagy genes in this context results in diminished tumor burden, primarily through activation of p53 as a result of intracellular stress signals (229). Whether autophagy is acting as tumor suppressive or oncogenic, these data make it clear that tight regulation of this process is essential for optimal cancer growth and progression; because of this fine balance, tipping the scale in one direction or another could be catastrophic for tumor cells. A
greater understanding of autophagy in each tumor type, how it is regulated, and the functional consequences of perturbing this process, are essential for developing novel therapeutic avenues for targeting cancer.

Section 1.6: Bioinformatics and TCGA

A decade and a half ago, the Human Genome Project successfully sequenced the entire human genome; today, the human genome can be rapidly sequenced for less than $5,000 (230). Moreover, new technologies such as RNA-seq, genome-wide methylation assays, and RPPA has necessitated scientists skilled at large-scale data analysis. In the era of personalized medicine, the disciplines of bioinformatics and biostatistics are becoming ever-more important. The emergence of this skillset has been particularly valuable for cancer, which consists of numerous mutations and network alterations that require a birds-eye view to piece together (231). This had led to the many important insights, such as mutational frequency among tumor types and tumors within a single type, among many others (232). Therefore, collaboration between molecular biologists, physicians, and bioinformaticians is a powerful combination that will facilitate discoveries that would not have been accomplished by each individual discipline.

The Cancer Genome Atlas (TCGA) is a publicly funded and worldwide collaborative endeavor that performs multiple forms of molecular profiling of patient normal and tumor tissue across over 30 different tumor types (233), including survival data. The molecular profiling includes RNA-seq, genome-wide methylation arrays,
DNA sequencing, RPPA, and more (233). These datasets are publicly available, and have served as the basis for numerous publications. Here, we have harnessed these datasets to investigate relationships between site-specific DNA methylation and gene expression across at least 10 tumor types. This has revealed two key discoveries: 3'UTR methylation as important in regulation of TIM-3, and the connection between \textit{CARD11} gene body methylation, gene expression, and kidney renal cell cancer progression.

**Overall Summary**

With the advent of large-scale epigenomic profiling, novel discoveries into how the epigenome changes during tumorigenesis and tumor progression are being uncovered. However, little is known about how these changes impact cancer pathogenesis, particularly in regard to regions outside of the promoter. Therefore, a comprehensive investigation into how epigenetic changes such as DNA methylation play a role in these processes is warranted, both within the tumor cells themselves, as well as within the stromal compartment. The following study seeks to add to the knowledge of how DNA methylation functions in regard to gene expression, and what the consequences of changes to these functionally relevant sites of DNA methylation are in relation to cancer pathogenesis.

**Section 2: Materials and Methods**

**Cell Culture**

The UOK111 renal cell carcinoma cells were a kind gift from Dr. Eric Jonasch. These cells were grown in DMEM (Clontech) +10% FBS +1% PenStrep, and were split upon
reaching 80% confluency. The H1975 lung adenocarcinoma cells were a kind gift from Dr. John Heymach. These were grown using RPMI (Clontech) +10% FBS +.1% gentamycin. These cells were split upon reaching 80% confluency. Each cell line was validated using STR testing, and were low passage.

**Western Blotting**

Cells were spun down at 1,500xRPM for 5 minutes, then the spin was repeated after washing with PBS. Cell pellets were then re-suspended in RIPA buffer, and quantified using the BCA assay (Thermo Scientific Pierce BCA Protein Assay Kit). Equal amounts of protein were run on an SDS-PAGE gel (8%-12%), then transferred to a nitrocellulose membrane, and incubated in 5% milk (in TBS-T) for one hour, then incubated in primary antibody overnight. Blots were then washed with TBS-T for 3x20 minutes, then incubated in secondary antibody (GE Healthcare, 1:2000) for one hour. ECL was then added to the blots for 3 minutes, then exposed to a film for quantification of protein. The following antibodies were used: anti-CARD11 (Cell Signaling Technology, 1:1000), anti-phosphoP70S6K (Cell Signaling Technology, 1:1000), anti-P70S6K (Cell Signaling Technology, 1:3000), anti-GAPDH (1:5000), anti-LC3B (Cell Signaling Technology, 1:1000).

**siRNA Transfection**

Cells to be transfected were plated in 6-well plates at 50% confluency the day of transfection. For each well, 1.3µg of siRNA was added to 50µL of serum-free media,
In a separate tube, 4µL of FuGene HD Transfection Reagent (Promega) was incubated in 50µL of serum-free media for 5 minutes. The siRNA/media mixture was then added dropwise to the transfection reagent mixture, mixed, then incubated for 15 minutes. Cells to be transfected were washed once with PBS, then 900µL serum-free media per well was added. 100µL of siRNA mixture was added dropwise to each well, which was then mixed, and placed in the incubator for 4-6 hours. Serum-free media was replaced with complete media. Reverse transfections were performed 2-3 days after forward transfections. 4µg of siRNA per well was added to 500µL serum-free media. In a separate tube, 8µL of RNAiMAX (Thermo Fiscer) was added to 500µL of serum-free media, and incubated for 5 minutes. The siRNA mixture was then added dropwise onto the RNAiMAX mixture. Cells were then lifted, suspended in 5% FBS media, and counted. 400,000 cells in 2 mL media per well were then added on top of the siRNA mixture. 4-6 hours later, the media was removed, and complete media was added.

**Virus production and transduction**

In order to make virus for transduction, 293T cells were plated in 10cm plates at 1.5x10⁶ cells per plate. The next day, 5µg of the lentiviral plasmid ptfLC3, which contained LC3 fused to a EGFP+RFP tag, was combined with 2.5µg of the pMD2.G (Addgene) envelope plasmid and 2.5µg of the psPAX2 (Addgene) packaging plasmid, into 250µL of DMEM SFM, per 10cm plate. Separately, 30µL of Lipofectamine 2000 (Thermo Fischer) was incubated in 250µL DMEM SFM. After 5 minutes, the lentiviral combination was pipetted onto the Lipofectamine solution
dropwise. This mixture was then incubated for 20 minutes, then added dropwise onto the 293T cells. After 4 hours, the media was changed to 10mL DMEM+15% FBS+0.1% gentamycin. After 3 days, the media was harvested, and passed through a 0.45µm filter, then added to the cell type that would be transduced. Two days later, the media was changed to that which matches the transduced cell type.

**Cellular Function Assays**

For both cell lines, cells were subject to an initial siRNA transfection at 50% confluency in a 6-well plate, followed by a reverse transfection using 400,000 cells 3 days later. 2 days after the reverse transfection, cells were harvested for analysis. For the proliferation assay, Click-iT EdU Alexa Fluor 488 assay (Thermo Fischer) was employed. Cells were incubated with washed with 10µM EdU for 1 hour, then fixed using 3.7% formaldehyde and permeabilized using 0.5% Triton X-100. Alexa Fluor-488 was then added, and flow cytometry was performed on cells to quantify the percent of cells that stained positive for EdU-Alexa Fluor-488 as a readout for proliferation. For cell cycle assay, the same batch of cells used for the proliferation assay were incubated in DAPI for 5 minutes, then flow cytometry for DAPI staining was used to quantify the stages of cell cycle for each cell population. The apoptosis assay was performed using the FITC Annexin V Apoptosis Detection Kit I (BD Biosciences). After being stained with FITC and Annexin V, cells were sorted based on staining to quantify percent of population in early stage and late stage apoptosis. For colony formation assay, 50,000 UOK111 and H1975 cells were transfected with siRNA as described, then transfected again two days later. The next day, the cells
were lifted and counted. They were then plated in a 6-well plate at 1,000 cells/well. They were left to grow for seven days, then fixed with ice-cold methanol for one hour, then stained in 1% crystal violet (Sigma Aldrich) in methanol for 45 minutes. The colonies were then counted by eye using a hemocytometer. Each group was plated in triplicate. For the CARD11 overexpression colony formation assay, 50,000 A498 cells/well were plated in a 6-well plate, then transfected with 1µg of empty vector pcDNA3.1 or pcDNA3.1-CARD11 vector. The next day, the cells were lifted, counted, and plated in a six-well plate at 1,000 cells/well. The group with Torin2 treatment had either DMSO vehicle or 150µM Torin2 added once the cells had attached. The colony formation assay was then conducted in the same manner as described above.

Quantitative Reverse Transcriptase PCR

RNA was extracted from cells using the Direct-zol RNA isolation kit (Zymo). Briefly, cells were washed with PBS, then re-suspended in TRIzol (Thermo Fischer), then run through a column, as instructed by the Direct-zol protocol. RNA was quantified using NanoDrop, then 500ng of RNA was used as a template for cDNA synthesis using the cDNA synthesis kit (Thermo Fischer). 25ng of cDNA was then used for qPCR analysis. Briefly, in each 96 well, 1µL of 100µM primer was added to 5µg cDNA and 5µL SYBR Green PCR Master Mix (Thermo Fischer). This mixture was then placed in the rt-PCR machine (Applied Biosystems), and the following program was undertaken: 50°C 2 minutes, 95°C 10 minutes, (95°C 15 seconds, 60°C 1 minutes) x 40 cycles. The Ct values were then compared, and the ΔΔCt was calculated. This was then used to quantify the relative changes in mRNA expression across samples.
**Autophagic Flux Assay**

Cells were first transfected with siRNA, then subject to reverse transfection, as outlined above. Two days after reverse transfection, 200nM of Bafilomycin A1 (Sigma Aldrich) was added to complete media. Cells were incubated with Bafilomycin A1-containing media for 4 hours, then protein was harvested, and run on an SDS-PAGE gel as outlined above. LC-3B (cleaved vs uncleaved) was determined by probing the membrane with anti-LC3B antibody.

**Live Cell Microscopy**

In order to measure differences in autophagic flux after knockdown of *CARD11*, 100,000 UOK111 cells expressing GFP/RFP-tagged LC3 were transfected with *CARD11* siRNA as listed above, then reverse transfected and plated on a 2-well chamber slide. One day later, cells were imaged using confocal fluorescent microscopy. Eight spots for control and treatment group that were similar in number of cells and fluorescent intensity were selected for imaging. A picture of each region was taken every 15 minutes. After 48h, the pictures were analyzed. Levels of late-stage autophagy were quantified by using ImageJ to quantify the amount of RFP expression within the cells. Additionally, the morphologies of the cells were evaluated by eye.
**In vivo mouse model**

500,000 A498 cells were plated in a 10cm dish. The next day, they were transfected with 7.5µg of either pcDNA3.1 empty vector, or *CARD11*-pcDNA3.1. One day later, 1mg/mL of G418 sulfate was added for selection. The cells were grown in G418 sulfate for 5 days, then 5x10⁶ cells were injected subcutaneously into the right flank of athymic nude mice. The tumors were allowed to grow until the mice became moribund, then were harvested for molecular and immunohistochemical analysis.

**DNA Methylation Analysis**

Cells undergoing methylation analysis had their whole genome DNA isolated using the DNeasy Blood & Tissue kits. 500ng of DNA was then subject to bisulfite treatment using the EZ DNA Methylation-Direct Kit (Zymo). PCR was then performed on the bisulfite-treated DNA using the method as described above. In order to differentiate between methylated and unmethylated CpG dinucleotides, primers for each CpG were designed with their 3’ end falling on the CpG of interest. The primers probing for unmethylated CpG had the 3’ end of their sequence be complementary to TpG, whereas the primers probing for methylated CpG were complementary to CpG, because after bisulfite treatment, all unmethylated CpGs become TpGs. The $\Delta\Delta$Ct of the PCR for each primer were then compared to determine relative levels of methylation in the cell population.

**Overexpression Constructs**
To overexpress the CARD11 natural antisense transcript, the cDNA of the spliced transcript was synthesized and inserted into the pMX vector by GeneArt. This plasmid was amplified in DH5α competent E. coli (Thermo Fischer) by adding 50ng into 50µL bacteria, incubating on ice for 30 minutes, incubating at 42°C for 45 seconds, then put on ice for 2 minutes. 50 µL of LB broth (Thermo Fischer) was then added, and cells were incubated for 1 hour at 37°C. Cells were then spread on an agar plate containing ampicillin. The plate was then incubated overnight, and colonies were picked the next morning. Colonies were grown in 3mL of LB broth for 8 hours, then transferred to 50mL of LB broth for 6 hours. Cells were then spun at 4,000xg for 15 minutes at 4°C, then the plasmids were isolated from the cell pellet using the QIAGEN midiprep kit. Plasmids were suspended in 50µL water, then quantified using NanoDrop (Thermo Fischer). 500ng of plasmid was digested with 2uL of EcoRI (Thermo Fischer) for 45 minutes. The digested product was run on a 1% agarose gel, and the band matching the size of the CARD11-NAT cDNA sequence was cut out and isolated using the QIAquick Gel Extraction Kit (Qiagen). Additionally, 500ng of the PCRII expression vector was cut with EcoRI, run on a gel, and isolated as described above. The cut PCRII vector and CARD11-NAT cDNA were then added together with 1uL T4 DNA ligase for 1 hour at 37°C, then transformed into DH5α cells as described above in order to obtain a PCRII construct with CARD11-NAT cDNA inserted within.

The CARD11 overexpression construct was ordered from Addgene (#44431), as well as the control empty vector pcDNA3.1.
Cas9-Tet1 fusion construct

The nickase-dead dCas9-VP64-T2A-GFP plasmid was ordered from Addgene (plasmid 61422). The VP64 was replaced by the TET1 catalytic domain using the following primers:

hTet1_CD_BamHI_F: ATTAGGATCCCTGCCCACCTGCAGCTGTCTT
hTet1_CD_Nhel_R: ATTAGCTAGCGACCCAATGGTTATAGGGCCCC

The PCR product was amplified using DNA isolated from HT-29 colorectal cells using the method described above, as these cells contain high TET1 expression. The program consisted of 95°C for 2m → (95°C for 15s → 60°C for 15s → 68°C for 45s)x30 cycles → 68°C for 5m. This product as well as the dCas9 plasmid were cut using BamHI and Nhel (Fermentas). Both were run on a 1% agarose gel, and the bands were isolated using a gel purification kit (Qiagen). The products were then incubated together with T4 ligase (NEB) at a 1:3 vector to insert ratio. This was then transformed into competent E.Coli and plated on an agarose plate containing ampicillin. The next day, four colonies were selected and grown in LB broth. After 8 hours, the culture was harvested and the plasmid was isolated using miniprep kit (Qiagen). The plasmid was then checked for proper insertion using the restriction digestion method described above. Once confirmed, the E.coli clone was grown in 50mL LB broth and the plasmid was isolated by midiprep kit (Qiagen). Virus was created from this plasmid using the method described above. H1975 cells were then incubated in viral media for 48h, then were replenished with normal media. After 7 days, these cells were sorted using flow cytometry for GFP expression. Those cells collected from this sort were checked for Cas9-Tet1 fusion protein expression by western using the Cas9 antibody (CST). Short guiding RNAs against the region of interest were selected using the Broad Institute CRISPRko design tool. These sgRNA were cloned using the SAM target sgRNA cloning
protocol. Site-specific methylation was evaluated using the DNA methylation direct kit as described above.

TCGA Tissue Selection

For this analysis, we worked with 10 tissue types with both tumor and normal samples with Illumina HiSeq RNASeqV2 and Illumina HumanMethylation 450k data available at the time of our initial download March 26, 2013 that had no publication restrictions according to the TCGA data policy (http://cancergenome.nih.gov/publications/publicationguidelines). These tumor types include bladder, breast, colon and rectal, head and neck squamous, kidney renal cell, liver, lung adeno, lung squamous, prostate adeno, thyroid, and uterine carcinomas. The corresponding clinical data used for the survival analysis was downloaded from the TCGA data portal current as of January 8, 2014.

Description of the methylation and gene expression data

According to the TCGA description file associated with the Illumina Human Methylation 450K array data, probes having a SNP within 10 nucleotide base pairs (bp) of the interrogated CpG site or having 15 bp from the interrogated CpG site overlap with a REPEAT element are masked as NA across all samples. There are 88058 probes that interrogate such sites (18.3% of all probes). While these beta values are not reported at Level 3, the methylated (M) and un-methylated (U) intensity values for these probes are recorded in the Level 2 data. Therefore, we used the
Level 2 data to reconstruct the beta values for all probes as $M/(M+U)$, also used by TCGA.

For gene expression data, we worked with the log2 transformed Level 3 RNASeqV2 data. To avoid errors for RNASeq raw counts of 0, all values are offset by 1 prior to taking logs.

**Methylation-expression correlation**

All analyses were performed using R, version 2.15.1. Using the complete set of probes targeting CpG dinucleotides, we performed a genome-wide analysis exploring the relationship between the proportion of methylation at various locations within and up to 1,500 bp upstream of a gene and the corresponding log transformed gene expression. We used the Spearman rank statistic to quantify the correlation for each pair. Because we expected these patterns to vary by tissue source site, we calculated coefficients individually using each of the 10 tissue types for which we had data from both the 450k methylation and RNASeq arrays.

**Survival analysis**

Tests for differences in survival were performed by comparing the overall survival of cancer patients in the top and bottom quartiles of 3'UTR methylation using the 'survival' package in R for all genes with correlations >0.5 between gene expression and methylation at the 3'UTR for each tissue type.
Pathway analysis

The pathway enrichment analysis was performed using Netwalker (http://www.netwalkersuite.org) on all genes with correlations >0.5 between gene expression and methylation at the 3'UTR, and <0.5 correlation between gene body methylation and gene expression, for each tissue type. An additional pathway analysis was run using Ingenuity Pathway Analysis (IPA, http://qiagenbioinformatics.com/products/ingenuity-pathway-analysis.com), which yielded similar results.

Network Identification and Construction

The network notes were obtained using Netwalker to determine how many nodes exist, and how the genes interact with one another. The resulting singular network node was identified from this method, and genes that have been shown to be associated with one another, but did not show up within the Netwalker database, were also isolated, then the node and other genes that associate it were exported to Cytoscape (https://www.cytoscape.org/). Connections were then manually included in Cytoscape based on published literature (234).

T cell Isolation, Activation, and Demethylation
The spleens of C57Bl/6 mice were excised, then ground on a 40µm filter while being repeatedly washed with RPMI (Sigma-Aldrich)+10% heat-inactivated FBS (Thermo Fischer Scientific)+1% PenStrep (Thermo Fischer Scientific)+0.01% Beta-Mercaptoethanol (Sigma-Aldrich). The resulting slurry was spun down at 4°C at 450xg for 5 minutes, supernatant was removed, then the pellet was washed with 50mL PBS (HyClone). The pellet was then incubated at room temperature for 5 minutes using 4mL ACK lysis buffer (Gibco), then washed with 50mL PBS and spun down at 4°C at 450xg. The pellet was then re-suspended in the media listed above. To obtain naïve T cells, cells were sent immediately for flow cytometry cell sorting. To activate the T cells, 2µL/mL of mouse CD3e (BD Biosciences) and 3.5 µL/10mL mouse CD28 (BioXCell) activating antibodies were added to the media, along with 1:10,000 100ng/µL mouse recombinant IL-2 (R&D Systems). To induce demethylation, activated T cells were treated with 500nM of 5’deoxy-azacytidine (Sigma-Aldrich) for 72h, or vehicle DMSO control.

**Flow Cytometry**

Spleen cells were grown in the conditions as outlined above were spun down at 450xg at 4°C for 5 minutes to pellet cells. The supernatant was removed, then cells were re-suspended in 2mL FACS buffer (PBS+2%FBS). Cells were then spun at 450xg at 4°C for 5 minutes. Supernatant was again removed, and cells were re-suspended in 200µL FACS buffer + 2µL mouse CD16/CD32 blocking antibody (BD Biosciences) and incubated on ice for 10 minutes. After incubation, 2µL each of mouse CD45-PE, CD4-eFlour 450, and CD8-APC-eFluor 780 (eBiosciences) were added, then cells
were incubated on ice for 30 minutes. All cells that were either CD45+ & CD4+, or CD45+ and CD8+, were sorted by flow cytometry and collected for molecular analysis.

Quantitative Polymerase Chain Reaction

Specimens were collected from flow cytometry, spun down at 450xg for 5 minutes at 4°C, then re-suspended in 350µL TRIzol (Thermo Fischer). RNA was isolated from specimens using the Direct-zol RNA Isolation kit (Zymo), and quantified by NanoDrop. 100ng of RNA was used for a cDNA template. cDNA was created using the Verso cDNA synthesis kit (Invitrogen). 5µg of cDNA was combined with 1µL of 100µM forward and reverse primers, and 5µL of SYBR Green PCR Master Mix (Thermo Fischer) each well. The resulting mixture was then run in a real time PCR machine (Applied Biosystems) using the following program: 50°C 2 minutes, 95°C 10 minutes, (95°C 15 seconds, 60°C 1 minutes) x 40 cycles. The resulting Ct values were then compared, and the ΔΔCt was obtained. This was used to quantify the relative change in mRNA across samples.

Primers

qPCR primers:

<table>
<thead>
<tr>
<th>Havcr2 F</th>
<th>CTCCAAGAACCCTAACCACG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Havcr2 R</td>
<td>AGCCCATGTGGAAATTTTTG</td>
</tr>
<tr>
<td>Primer</td>
<td>Sequence</td>
</tr>
<tr>
<td>--------</td>
<td>----------</td>
</tr>
<tr>
<td>Itk F</td>
<td>TGTCTTAACGAAAGCCAGCC</td>
</tr>
<tr>
<td>Itk R</td>
<td>TCCACACACTTGATTCTGGAG</td>
</tr>
<tr>
<td>Vav1 F</td>
<td>TTCCCCAGGCCATTAATCTT</td>
</tr>
<tr>
<td>Vav1 R</td>
<td>GCCGAACTTTCACAGCA</td>
</tr>
</tbody>
</table>

**Methylation analysis primers:**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>mVAV/Fbio2</td>
<td>AGTTTGTTATTGATGGAAGTTAG</td>
</tr>
<tr>
<td>mVAV/R2</td>
<td>ACATAAAACCAATAATAACCAATAAA</td>
</tr>
<tr>
<td>mVAV/S2</td>
<td>CCAATAAAAACATCCCC</td>
</tr>
<tr>
<td>mITK/Fbio1</td>
<td>TTGTTTATTTTTTTGGGAAATTTT</td>
</tr>
<tr>
<td>mITK/R1</td>
<td>TCCTAAACACACCACAAACCTT</td>
</tr>
<tr>
<td>mITK/S1</td>
<td>CACCACAAACTTCTCT</td>
</tr>
<tr>
<td>mITK/F2</td>
<td>ATTTTTGGGATATTTTAATGGAAGATA</td>
</tr>
<tr>
<td>mITK/Rbio2</td>
<td>TCTATAACACAAAAACACCATTACA</td>
</tr>
<tr>
<td>mITK/S2</td>
<td>GGTGTTTAGTTTTTTTTAA</td>
</tr>
<tr>
<td>mHAVCR2/F1</td>
<td>AGTATTAGGAAGTTGGAGTAGA</td>
</tr>
<tr>
<td>mHAVCR2/Rbio1</td>
<td>TTCCCCCTAAACACACAACAT</td>
</tr>
<tr>
<td>mHAVCR2/S1</td>
<td>TTTTTTATAGAAGAAAAAGGTTT</td>
</tr>
<tr>
<td>mHAVCR2/F2</td>
<td>TGTTTTAAAGGGAATTGTAGATA</td>
</tr>
<tr>
<td>mHAVCR2/Rbio2</td>
<td>ACTATAACAAAATAATAACTAAATCCAC</td>
</tr>
<tr>
<td>mHAVCR2/S2</td>
<td>GGAGTTGTATTGGATAGAA</td>
</tr>
<tr>
<td>mHAVCR2-E/F1</td>
<td>AGGTTTGGTAATGTTGAGTAG</td>
</tr>
<tr>
<td>mHAVCR2-E/Rbio1</td>
<td>ACCTCCAATAAAATTACAAATAACAT</td>
</tr>
<tr>
<td>mHAVCR2-E/S1</td>
<td>AGGATTAGTTTGAGGAAATATT</td>
</tr>
<tr>
<td>mHAVCR2-P/F1</td>
<td>GGGGATTAGAGGTAGAGGAGT</td>
</tr>
<tr>
<td>mHAVCR2-P/Rbio1</td>
<td>AAAAAACCCAAACCCAAACTT</td>
</tr>
</tbody>
</table>
Methylation analysis

One microgram of genomic DNA was treated with sodium bisulfite using the EZ DNA Methylation-Gold Kit (Zymo Research, Irvine, CA) according to the manufacturer’s protocol. The samples were eluted using 40µl of M-Elution Buffer, then 2µl of the eluted product were subjected to PCR analysis. The pyrosequencing analysis as well as the bisulfite conversion were carried out at the DNA Methylation Analysis Core, UT MD Anderson Cancer Center.

The Pyromark Assay Design SW 1.0 (Qiagen) software was used for design of the Havcr2, Itk, and Vav1 primers. These primers were designed using the following parameters: approximately 5bp away from the CpG to be analyzed, a primer with an annealing temperature of within 5°C of 40°C was designed for sequencing. Next, two primers representing the forward and the reverse flanking the sequencing primer were created. To determine the best annealing temperature, gradient PCR was performed. The combined volume totaled to be 20µL, and all 20µL was employed for a single reaction, as described (235). Streptavidin-sepharose high-performance beads (GE Healthcare Life Sciences) were used in purification of the PCR product. The biotinylated products of the PCR reaction and primer used for sequencing
(3.6 pmol per reaction) were then co-denatured using the guide for sample preparation (PSQ96). The sequencing was then carried out using a PyroMark Q96 ID instrument. The reagents used for this reaction are the Pyromark Gold Q96 variety (Qiagen). The PyroMark Q96 software was used to determine the amount of methylation for each analyzed CpG. In each sample, the methylation average, as well as the duplicates, were recorded.

Overarching Hypothesis and Specific Aims

**Hypothesis:** DNA methylation at sites outside the promoter is critical for cancer pathogenesis. Targeting these sites will result in decreased tumorigenesis and tumor progression.

**Specific Aim 1:** To Investigate the relationship between non-promoter DNA methylation and gene expression in genes relevant to cancer

**Specific Aim 2:** To determine the function(s) of relevant sites of non-promoter DNA methylation

**Specific Aim 3:** To elucidate the molecular and cellular consequences of gene body DNA methylation on cancer pathogenesis
Results

Section 3: Identification of methylation sites correlated with gene expression, and investigating the function and regulation of *CARD11* in cancer

Methylation is the most common DNA modification in the genome, and plays an important role in regulating gene expression. This process is a critical means by which cells containing the exact same DNA template can differentiate into a spectacular array of different cell types and functions. The most studied type of DNA methylation is that which occurs on CpG islands within a gene promoter region. This has been shown to powerfully repress transcription through multiple mechanisms, particularly through inducing formation of heterochromatin by recruiting histone modifiers. However, promoter DNA methylation only accounts for a tiny fraction of genomic methylation. Recent studies have begun to unravel the impact of DNA methylation outside of the promoter region. Enhancer, gene body, and intergenic DNA methylation have all been implicated in various forms of gene regulation. Despite these recent advancements, much remains to be understood about region-specific DNA methylation.

Because DNA methylation is essential for numerous cellular functions, aberrant DNA methylation is frequently observed in a huge variety of unrelated diseases. This phenomenon has been implicated in conditions such as neurodegenerative disorders, cardiovascular disease, autoimmune disease, and many others. Most prominently, aberrant DNA methylation is closely tied to cancer
pathogenesis. Since cancer involves a fundamental alteration of a cell’s developmental state, global changes to DNA methylation are ubiquitously observed across nearly every type of cancer. These changes are involved in early phases of tumorigenesis through cellular de-differentiation, suppression of tumor suppressor genes, and enhanced expression or reduced repression of oncogenes. Furthermore, changes to the DNA methylome have been causally linked to metastasis, emergence/maintenance of chemoresistance, and alterations to the tumor microenvironment.

Considering the immense involvement of this epigenetic modification on cancer formation and progression, elucidating how changes to the DNA methylome drive these oncological outcomes is critical for a greater understanding of these deadly diseases. Moreover, investigating how aberrant DNA methylation occurs is crucial for designing novel therapeutic approaches. In particular, gene body methylation has emerged as a major player in cancer pathogenesis. Contrary to promoter DNA methylation, this modification has been shown to promote gene expression by multiple mechanisms, depending on the specific locus and context. For example, DNMT3B-mediated gene body methylation, especially within myc-regulated genes, has been demonstrated to activate oncogenes in colorectal cancer. Despite these recent advances, little is understood about how gene body methylation is intertwined in cancer pathogenesis.

**Subsection 3.1:** The T- and B-cell specific gene *CARD11* has robust correlation between gene expression and gene body methylation in 19 of 32 cancer types.
To address the question of how gene body methylation plays a role in cancer pathogenesis, the following analysis was performed: the correlation between gene body methylation and gene expression of a list of highly likely tumor suppressors and oncogenes (236) across 32 tumor types were examined. Those genes with a correlation coefficient of > 0.5 between gene body methylation and gene expression in at least one tumor type were selected. The gene that matched this criterion in the highest number of tumors, \textit{CARD11}, was found to have highly correlated gene body methylation and gene expression in 19 of the 32 tumor types examined (Fig26.).
Figure 26: Heat map of all high confidence cancer-related genes that have a correlation coefficient of $>$0.5 (represented by red boxes) between gene body methylation and gene expression across 32 cancer types.
Based on this striking observation, the impact of CARD11 was examined in more detail. This gene is expressed in T and B cells, and functions as a scaffold protein that links activation of the T and B cell receptors to the downstream effector pathway NF-kB (237, 238) (Fig 27).

**Figure 27:** CARD11 is phosphorylated after B- and T-cell receptor activation, causing oligomerization and recruitment of Bcl-10/MALT1 complex, which induces degradation of IKK, causing NFkB activation. <Taken with permission from Gastric MALT lymphoma: a model of chronic inflammation-induced tumor development. Sagaert 2010 Nature reviews(3)>
CARD11 is well-known to be a driver of diffuse large B cell lymphoma, in which activating mutations drive constitutive NF-kB activation, leading to pro-cancerous effects(239), but its role in epithelial cancers has never been explored. To address this, the difference in expression of CARD11 between 15 of the most common epithelial cancers and their corresponding normal tissues was evaluated using TCGA datasets. This analysis revealed that CARD11 is significantly overexpressed in 9 of the 15 tumor types, most substantially in kidney renal cell carcinoma (KIRC) and lung adenocarcinoma (LUAD) (Fig28.).

**Figure 28:** Expression of CARD11 in normal versus tumor in 15 tumor types. KIRC and LUAD show the greatest difference in expression when comparing normal and tumor samples.
The association between **CARD11** expression and patient overall survival was then evaluated by determining the Z-score between expression and overall survival. The Z-score tells us how many standard deviations above the mean a group is. In this case, the **CARD11** gene expression of patients that were in the bottom quartile of survivors was compared to the mean. This revealed that high expression is strongly correlated with shorter patient survival in KIRC, with a Z-score of approximately 4 (Fig29.. Of note, BRCA and BLCA showed the opposite effect, indicating tissue-specific effects of **CARD11** expression on overall survival.

**Figure 29:** Z-score of **CARD11** expression in 32 tumor types. KIRC has a z-score of nearly 4, indicating that **CARD11** expression may negatively impact patient overall survival.
Plotting gene expression of *CARD11* on a Kaplan-Meier curve revealed that patients who overexpress this gene had significantly lower overall survival, relative to those who expressed *CARD11* at a low level, in KIRC (Fig30.).

**Figure 30:** Patient overall survival in those patients with the highest 20% of *CARD11* expression versus the lowest 20%. High expression of *CARD11* is associated with shorter overall survival.
These data indicate that expression of *CARD11* is highly correlated with methylation of the gene body, that higher expression is observed in 9 of 15 epithelial tumor types examined, and that overexpression of *CARD11* may negatively impact survival of patients with KIRC.

**Subsection 3.2: *CARD11* regulates autophagy in kidney renal cell carcinoma and lung adenocarcinoma through the mTOR pathway.**

Given the observed overexpression of *CARD11* across numerous epithelial cancers, and considering that expression of *CARD11* is inversely correlated with patient overall survival in KIRC, an investigation into the functional impact of *CARD11* expression in epithelial cancers, as well as the mechanistic underpinning, was undertaken. *CARD11* is an adaptor protein which links stimulation of the T and B cell receptors to activation of the NF-κB pathway. Because *CARD11* is known to impact cellular signaling, an RPPA was conducted in which *CARD11* was knocked down

![Figure 31: Western blot of the knockdown efficiency of CARD11 protein in the samples used for RPPA.](image)
using siRNA in the H1975 lung adenocarcinoma cell line, which contains high expression of \textit{CARD11} (Fig31.).

According to the RPPA, contrary to expectations, the NF-kB pathway remained unchanged after \textit{CARD11} knockdown (Fig32.).
Figure 32: RPPA heatmap comparing siControl vs siCARD11 (bottom 3 rows) in H1975 lung adenocarcinoma cells that shows upregulated (red) or downregulated (green) genes.
To determine which pathway was most impacted by knockdown of *CARD11*, the changes to protein expression and phosphorylation detected by RPPA were analyzed using the NetWalker platform. This demonstrated that the pathway most affected by knockdown was the mTOR pathway (Table 1).

![Table 1: NetWalker analysis of the top pathways affected by CARD11 knockdown in H1975 lung adenocarcinoma cells.](image)
The mTOR pathway as the control center of cell growth, can affect a multitude of cellular processes, ranging from metabolism, to transcription, to autophagy(240). Therefore, the impact of CARD11 on the functions involved in cancer pathogenesis were measured. To find the most optimal renal cell carcinoma cell line for studying the effects of CARD11, a panel of renal cell carcinoma lines were probed for CARD11 protein expression by western blot. This revealed that UOK111 contained the highest expression of CARD11 among these cell lines (Fig33.); therefore, UOK11 was selected for further analysis.

![Image of western blot showing CARD11 and GAPDH expression in different renal cell carcinoma lines](image)

**Figure 33:** Cell panel of 8 KIRC lines to determine which cell line contained the highest expression of CARD11.
CARD11 has been shown to be necessary for proper mTOR activation in lymphocytes (188), so to confirm this role in epithelial cancers as well, H1975 cells and UOK111 renal cell carcinoma cells were transfected with CARD11 siRNA or control siRNA. S6, a downstream node of the mTOR pathway that showed decreased activation at both major phosphorylation sites from the RPPA, was probed for phosphorylation as a measure of activation. After CARD11 knockdown, the amount of phosphorylated S6 decreased in both UOK111 and H1975 cell lines (Fig34.), validating the finding that the mTOR pathway is positively regulated by CARD11.
Figure 34: The phosphorylation state of the S6 ribosomal protein, a downstream component of the mTOR pathway that was shown by RPPA to be decreased after \textit{CARD11} knockdown.
UOK111 and H1975 cells were subject to proliferation and cell cycle assays after CARD11 knockdown, however, no significant change in either of these processes was observed (Fig 35).

Figure 35: UOK111 (top) and H1975 (bottom) showed no significant difference in cell cycle (left) or proliferation (right) after CARD11 knockdown.
Next, the rate of apoptosis was determined after CARD11 knockdown, but there was no difference between control siRNA and CARD11 siRNA groups. Finally, the ability of the cells to establish single-cell colonies after CARD11 knockdown was measured. Both UOK111 and H1975 cell lines showed a significantly impaired capability of forming colonies after CARD11 knockdown (Fig36.).
Figure 36: UOK111 (top) and H1975 (bottom) showed no significant difference in apoptosis (left) after CARD11 knockdown; however, both cell lines exhibited significantly diminished numbers of colonies formed (right) after CARD11 knockdown.
Conversely, the KIRC cell line A498, which does not express \textit{CARD11}, was transfected with a \textit{CARD11} overexpression plasmid. This resulted in a nearly 4-fold increase in colony formation. Furthermore, treating the A498 cells with Torin2, an mTOR inhibitor, abrogated the increase in colony formation of the \textit{CARD11} overexpressing line, relative to the empty vector-transfected line (\textbf{Fig37.}). This supports the conclusion that \textit{CARD11}-mediated mTOR activation is important for the observed capacity of cancer cells to form colonies.

\textbf{Figure 37:} (Top) \textit{CARD11} protein expression increases pS6, and treatment with 50nM Torin2 abrogates S6 phosphorylation (Left) \textit{CARD11} protein overexpression confirmed by western blot. (Right) Overexpression of \textit{CARD11} in A498 KIRC cells resulted in a three-fold increase in number of colonies, whereas treating with mTOR inhibitor Torin2 reduced the difference in number of colonies formed between empty vector and \textit{CARD11} overexpressing groups. EV= empty vector; CARDOX=\textit{CARD11} overexpression.
mTOR activation negatively regulates autophagy during situations in which the cell has access to nutrients and can grow and divide. However, autophagy is activated in stressful situations such as when isolated as single cells such as in a colony formation assay, to allow the cells to survive. Furthermore, diminished autophagy is known to play a role in renal cell carcinoma pathogenesis (241). Based on this rationale, expression of CARD11 was hypothesized to decrease autophagy. To test this hypothesis, the level of autophagy in UOK111 cells after CARD11 knockdown was determined using lipidated LC3 as a marker. CARD11 knockdown resulted in increased amounts of lipidated LC3 in both cell lines tested (Fig38.).

Figure 38: Knockdown of CARD11 results in increased lipidated LC3 in UOK111 (left) and H1975 (right) cells.
Because an increase in lipidated LC3 can be due to increased autophagy, or to decreased fusion of the autophagosome with the lysosome, the fusion inhibitor bafilomycinA1 was employed to ascertain the cause of increased lipidated LC3. If fusion was compromised, the addition of bafilomycinA1 should not result in any change in lipidated LC3 after CARD11 knockdown. However, after adding bafilomycinA1, the amount of lipidated LC3 increased substantially in CARD11 siRNA cells, indicating that knocking down CARD11 causes increased autophagy in UOK111 (Fig 39.).

Figure 39: In UOK111, knockdown of CARD11 causes increased lipidated LC3 (bottom band), indicating an increase in autophagy. Treatment with Bafilomycin further increases lipidated LC3, indicating that autophagic flux has increased after CARD11 knockdown, rather than a decreased ability for autophagosomes to fuse with lysosomes.
Similar results were observed in H1975 LUAD cells, in which knockdown of *CARD11* caused an increase in lipidated LC3, and treatment with bafilomycin in this context resulted in even greater accumulation of lipidated LC3, suggesting that the impact of *CARD11* knockdown on autophagic flux was not limited to a singular tumor type (Fig 40.).

**Figure 40:** In H1975, knockdown of CARD11 causes increased lipidated LC3 (bottom band), indicating an increase in autophagy. Treatment with Bafilomycin further increases lipidated LC3, indicating that autophagic flux has increased after CARD11 knockdown, rather than a decreased ability for autophagosomes to fuse with lysosomes.
To further test this hypothesis, a GFP/RFP-tagged LC3 protein was expressed in UOK111 cells, allowing for visualization of autophagic flux. When the autophagosome forms, the GFP fluorescence of LC3 is converted into RFP. Therefore, the red punctates within a cell represent the level of late-stage autophagy. After transfecting UOK111 cells, they were sorted for expression of GFP. The cells received from this sort were then subject to transfection of control siRNA or CARD11 siRNA. Two days later, they were again transfected, and plated in a slide that allows for visualization. One day later, the autophagy levels of the cells were visualized using live cell confocal microscopy. In order to best compare these two groups, two regions in which both groups had low RFP expression were selected to record. Cells were then recorded for 48 hours, and the changes to RFP expression and localization, as well as the cell morphology, were examined. The control siRNA-treated cells remained healthy, and showed low RFP-labeled punctates, whereas the CARD11 siRNA-treated cells had high numbers of RFP-labeled punctates form, and showed high cell death after late-stage autophagy formation (Fig41.). This strongly supports the conclusion that CARD11 suppresses autophagy, and that knockdown of CARD11 results in increased autophagic flux.
**Figure 41:** UOK111 cells were stably transfected with LC3-GFP/RFP as a measure of autophagic flux. RFP indicates that cells are undergoing late-stage autophagy. Knockdown of *CARD11* resulted in a much greater proportion of cells to undergo late-stage autophagy.
Subsection 3.3: CARD11 Natural Antisense Transcript does not regulate CARD11 expression.

As noted, CARD11 gene body methylation is strongly correlated with gene expression. Therefore, to ascertain whether this modification is a driving factor in overexpression of CARD11 in epithelial cancers, gene body methylation in tumor samples or corresponding normal tissues was compared, revealing an increase in gene body methylation of CARD11 in KIRC, relative to normal tissues (Fig 42). However, this increase was not statistically significant.

![Figure 42](image-url)

Figure 42: Average of the top 10 most significantly CARD11 gene body methylation sites between normal and tumor.
Treating UOK111 cells with decitabine resulted in a decrease in *CARD11* expression, supporting the notion that methylation increases its expression (Fig43.).

**Figure 43:** Treatment of UOK111 cells with decitabine resulted in a 30% decrease in *CARD11* expression.
To understand how gene body methylation may be positively regulating \textit{CARD11} gene expression, the gene locus was analyzed using Santa Cruz Genome Browser. This revealed an unexplored natural antisense transcript (NAT) that falls within the last 2/3rds of the gene body (Fig 44.).

\textbf{Figure 44:} \textit{CARD11} gene layout (taken from Ensembl). \textit{CARD11} gene body contains a natural antisense transcript (red box) composed of two exons and one intron.
Because NATs have been shown to commonly mediate site-specific DNA methylation, as well as to regulate expression of their corresponding sense transcript (242), this NAT was selected for further analysis. NATs can either positively or negatively impact sense gene expression (242); therefore, the first step was to investigate the correlation between CARD11 and CARD11-NAT. This revealed a robust positive correlation between this pair in every tissue type examined, including in KIRC and LUAD (Fig45., Table 2).

**Figure 45:** CARD11 and its NAT are robustly correlated in both KIRC and LUAD.
### Table 2: The correlation between *CARD11* sense and NAT across 19 tumor types.

<table>
<thead>
<tr>
<th>Tissue Type</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLCA</td>
<td>0.84</td>
</tr>
<tr>
<td>BRCA</td>
<td>0.75</td>
</tr>
<tr>
<td>CESC</td>
<td>0.70</td>
</tr>
<tr>
<td>GBM</td>
<td>0.65</td>
</tr>
<tr>
<td>HNSC</td>
<td>0.81</td>
</tr>
<tr>
<td>KICH</td>
<td>0.77</td>
</tr>
<tr>
<td>KIRC</td>
<td>0.75</td>
</tr>
<tr>
<td>KIRP</td>
<td>0.82</td>
</tr>
<tr>
<td>LAML</td>
<td>0.90</td>
</tr>
<tr>
<td>LGG</td>
<td>0.69</td>
</tr>
<tr>
<td>LIHC</td>
<td>0.62</td>
</tr>
<tr>
<td>LUAD</td>
<td>0.80</td>
</tr>
<tr>
<td>LUSC</td>
<td>0.79</td>
</tr>
<tr>
<td>OV</td>
<td>0.83</td>
</tr>
<tr>
<td>PAAD</td>
<td>0.65</td>
</tr>
<tr>
<td>PRAD</td>
<td>0.71</td>
</tr>
<tr>
<td>SKCM</td>
<td>0.81</td>
</tr>
<tr>
<td>STAD</td>
<td>0.79</td>
</tr>
<tr>
<td>THCA</td>
<td>0.83</td>
</tr>
</tbody>
</table>
Furthermore, \textit{CARD11-NAT} expression across 11 tissue types was evaluated using TCGA datasets. This revealed that \textit{CARD11-NAT} was overexpressed in 5 of the 11, with KIRC and LUAD having the most significant difference in expression between normal and tumor (Fig 46).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{card11_nat_expression.png}
\caption{Comparing expression in tumor versus normal of \textit{CARD11-NAT} across 12 tissue types. \textit{CARD11-NAT} shows greatest difference in expression between normal and tumor in KIRC and LUAD.}
\end{figure}
Therefore, it was hypothesized that \textit{CARD11-NAT} is positively regulating expression of \textit{CARD11} by promoting methylation of the \textit{CARD11} gene body. To test this hypothesis, TCGA samples for KIRC and LUAD were segregated based on high \textit{versus} low expression of \textit{CARD11-NAT}, and \textit{CARD11} methylation was compared between the two groups. This revealed that of 79 methylation sites examined, 31 showed a significant difference in methylation in KIRC, and 63 showed a significant difference in LUAD. Within the gene body, 23 out of 25 loci in KIRC with differential methylation showed higher methylation in the samples with high \textit{CARD11-NAT}. Likewise, 54 of 55 loci in LUAD with differential methylation showed higher methylation in the samples with high \textit{CARD11-NAT} (\textbf{Table 3}). In particular, the probes displaying the most significant difference in methylation were located at a CpG island within the \textit{CARD11} gene body (Fig47.).
<table>
<thead>
<tr>
<th></th>
<th>KIRC</th>
<th>LUAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total (CARD11) DNA methylation probes</td>
<td>79</td>
<td>79</td>
</tr>
<tr>
<td># Significant difference in methylation based on (CARD11)-(NAT) expression</td>
<td>31 (35%)</td>
<td>63 (80%)</td>
</tr>
<tr>
<td># Significantly differentially methylated loci within the (CARD11) gene body</td>
<td>25</td>
<td>55</td>
</tr>
<tr>
<td># of probes showing increased methylation in (CARD11)-(NAT) High group</td>
<td>23/25</td>
<td>54/55</td>
</tr>
</tbody>
</table>

**Table 3:** The difference in \(CARD11\) gene body methylation after segregating samples based on high versus low expression of \(CARD11\)-\(NAT\).
Figure 47: (Top) CARD11 layout (taken from Santa Cruz Genome Browser) showing a CpG island (red box) located approximately 2/3rds down the gene body.

(Bottom) Expression level of CARD11-NAT (T317639) in relation to CpG island methylation. In both KIRC and LUAD, methylation of this region is associated with higher expression of CARD11-NAT.
To determine whether the correlation between high \textit{CARD11-NAT} and high \textit{CARD11} gene body CpG island methylation is due to \textit{CARD11-NAT} inducing DNA methylation, or \textit{vice versa}, \textit{CARD11-NAT} was knocked down using siRNA, then the methylation of the \textit{CARD11} CpG island was measured. This showed that knockdown of \textit{CARD11-NAT} had no effect on the methylation of this locus (Fig48.).

\textbf{Figure 48}: Knockdown of \textit{CARD11-NAT} did not impact the level of methylation of the \textit{CARD11} gene body CpG island.
Next, specific demethylation of the CpG island was induced by expressing a fusion protein consisting of Cas9 combined with the catalytic domain of Tet1, the enzyme responsible for DNA demethylation. Two guide RNAs were designed to target this fusion protein to the CpG island, and the Cas9-Tet1 fusion protein along with the sgRNAs were transduced into H1975 cells by lentiviral transduction to induce stable demethylation of this locus (Fig49.).

**Figure 49:** (Left) illustration of the Cas9-Tet1 fusion protein that allows for specific demethylation of a locus based on guide RNA binding. (Right) expression of the Cas9-Tet1 protein can be detected by western blot.
This resulted in a significant reduction of both CARD11 and CARD11-NAT expression, strongly indicating that CARD11-NAT is downstream of CARD11 CpG island methylation (Fig50.).

**Figure 50:** (Top) CARD11 CpG island methylation after specific demethylation. (Bottom) Expression of CARD11 (left) and CARD11-NAT (right) both exhibited a substantial decrease in expression after specific demethylation of the CARD11 gene body CpG island induced by Cas9-Tet1.
To determine whether the corresponding decrease in CARD11-NAT was the cause of decreased CARD11 expression after site-specific demethylation, CARD11-NAT was overexpressed, and the expression of CARD11 was quantified by qPCR. This showed no difference in CARD11 expression after overexpression of CARD11-NAT, suggesting that the methylation of this region is regulating CARD11 expression through an alternate mechanism (Fig51.).

**Figure 51:** Expression of CARD11-NAT (left) and CARD11 (right) after CARD11-NAT overexpression. Higher CARD11-NAT did not impact the expression of CARD11 in this context.

Therefore, this data discounts the hypothesis that the NAT regulates CARD11, and that the decrease in CARD11 after demethylation is due to decreased NAT expression.
One alternate hypothesis that is currently under investigation is whether the intragenic CpG island is the location of a cryptic alternative promoter. Recent studies have identified a wide number of intragenic CpG islands as being methylated as a means of repressing aberrant downstream transcription initiation\((243)\). In this context, gene body CpG island methylation results in increased transcription of the full-length mRNA; loss of methylation causes transcription to occur downstream, thereby inhibiting expression of the full length transcript. Furthermore, the \textit{CARD11} CpG island contains multiple binding sites for SP1, a transcription factor that is sensitive to DNA methylation. To test this hypothesis, the H3K9ac status of this region will be tested, which indicates when a specific site adopts the chromatic architecture of an actively transcribing promoter. Furthermore, the binding of SP1 to \textit{CARD11} gene body CpG island will be evaluated after demethylation.
Section 4: 3’UTR DNA methylation is a functionally relevant epigenetic modification distinct from gene body methylation, and plays a role in regulating T cell activation genes.

Subsection 4.1: DNA methylation probes positively correlated with increased gene expression are enriched in 3’UTRs.

The Cancer Genome Atlas (TCGA) provides a unique opportunity to harness large-scale molecular profiling datasets across numerous tumor types and over 11,000 patient samples (244). In order to explore the relationship between poorly understood sites of DNA methylation and gene expression, genome-wide RNA-seq and methylation array datasets from this platform were utilized. Initially, the proportion of probes achieving negative Spearman correlations of < -0.5 between gene expression and DNA methylation were examined within each of the 6 gene regions included in the Illumina methylation probe annotation after normalizing for the total number of probes interrogating each region.

We note that the sample sizes associated with TCGA are such that these correlations are essentially certain to be real (p <0.0004, assuming n=50). With sample sizes per tissue in excess of a few hundred, absolute correlations > 0.5 were considered highly significant for this analysis. Approximately a third of the strong
associations were positive (at the probe level there were 44,309 negative and 29,043 positive associations; at the gene level there were 6,287 negative and 3,200 positive associations). As expected, the majority (> 75%) of the negatively correlated probes across all 10 tissue types were concentrated within the first exon, 5’-UTR, and upstream of the transcription start site (Fig 52.).

**Figure 52:** Proportion heat map representing the distribution of probes in which the correlation between methylation and gene expression was < -0.5 across 10 tissue types.
Upon segregating the positively correlated probes based on region, it was noted that approximately 3% of these probes were interrogating the 3'UTR. Gene body DNA methylation is a known feature of highly transcribed genes (245); however, previous studies have grouped the 3'UTR along with the rest of the gene body, rather than investigating it as a distinct region. To determine whether there existed a statistical rationale to separate the 3'UTR from the gene body, the proportion of probes within the 3'UTR that exhibited a positive correlation of > 0.5 between DNA methylation and gene expression was calculated (Fig 53.).

**Figure 53:** Proportion heat map representing the distribution of probes in which the correlation between methylation and gene expression was > 0.5 across 10 tissue types.
This revealed a substantial net enrichment in the proportion of positively correlated probes in the 3'UTR when compared to the entire gene region across all 10 tissue types (Fig54.), with up to 590 genes exhibiting a correlation coefficient of > 0.5 between 3'UTR methylation and gene expression (Table 4).
<table>
<thead>
<tr>
<th>BLCA</th>
<th>Correlation Coefficient</th>
<th>BRCA</th>
<th>Correlation Coefficient2</th>
<th>CORE</th>
<th>Correlation Coefficient3</th>
</tr>
</thead>
<tbody>
<tr>
<td>AASS</td>
<td>0.596732</td>
<td>ABLIM3</td>
<td>0.6087407</td>
<td>ABCA3</td>
<td>0.5134683</td>
</tr>
<tr>
<td>ABCA2</td>
<td>0.5940652</td>
<td>ACE2</td>
<td>0.5152063</td>
<td>ABLIM3</td>
<td>0.5071737</td>
</tr>
<tr>
<td>ABCA3</td>
<td>0.7429512</td>
<td>ADCY2</td>
<td>0.5784943</td>
<td>ALOX5</td>
<td>0.5318469</td>
</tr>
<tr>
<td>ABLIM2</td>
<td>0.5834121</td>
<td>AGAP2</td>
<td>0.6427411</td>
<td>ANKR5</td>
<td>0.5128174</td>
</tr>
<tr>
<td>ABLIM3</td>
<td>0.639832</td>
<td>AMH</td>
<td>0.5517516</td>
<td>BAHCC1</td>
<td>0.6107377</td>
</tr>
<tr>
<td>ACOXL</td>
<td>0.6250692</td>
<td>ANO5</td>
<td>0.6308839</td>
<td>C10orf72</td>
<td>0.5160956</td>
</tr>
<tr>
<td>ACSS3</td>
<td>0.5256172</td>
<td>ASAM</td>
<td>0.5593502</td>
<td>C17orf77</td>
<td>0.5037266</td>
</tr>
<tr>
<td>ADAM33</td>
<td>0.520791</td>
<td>ASB2</td>
<td>0.5050138</td>
<td>C2CD4C</td>
<td>0.5362058</td>
</tr>
<tr>
<td>ADAMTS17</td>
<td>0.5905777</td>
<td>ASPH</td>
<td>0.5151348</td>
<td>C6orf97</td>
<td>0.5073339</td>
</tr>
<tr>
<td>ADAMTS1L3</td>
<td>0.615171</td>
<td>ATP2B2</td>
<td>0.5260176</td>
<td>CADM3</td>
<td>0.5612771</td>
</tr>
<tr>
<td>ADAMTS1L5</td>
<td>0.603206</td>
<td>B3GAT1</td>
<td>0.5220495</td>
<td>CARD11</td>
<td>0.702967</td>
</tr>
<tr>
<td>ADARB1</td>
<td>0.514069</td>
<td>BACH2</td>
<td>0.5025892</td>
<td>CD1D</td>
<td>0.5386125</td>
</tr>
<tr>
<td>ADCY2</td>
<td>0.6243769</td>
<td>BAHCC1</td>
<td>0.5476259</td>
<td>CDO1</td>
<td>0.5375716</td>
</tr>
<tr>
<td>ADCY9</td>
<td>0.6462192</td>
<td>BCL11A</td>
<td>0.538716</td>
<td>CPE</td>
<td>0.6422394</td>
</tr>
<tr>
<td>ADD2</td>
<td>0.5051489</td>
<td>BIVM</td>
<td>0.5357103</td>
<td>CPLX2</td>
<td>0.5389136</td>
</tr>
<tr>
<td>ADH1B</td>
<td>0.5186804</td>
<td>BMP2</td>
<td>0.6188118</td>
<td>CRIP2</td>
<td>0.5943179</td>
</tr>
<tr>
<td>ADORA1</td>
<td>0.5027623</td>
<td>BOC</td>
<td>0.5102699</td>
<td>CSNK2A1</td>
<td>0.5498119</td>
</tr>
<tr>
<td>ADPRH</td>
<td>0.5103907</td>
<td>C10orf128</td>
<td>0.571714</td>
<td>CTXN1</td>
<td>0.542705</td>
</tr>
<tr>
<td>ADRA1A</td>
<td>0.5428149</td>
<td>C10orf90</td>
<td>0.5685527</td>
<td>CUL4A</td>
<td>0.5009405</td>
</tr>
<tr>
<td>AJAP1</td>
<td>0.5437472</td>
<td>C11orf21</td>
<td>0.5128085</td>
<td>DCDC2</td>
<td>0.6050958</td>
</tr>
<tr>
<td>ALDH1L2</td>
<td>0.514923</td>
<td>C11orf63</td>
<td>0.5548425</td>
<td>DCLK2</td>
<td>0.5045169</td>
</tr>
<tr>
<td>ALDH2</td>
<td>0.591384</td>
<td>C17orf51</td>
<td>0.6420734</td>
<td>DDX3Y</td>
<td>0.8002091</td>
</tr>
<tr>
<td>AMOT</td>
<td>0.6702877</td>
<td>C17orf75</td>
<td>0.5039303</td>
<td>DGK6</td>
<td>0.5271233</td>
</tr>
<tr>
<td>ANOS</td>
<td>0.5600789</td>
<td>C19orf18</td>
<td>0.5528789</td>
<td>DOCK2</td>
<td>0.5898943</td>
</tr>
<tr>
<td>APBB1P</td>
<td>0.5601641</td>
<td>C20orf160</td>
<td>0.550029</td>
<td>DOK2</td>
<td>0.5577691</td>
</tr>
<tr>
<td>ARHGAP28</td>
<td>0.7179572</td>
<td>C8orf79</td>
<td>0.6529395</td>
<td>EIF1AY</td>
<td>0.7646542</td>
</tr>
<tr>
<td>ARMC4</td>
<td>0.7332187</td>
<td>C9orf125</td>
<td>0.5540748</td>
<td>EMP3</td>
<td>0.5646654</td>
</tr>
<tr>
<td>ARSJ</td>
<td>0.67742</td>
<td>CA8</td>
<td>0.5158163</td>
<td>FAM189A1</td>
<td>0.624909</td>
</tr>
<tr>
<td>ASAM</td>
<td>0.6373837</td>
<td>CACNA1H</td>
<td>0.5302311</td>
<td>FBLN2</td>
<td>0.5505962</td>
</tr>
<tr>
<td>ASB12</td>
<td>0.5004157</td>
<td>CADM3</td>
<td>0.6157453</td>
<td>FGF9</td>
<td>0.5290764</td>
</tr>
<tr>
<td>ASB2</td>
<td>0.5098946</td>
<td>CAMK2D</td>
<td>0.5295647</td>
<td>FSCN1</td>
<td>0.5490571</td>
</tr>
<tr>
<td>ATP10A</td>
<td>0.6582176</td>
<td>CBFA2T3</td>
<td>0.5951633</td>
<td>GABBR1</td>
<td>0.5036603</td>
</tr>
<tr>
<td>ATP10B</td>
<td>0.5056743</td>
<td>CBLN2</td>
<td>0.5188389</td>
<td>GNA15</td>
<td>0.546698</td>
</tr>
<tr>
<td>AXL</td>
<td>0.537961</td>
<td>CDK6</td>
<td>0.5623088</td>
<td>GPR124</td>
<td>0.5373269</td>
</tr>
</tbody>
</table>

Table 4: Partial list of correlation between 3'UTR methylation and gene expression by tissue type.
This prompted a closer look at how methylation of the 3’UTR relates to gene expression, particularly whether this is a passenger modification and its association with gene expression could be attributed to another variable, or whether levels of 3’UTR methylation are the key differentiators of how highly genes are expressed.

Subsection 4.2: Differences in expression between normal and cancer unexplained by promoter methylation or copy number variation.

To examine whether other drivers of gene expression more suitably explained differences in gene expression, we focused on HOXC13 for several reasons, including a strong positive correlation (> 0.5 between 3’UTR methylation and gene expression in all 10 examined tissues), a comprehensive set of probes interrogating each region of the gene, and significantly higher expression in tumor tissues than in corresponding normal tissues (Fig55.).
Figure 55: Dotplot of log2 RNA-seq expression values across 10 TCGA tissue samples; black stars indicate normal tissue samples. The gene was overexpressed in tumors compared with in the corresponding normal samples.
Furthermore, *HOXC13* exhibits significantly elevated levels of methylation at the 3'UTR (all p-values < 0.001) when comparing tumors to normal samples in bladder, breast, colorectal, head and neck, lung, and uterine tissue types, suggesting that 3'UTR methylation may be a primary driver of this differential expression. To further investigate the nature of this variation, we examined common processes known to account for differential gene expression, primarily promoter methylation and copy number variation. First, we addressed possible allelic gain or loss using the TCGA copy number data for these tissue types (*Fig56.*).

**Figure 56:** Copy number alterations in HOXC13 for the same samples. There is no evidence of copy number gain or loss in HOXC13.
There was no significant *HOXC13* gene amplification or deletion across all 10 tumor types. Specifically, $< 5\%$ of cases had any reported copy number alterations in the *HOXC13* gene region, making it unlikely that copy number is driving the observed variation in expression.

We also considered whether the difference in *HOXC13* expression between normal and tumor tissues could be due to divergent promoter methylation. Two tumor types, lung squamous and bladder cancer were selected as representative examples, as they demonstrated the greatest variation in gene expression between tumor and normal samples (Fig57.).

Less than 10% of samples in the lowest quartile of gene expression have any evidence of methylation in the promoter region of *HOXC13*, strongly suggesting that differences in promoter methylation cannot account for the observed differences in expression. Next, we characterized the 3'UTR methylation pattern of high and low *HOXC13* expressers and discovered that $> 90\%$ of samples in the highest quartile of
expression have methylated 3'UTRs, further implicating 3'UTR methylation as a potential driver of expression.

**Figure 57**: Heatmaps representing the methylation patterns across HOXC13 in bladder and lung squamous cell carcinomas. The row color bar depicts the gene structure: purple, promoter region; light green, 5'UTR; dark green, first exon; grey, body; red, 3'UTR. The column color bar represents the top 25% gene expressers in pink and the bottom 25% expressers in black.
Subsection 4.3: Extent of 3’UTR methylation can predict overall survival.

Considering the genome-wide changes in methylation patterning exhibited by tumors, and the clinically actionable nature of these modifications due to the reversible nature of DNA methylation\((246)\), genome-wide differences in 3’UTR methylation between tumor and normal tissues in genes with a > 0.5 correlation coefficient between 3’UTR methylation and gene expression were examined. Because of observed divergence in 3’UTR methylation between tumor and normal tissues in certain genes, it was hypothesized that these genes might play roles in tumorigenesis and progression; therefore, we sought to determine whether these genes track with overall survival. Within five of the tumor types, including head and neck, lung adenocarcinoma, lung squamous, bladder, and kidney renal cell, we saw significant changes in survival associated with methylation at the 3’UTR (Table 5).

<table>
<thead>
<tr>
<th>Tissue</th>
<th>BLCA</th>
<th>BRCA</th>
<th>CORE</th>
<th>HNSC</th>
<th>KIRC</th>
<th>LUAD</th>
<th>LUSC</th>
<th>PRAD</th>
<th>THCA</th>
<th>UCEC</th>
</tr>
</thead>
<tbody>
<tr>
<td># significant</td>
<td>35</td>
<td>7</td>
<td>0</td>
<td>20</td>
<td>22</td>
<td>17</td>
<td>11</td>
<td>0</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td># expected</td>
<td>17.15</td>
<td>10.1</td>
<td>1.3</td>
<td>7.3</td>
<td>1.9</td>
<td>5.9</td>
<td>6.25</td>
<td>5.65</td>
<td>1.9</td>
<td>3.8</td>
</tr>
<tr>
<td>total</td>
<td>343</td>
<td>202</td>
<td>26</td>
<td>146</td>
<td>38</td>
<td>118</td>
<td>125</td>
<td>113</td>
<td>38</td>
<td>76</td>
</tr>
<tr>
<td>p-value</td>
<td>6.25E-05</td>
<td>0.882</td>
<td>1</td>
<td>4.38E-05</td>
<td>&lt;2.2e-16</td>
<td>8.43E-05</td>
<td>0.049</td>
<td>1</td>
<td>0.573</td>
<td>0.729</td>
</tr>
</tbody>
</table>

**Table 5:** The first row of the table represents the total number of differentially methylated genes that had significant separation in overall survival \((p < 0.05)\). The second column represents the number of genes we would have expected to be significantly associated with survival on the basis of Poisson distribution if the genes had been chosen at random. The final row summarizes the total number of differentially methylated genes that were positively correlated with a 3’UTR > 0.5.
Using kidney renal cell carcinoma (KIRC) as an example, of the 38 genes with differential 3'UTR methylation between normal and cancer, 22 showed significant separation in overall survival at a p-value of 0.05. For example, methylation at the 3'UTR of the myosin 1G (MYO1G) gene in KIRC is highly correlated with gene expression, and patients with the lowest 20% methylation levels have significantly better survival (p<0.001) than those with the highest 20% (Fig58.).

**Figure 58:** Differences in patient overall survival on the basis of methylation status using an example gene, MYO1G, in the context of kidney renal clear cell carcinoma. The red and blue points in the dotplot (B, left panel) distinguish the samples with the highest and lowest 20% of methylation, respectively. (right panel) The same samples were used to evaluate the difference in overall survival between these groups.
This proportion is significantly greater than the 2 genes expected to be associated with survival had 38 genes been selected at random. In summary, extent of 3’UTR methylation in a significant number of genes is associated with overall survival in 5 of 10 tumor types examined, suggesting that 3’UTR methylation may be a functionally relevant modification in cancer pathogenesis.

Subsection 4.4: Genes in which expression is uniquely correlated with 3’UTR DNA methylation are enriched for those involved in T cell activation.

Because the majority of genes from this analysis exhibit a positive correlation between gene body methylation and expression, those genes in which 3’UTR methylation had a > 0.5 correlation coefficient with gene expression and < 0.5 in the gene body were selected for further analysis. The most exemplary gene exhibiting this phenomenon, ITK, contains robust positive correlation in the 3’UTR, but negative correlation in the gene body (Fig59.).
Figure 59: A representative plot of the relationship between methylation and gene expression on the ITK gene body. The top box represents the gene layout, with the blue portion highlighting the 3'UTR. In the “correlations” box, each line represents a tissue type.
Filtering using these criteria yielded a list of 156 genes (Table 6).

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ABLM3</td>
<td>CRTAM</td>
<td>GIMP4</td>
<td>IGSF9B</td>
<td>MICA</td>
<td>PKHD1</td>
<td>SLC7A4</td>
</tr>
<tr>
<td>2</td>
<td>ADAMTSL3</td>
<td>CYP4V2</td>
<td>GLYAT1</td>
<td>IKZF1</td>
<td>MRAP2</td>
<td>PLCL1</td>
<td>SLFN5</td>
</tr>
<tr>
<td>3</td>
<td>ASAM</td>
<td>DLGAP2</td>
<td>GMFG</td>
<td>IRF4</td>
<td>MSN</td>
<td>PNMA2</td>
<td>SPATA6</td>
</tr>
<tr>
<td>4</td>
<td>ASB2</td>
<td>DLX3</td>
<td>GNA15</td>
<td>ISL2</td>
<td>MTMR9</td>
<td>POU4F1</td>
<td>ST6GALNA</td>
</tr>
<tr>
<td>5</td>
<td>B3GNT3</td>
<td>DOK2</td>
<td>GPR114</td>
<td>ITK</td>
<td>MUM1L1</td>
<td>PTPLA</td>
<td>STC2</td>
</tr>
<tr>
<td>6</td>
<td>BNC1</td>
<td>E2F7</td>
<td>GPR55</td>
<td>KBTBD11</td>
<td>MXRA8</td>
<td>RAB11FIP4</td>
<td>SYT7</td>
</tr>
<tr>
<td>7</td>
<td>BTC</td>
<td>EDA2R</td>
<td>GREM2</td>
<td>KCNA1</td>
<td>NF1K</td>
<td>RANBP17</td>
<td>TBX18</td>
</tr>
<tr>
<td>8</td>
<td>C10orf128</td>
<td>EMI1</td>
<td>GSTO2</td>
<td>KCNC3</td>
<td>NKX2-8</td>
<td>RASGEF1C</td>
<td>TCF15</td>
</tr>
<tr>
<td>9</td>
<td>C13orf18</td>
<td>FAM110B</td>
<td>GSTT1</td>
<td>KCNG1</td>
<td>NOV</td>
<td>RIN3</td>
<td>TCTEX1D1</td>
</tr>
<tr>
<td>10</td>
<td>C1orf115</td>
<td>FAM198A</td>
<td>HAVCR2</td>
<td>KCNJ12</td>
<td>NPTX1</td>
<td>RTN4RL1</td>
<td>THSD4</td>
</tr>
<tr>
<td>11</td>
<td>C9orf125</td>
<td>FAM5B</td>
<td>HERC5</td>
<td>KCNB4</td>
<td>NPY1R</td>
<td>SDR16C5</td>
<td>TIP2</td>
</tr>
<tr>
<td>12</td>
<td>CBLN1</td>
<td>FAM64A</td>
<td>HLA-DOA</td>
<td>KDM6A</td>
<td>NR2F1</td>
<td>SECTM1</td>
<td>TMEM200/ZNF677</td>
</tr>
<tr>
<td>13</td>
<td>CCBE1</td>
<td>FAM78A</td>
<td>HLA-DPA1</td>
<td>KHDC1L</td>
<td>NT5E</td>
<td>SELENBP1</td>
<td>TMSB4Y</td>
</tr>
<tr>
<td>14</td>
<td>CD300A</td>
<td>FAM84A</td>
<td>HLA-DQA1</td>
<td>KIAA1644</td>
<td>OSR1</td>
<td>SERPINA3</td>
<td>TRIM68</td>
</tr>
<tr>
<td>15</td>
<td>CD300LB</td>
<td>FCER2</td>
<td>HNF4G</td>
<td>KLHL6</td>
<td>OTX1</td>
<td>SH3BGR1L</td>
<td>TTL2</td>
</tr>
<tr>
<td>16</td>
<td>CHODL</td>
<td>FMO2</td>
<td>HOXA13</td>
<td>LCP1</td>
<td>PABPC4L</td>
<td>SIGLEC1</td>
<td>UBF2QL1</td>
</tr>
<tr>
<td>17</td>
<td>CHST1</td>
<td>FOXK2</td>
<td>HOXB9</td>
<td>LIN7A</td>
<td>PAQR8</td>
<td>SIGLEC5</td>
<td>UCL1</td>
</tr>
<tr>
<td>18</td>
<td>CLCN4</td>
<td>FPR1</td>
<td>HOXC11</td>
<td>LOC391322</td>
<td>PAK9</td>
<td>SIX2</td>
<td>VAV1</td>
</tr>
<tr>
<td>19</td>
<td>CLEC4D</td>
<td>GABRE</td>
<td>HOXC8</td>
<td>LRAT</td>
<td>PIK3CG</td>
<td>SLC15A5</td>
<td>VAX1</td>
</tr>
<tr>
<td>20</td>
<td>CPE</td>
<td>GIMP1</td>
<td>HSP88</td>
<td>MAL2</td>
<td>PIP5K1B</td>
<td>SLC5A1</td>
<td>VPS37D</td>
</tr>
</tbody>
</table>

**Table 6:** List of genes that exhibit a correlation coefficient of > 0.5 between gene expression and 3’UTR methylation, but a correlation coefficient of < 0.5 between gene expression and gene body methylation, in at least two tissue types.
These genes were then subjected to a pathway analysis, which revealed an enrichment of genes involved in regulating various aspects of T cell activation (Table 7).

<table>
<thead>
<tr>
<th>Functional Annotation</th>
<th>$\Delta$ Hyper-geometric p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>T cell activation</td>
<td>7.73630258758785E-8</td>
</tr>
<tr>
<td>MHC class II receptor activity</td>
<td>1.76875556622534E-7</td>
</tr>
<tr>
<td>mesoderm development</td>
<td>4.468296831697853E-7</td>
</tr>
<tr>
<td>MHC class II protein complex</td>
<td>9.570774587081688E-7</td>
</tr>
<tr>
<td>Generation of second messenger molecules</td>
<td>1.155660531728697E-6</td>
</tr>
<tr>
<td>antigen processing and presentation of peptide or polysaccharide an...</td>
<td>2.015084530119795E-6</td>
</tr>
<tr>
<td>Type I diabetes mellitus</td>
<td>2.17304576997906E-6</td>
</tr>
<tr>
<td>Type I diabetes mellitus</td>
<td>2.17304576997906E-6</td>
</tr>
<tr>
<td>Staphylococcus aureus infection</td>
<td>5.228211575080663E-6</td>
</tr>
<tr>
<td>Staphylococcus aureus infection</td>
<td>5.228211575080663E-6</td>
</tr>
<tr>
<td>urocod</td>
<td>5.31237447564414E-6</td>
</tr>
<tr>
<td>trailing edge</td>
<td>6.31237447564414E-6</td>
</tr>
<tr>
<td>Translocation of ZAP-70 to Immunological synapse</td>
<td>5.427965512623229E-6</td>
</tr>
<tr>
<td>Phosphorylation of CD3 and TCR zeta chains</td>
<td>3.850201250072859E-6</td>
</tr>
<tr>
<td>Potassium Channels</td>
<td>2.019338082061386E-6</td>
</tr>
<tr>
<td>ear development</td>
<td>1.043368712871917E-5</td>
</tr>
<tr>
<td>Asthma</td>
<td>1.188640460694963E-5</td>
</tr>
<tr>
<td>Asthma</td>
<td>1.188640460694963E-5</td>
</tr>
<tr>
<td>innervation</td>
<td>1.343141269034414E-5</td>
</tr>
<tr>
<td>neuromuscular process controlling posture</td>
<td>1.343141269034414E-5</td>
</tr>
<tr>
<td>PD-1 signaling</td>
<td>1.3663025719317E-5</td>
</tr>
<tr>
<td>peripheral nervous system neuron differentiation</td>
<td>1.840186402368715E-5</td>
</tr>
<tr>
<td>peripheral nervous system neuron development</td>
<td>1.840186402368715E-5</td>
</tr>
<tr>
<td>TCR signaling</td>
<td>1.983781104951186E-5</td>
</tr>
<tr>
<td>transcription regulatory region sequence-specific DNA binding</td>
<td>2.038421410677303E-5</td>
</tr>
<tr>
<td>interferon-gamma-mediated signaling pathway</td>
<td>2.290437855134719E-5</td>
</tr>
<tr>
<td>renal vesicle development</td>
<td>2.44776863334983E-5</td>
</tr>
<tr>
<td>Allograft rejection</td>
<td>2.870014919032165E-5</td>
</tr>
<tr>
<td>Allograft rejection</td>
<td>2.870014919032165E-5</td>
</tr>
<tr>
<td>Costimulation by the CD28 family</td>
<td>3.013461928604233E-5</td>
</tr>
<tr>
<td>lymphocyte costimulation</td>
<td>3.013461928604233E-5</td>
</tr>
<tr>
<td>T cell costimulation</td>
<td>3.013461928604233E-5</td>
</tr>
</tbody>
</table>

**Table 7:** A NetWalker analysis of the 156 genes shows the correlation between gene expression and 3'UTR methylation of > 0.5 and gene body methylation of < 0.5. T cell activation and antigen presentation were the most overrepresented pathways.
This includes genes involved in T cell receptor signaling (*ITK*, *VAV1*) and T cell exhaustion (*HAVCR2*), as well as antigen presentation, particularly regarding the MHC class II complex (*HLA-DQA1*, *HLA-DOA*). These genes were then subject to a network analysis using NetWalker, which uncovered a single interconnected node of 23 genes, of which 7 genes are involved in T cell activation (Fig60.).

![Network Diagram](image)

**Figure 60:** A single network node with more than 1 interaction was identified using a Netwalker analysis. The genes in the red box represent genes that are canonically expressed in T cells and are known to be integral to T cell activation.
The protein products of three genes (*ITK*, *HAVCR2*, *VAV1*) interact with each other, and are primarily expressed in T cells. Furthermore, these genes contain a positive correlation between gene expression and 3’UTR methylation of > 0.5 and gene body methylation of < 0.5 in 2 or more tissue types, as is observed with *VAV1* (Fig 61).

**Figure 61:** A representative plot of the relationship between methylation and gene expression on the *VAV1* gene body. The top box represents the gene layout, with the blue portion highlighting the 3’UTR. In the “correlations” box, each line represents a tissue type.
Additionally, the immune checkpoint gene \textit{HAVCR2} also exhibits this phenomenon, and interacts with both \textit{VAV1} and \textit{ITK} (Fig62.).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{Figure62.png}
\caption{A representative plot of the relationship between methylation and gene expression on the \textit{HAVCR2} gene body. The top box represents the gene layout, with the blue portion highlighting the 3'UTR. In the "correlations" box, each line represents a tissue type.}
\end{figure}
Because the samples analyzed are tumor tissues that contain immune cells, we wondered whether methylation and expression differences in T cell-related genes were occurring within tumor cells or within T cells present in the tumor samples.

To determine whether the 3'UTR methylation phenomenon should be investigated within tumor cells, or within T cells, the expression of the three T cell-specific genes listed above (ITK, VAV1, and HAVCR2) were plotted against estimated levels of T cells using previously established methodology for ascertaining T cell count in TCGA samples (244). For each of these three genes, a highly significant association was discovered between expression and T cell count was observed for ITK (Fig63.).
Figure 63: The expression of *ITK* (x-axis) was compared to the estimated number of T cells (y-axis) in each TCGA sample, revealing a robust correlation.
VAV1 expression also showed a very strong correlation with T cell count (Fig 64.).

**Figure 64:** The expression of VAV1 (x-axis) was compared to the estimated number of T cells (y-axis) in each TCGA sample, revealing a robust correlation.
Additionally, *HAVCR2* expression was robustly associated with T cell count (**Fig65**).

**Figure 65:** The expression of *HAVCR2* (x-axis) was compared to the estimated number of T cells (y-axis) in each TCGA sample, revealing a robust correlation.
These data support the conclusion that for the analyzed genes, 3'UTR methylation being connected to gene expression is occurring within T cells. Based upon these findings, the relationship between 3'UTR methylation and gene expression of these genes was examined in T cells.

Subsection 4.5: Increased DNA methylation of Havcr2 specific to the 3'UTR occurs in conjunction with upregulated gene expression after T cell activation.

T cells are known to dynamically modulate DNA methylation when changing activation states (247), and a specific epigenetic profile is essential for proper function. Therefore, changes to 3'UTR methylation and gene expression of Itk, Vav1, and Havcr2 were examined after isolating T cells from c57Bl/6 mice and activating the T cells ex vivo. 72 hours after stimulation, expression of Vav1 was unchanged,
and *Itk* expression decreased slightly; however, *Havcr2* expression increased by nearly 50-fold (Fig66.).

**Figure 66:** T cells isolated from C57Bl/6 mice and activated *ex vivo* had no significant differences in *Itk* or *Vav1* gene expression, whereas the gene expression of *Havcr2* was substantially increased. Likewise, 3'UTR methylation of *Itk* and *Vav1* exhibited minimal differences after T cell activation, whereas *Havcr2* 3'UTR methylation increased by 2.5-fold.
Next, two CpG sites within the 3’UTR for each gene were examined for changes to DNA methylation. Similar to gene expression, Vav1 3’UTR methylation did not differ between naïve and activated T cells, whereas Itk 3’UTR methylation decreased slightly; however, Havcr2 3’UTR methylation increased substantially, approximately 2.5-fold higher at both sites in activated T cells relative to naïve (Fig 67.).

**Figure 67:** Methylation of the Havcr2 promoter region in naïve T cells was low, and activated T cells displayed a significant decrease; however, this promoter lacks a CpG island and is not likely to regulate Havcr2 gene expression. Methylation of the adjacent exon increases slightly, and methylation of the intron immediately adjacent to the 3’UTR shows no significant changes.
To determine whether this increase in *Havcr2* expression was due to changes in promoter methylation, as is the case for many genes switched on after T cell activation, methylation of the promoter region was assayed. While a significant decrease was observed in methylation of the promoter region after activation, very low promoter methylation was observed in the naïve state, and the *Havcr2* promoter lacks a CpG island, indicating that this modification is likely not the major regulator of gene expression. Next, methylation of the exon and intron immediately adjacent to the 3'UTR of *Havcr2* was examined. Methylation of the intron did not change upon activation, and methylation of the adjacent exon showed a slight increase in methylation, however, not nearly as substantial as that of the 3'UTR (Fig68.). These data demonstrate the exquisite specificity with which robust increases in methylation are targeted to the 3'UTR.

**Figure 68:** Schematic of the *Havcr2* gene, with the location of the interrogated sites in each discrete region of the gene highlighted.
Subsection 4.6: Treatment with decitabine, or knockout of Dnmt3a results in reduced *Havcr2* gene expression.

To investigate whether increases in DNA methylation were necessary for the increase in *Havcr2* gene expression, T cells were activated *ex vivo*, then treated with the DNMT inhibitor decitabine, or DMSO vehicle as a control. After 72 hours, the activated T cells treated with decitabine showed a 4-fold decrease in *Havcr2* gene expression relative to the DMSO-treated T cells, indicating that DNA methylation is upstream of upregulated *Havcr2* gene expression after T cell activation (**Fig69.**).

**Figure 69:** *Havcr2* gene expression after *ex vivo* activation of T cells and subsequent treatment with the demethylating agent decitabine. Treatment with decitabine causes a four-fold decrease in *Havcr2* expression.
These data highlight *de novo* DNA methylation as an important component of modulating gene expression after activation, as observed in previous studies (1). As further evidence of this observation, the gene expression profile of exhausted T cells from a *Dnmt3a* knockout mouse were examined using gene array and genome-wide bisulfite sequencing data obtained from Ghoneim *et al*(1). From this gene array, *Havcr2* was fifth-most downregulated gene in exhausted CD8 T cells lacking *Dnmt3a* relative to wild-type (*Fig70.*), supporting the conclusion that *de novo* methylation of *Havcr2* 3'UTR results in increased gene expression.
Figure 70: Heat map of the top genes whose expression was most altered in naïve or exhausted T cells after Dnmt3a knockout. Havcr2 expression was the 5th-most decreased in Dnmt3a knockout exhausted T cells(1).
The difference in *Havcr2* expression after *Dnmt3a* knockout in naïve and exhausted T cells was then quantified, revealing a significant decrease in *Havcr2* expression after *Dnmt3a* knockout in exhausted, but not naïve, T cells (Fig71.).

**Figure 71:** Quantitative expression data for *Havcr2* expression in naïve or exhausted T cells after *Dnmt3a* knockout.
Next, using the bisulfite array data generated from this study, the gene body methylation of *Havcr2* was compared between wild-type and *Dnmt3a* knockout CD8+ T cells. This revealed that a substantial decrease in *Havcr2* methylation occurred within the 3'UTR, whereas most sites within the rest of the gene body remained relatively unchanged (Fig72.), independently confirming the site specificity of *Havcr2* methylation.

**Figure 72:** Visualization of bisulfite sequencing array data that shows extent of methylation of *Havcr2* gene. The top portion is the entire gene, and the bottom portion is the 3'UTR. The 3'UTR shows the most substantial change in methylation after *Dnmt3a* knockout in exhausted T cells.
Section 5: Discussion

Subsection 5.1: The implications of CARD11 gene body methylation, and role in KIRC and LUAD pathogenesis

The results of this analysis have revealed two important findings: 1) aberrant expression of CARD11 in renal cell carcinoma and lung adenocarcinoma can promote cancer pathogenesis through changes to mTOR signaling and autophagy, and 2) CARD11 gene body methylation results in increased expression.

This study expands our understanding of the creative means by which tumor cells activate genes otherwise unexpressed in their tissue of origin, and to co-opt downstream signaling pathways that serve to propel tumor malignancy forward. In the case of CARD11, this lymphocyte-specific gene canonically links the T- and B-cell receptors to the NF-kB pathway, thereby mediating lymphocyte activation upon antigen stimulation. Based upon this knowledge, CARD11 overexpression in the context of renal cell carcinoma and lung adenocarcinoma was hypothesized to increase NF-kB activation even in the absence of antigen receptors. Surprisingly, knockdown of CARD11 did not affect the NF-kB pathway, instead conveying its effects primarily through the mTOR axis. CARD11 has been shown previously to be involved in proper mTOR activation in T cells, but this is the first time this phenomenon has been observed in cells of epithelial origin.

While localized renal cell carcinoma can be eliminated through procedures such as resection of the kidney or heat/cold ablation therapy, approximately one third of patients with clear cell renal cell carcinoma develop metastasis(248). Stage IV
renal cell carcinoma has a survival rate of only 10%, and is highly resistant to chemotherapy(249), necessitating more efficacious therapeutic approaches. The mTOR pathway, which is an important component of damage response in the kidney, is a key mediator of cancer pathogenesis in renal cell carcinoma(250). It is hypothesized that this is due to the reliance of renal cancer on HIF-1α and HIF-2α; these two proteins depend on mTOR activation for efficient translation(251). Based on this observation, small molecule inhibitors of mTOR such as Temsirolimus and Everolimus have been employed to treat advanced stage metastatic renal cell carcinoma, with patients showing longer overall survival(252) and progression-free survival(253) from these drugs, respectively. However, resistance to these therapies emerges over time, primarily through compensatory activation of the MAPK pathway. Current clinical trials combining mTOR inhibitors with MAPK inhibitors have been limited by toxicity(254); therefore, new, less toxic approaches in targeting the mTOR pathway are needed. Because CARD11 is not expressed by normal kidney cells, this offers an opportunity to suppress mTOR activation while not impacting the kidney. Moreover, targeting CARD11 gene body methylation may provide a highly selective means by which CARD11 expression, and downstream mTOR signaling, may be suppressed. However, the prevalence of CARD11 gene body methylation in T and B cells, and how demethylation impacts the function of these cells, requires to be investigated before targeting this region with site-specific demethylation can achieve clinical adoption.

In addition to further supporting the finding that tumor cells can utilize non-canonical pathways, this analysis also reveals how tumor cells employ non-canonical
mechanisms to induce aberrant expression of genes that are otherwise silenced by their tissue of origin. Previous studies have uncovered various means by which this process is undertaken. These include promoter demethylation, upregulation of transcription factors, changes to chromatin architecture, and downregulation of microRNAs, among many others. Here, we have provided further evidence that gene body DNA methylation is an alternative mechanism by which tumor cells can alter their transcriptional profile during tumorigenesis and tumor progression. This serves to reiterate the importance of examining gene body DNA methylation as a means of fully understanding transcriptional regulation, and that targeting this type of epigenetic modification is a newly emerging route to modulate aberrant gene expression.

The past 20 years has seen gene regulation by RNAs becoming more prominent and expansive. The most famous, miRNAs, have been shown to entire axes of cellular function. Lnc-RNAs can induce changes to chromatin, recruit transcription factors, or bind mRNAs. Importantly, they are also known to modulate DNA methylation by associating with DNMT1, DNMT3A & B, or GADD45A/TDG, among others. Because antisense RNAs contain part or all of their sequence that is reverse complement to the sense strand, this allows for remarkable specificity in localizing proteins, such as methylation enzymes, to select loci. Furthermore, over 30% of the genome produces some form of antisense transcript. Despite these unique properties, the vast majority of the antisense RNAome has never been studied. A certain portion of these antisense RNAs may not serve a function at all, and can be discarded as transcriptional noise, but this study highlights the need to catalog the thousands of antisense RNAs that exist, and how they are involved in
normal and abnormal cellular processes. As systemic siRNA therapies become more accepted in the clinic, antisense RNAs make promising candidates for a new frontier of targets in cancer. However, in this study, we have determined that the CARD11-NAT does not regulate CARD11. Despite this, we have concluded that these two transcripts are co-regulated by the gene body CpG island. Considering CARD11-NAT is spliced, and is highly overexpressed in multiple cancer types, it could be functionally relevant both for normal cellular processes as well as in cancer. On the other hand, its increased expression in cancer may simply be a byproduct of increased CARD11 expression. The role this NAT plays in cellular biology must be further explored.

What prompts the increase in CARD11 DNA methylation during tumorigenesis and tumor development still remains to be understood. As mentioned, tumor cells exhibit global hypomethylation, but hypermethylation in regions such as the promoters of tumor suppressor genes. What specific factors guide DNA methylation enzymes to establish this site-specific hypermethylation, and whether there is a mechanistic overlap between gene body hypermethylation and promoter methylation, necessitate investigation.

In conclusion, these data have revealed the role that the lymphocyte-specific gene CARD11 is upregulated in certain epithelial cancers, and overexpression is strongly associated with decreased patient overall survival in kidney renal cell carcinoma. CARD11 can activate the mTOR pathway in the context of epithelial cancer, while not impacting the canonical NF-kB pathway. The expression of CARD11 is regulated by methylation of the CpG island within the gene body, which
increases expression. Therefore, this study emphasizes the importance of considering how methylation can increase expression of oncogenes, and provides novel insight into how oncogenes are activated during tumorigenesis and tumor progression.

Subsection 5.2: The relationship between 3'UTR methylation and expression, and the implications of 3'UTR methylation in T cell activation genes and beyond

Ascertaining the functional and clinical effects of site-specific DNA methylation remains an important step in unraveling the many layers of epigenetic regulation. Here, we found that the 3’UTR is a functionally distinct site for epigenetic modification. DNA methylation of the gene body is known to be associated with increased gene expression, but by separately examining the 3’UTR across both normal and tumor tissue samples, we revealed an enrichment of DNA methylation sites in this region that are uniquely correlated with increased gene expression. Moreover, we identified several genes that exhibited divergent gene expression between normal and tumor tissues; they lacked significant alterations in copy number or promoter methylation that would explain the differences in expression independently of changes in 3’UTR methylation. In 5 of the 10 tumor types examined, 3’UTR methylation was associated with patient overall survival in a significant number of genes.

Interestingly, by separating out the 3’UTR as a distinct functional region for the first time, an unexpected link between DNA methylation of this region and T cell regulation was observed. For certain genes in this category, particularly those related to T cell receptor activation (Itk and Vav1), the extent of 3’UTR DNA methylation was
correlated with both the presence of T cells in a tumor and with patients’ overall survival. For the progressively important immune checkpoint gene *Havcr2*, DNA methylation of the 3’UTR may serve as a means by which T cell exhaustion occurs. TIM-3-expressing T cells exhibit a severely exhausted phenotype (255, 256), and this protein is frequently found to be expressed in tumor-infiltrating lymphocytes (257). In addition, the expression of TIM-3 promotes resistance to PD-1/PD-L1 blockade (258), and de-methylating agents, in combination with PD-1/PD-L1 blockade, exhibit greater sensitivity (1). DNA methylation inhibition and the subsequent sustained or increased expression of *Havcr2* may be an underlying reason for this observation.

These data suggest two new avenues of exploration that will broaden our understanding of this epigenetic modification. The first is ascertaining how 3’UTR methylation influences gene expression. It is well established that DNA methylation affects the binding of regulatory proteins (41). In the case of proteins with methylation-binding domains, DNA methylation can increase binding (259, 260). On the other hand, DNA methylation can inhibit protein binding or mask sequence recognition, as is the case for many transcription factors (30). Another potential explanation for the effect of the 3’UTR on gene expression may be differential alternative splicing and alternative polyadenylation. Gene body methylation has already been shown to affect exon inclusion (261); alternative polyadenylation has not been linked with DNA methylation, but if different lengths of the 3’UTR are dependent on methylation, transcripts with shorter 3’UTRs would have greater mRNA stability and thereby more gene expression (262). However, gene body methylation has also been observed as
a consequence of higher gene expression, rather than as a cause (263); therefore, 3'UTR methylation may occur downstream of higher gene expression.

The second question arises in regard to how 3'UTR methylation is regulated. *De novo* DNA methylation is deposited by the DNMT3 enzymes (264), and *de novo* demethylation is handled by the TET family of enzymes; therefore, these are likely to play a role in producing the differential methylation observed in tumor and normal tissues and in activated versus naïve T cells. Indeed, given the substantial decrease in *Havcr2* gene expression after *Dnmt3a* knockout, this enzyme in particular seems to be involved. However, the co-factors that position these enzymes in a tightly controlled spatial and temporal context are currently unknown.

These data have multiple implications. First, these genes may play a previously unidentified role in cancer pathogenesis. Second, 3'UTR methylation of these genes may serve as a biomarker for disease presence and progression. Finally, demethylating this region may serve as a target for cancer therapy; how demethylating agents affect the 3'UTR should be taken into account when evaluating the mechanism and efficacy of these therapies.

Taken together, our findings indicate that the 3'UTR is a region of epigenetic importance. These data raise the possibility of a novel component of epigenetic regulation that operates during T cell development and activation, as well as other cellular processes. Furthermore, they shed light on a potential novel mechanism by which T cells upregulate immune checkpoint mediators. In addition, the disrupted patterns of 3’UTR methylation observed in cancer suggest that alterations in 3’UTR methylation play a role in tumorigenesis and tumor progression. Coupled with the
Cas9-Tet1 technology developed here and elsewhere, the capacity to specifically modulate the epigenome has finally come into reach. Given this advancement, decoding the functional implications of site-specific modifications to the DNA and histones has taken on even greater importance. The findings reported here broaden our understanding of the effect of DNA methylation on cellular processes, and most importantly, they highlight novel components of cancer pathogenesis, opening new avenues for clinical therapy. Future research should focus on which modification sites are most critical for cancer pathogenesis, as well as optimizing both the efficiency and delivery of constructs that can target these sites. Achieving these aims hold the potential to grant exquisite control over the epigenome.
Bibliography


41. Tate, P. H., and A. P. Bird. 1993. Effects of DNA methylation on DNA-binding proteins and
42. Li, B., M. Carey, and J. L. Workman. The Role of Chromatin during Transcription. Cell 128:
707-719.
Structural & Molecular Biology 20: 274.
44. Shlyueva, D., G. Stampfel, and A. Stark. 2014. Transcriptional enhancers: from properties to
45. Neph, S., A. B. Stergachis, A. Reynolds, R. Sandstrom, E. Borenstein, and J. A.
Stamatoyannopoulos. 2012. Circuitry and dynamics of human transcription factor
factors shape the mouse methylome at distal regulatory regions. Nature 480: 490-495.
Lineage-specific DNA methylation in T cells correlates with histone methylation and
Reversible Regulation of Promoter and Enhancer Histone Landscape by DNA Methylation in
49. Stroud, H., S. Feng, S. Morey Kinney, S. Pradhan, and S. E. Jacobsen. 2011. 5-
Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic
stem cells. Genome biology 12: R54.


Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J.
Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C.
Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M.
Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A.
P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V.
Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M.
Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R.
McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala,
L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge,
L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler,
Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P.
Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. Mclysaght, T. Mikkelsen, J. V.
Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A.
Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K.
Osoegawa, H. Shizuya, S. Choi, Y. J. Chen, and J. Szustakowski. 2001. Initial sequencing and


87. Sutherland, G. R. 2003. Rare fragile sites. *Cytogenetic and genome research* 100: 77-84.


of the putative oncogene ELMO3 are associated with lung cancer development and metastases formation. *Oncoscience* 1: 367-374.


hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines.  
Cancer Res 57: 808-811.


*Nature Reviews Immunology.*


212. Du, W., M. Yang, A. Turner, C. Xu, R. L. Ferris, J. Huang, L. P. Kane, and B. Lu. 2017. TIM-3 as 
a Target for Cancer Immunotherapy and Mechanisms of Action. *International journal of 
molecular sciences* 18.


215. Zare-shahabadi, A., E. Masliah, G. V. W. Johnson, and N. Rezaei. 2015. Autophagy in 

216. Bento, C. F., M. Renna, G. Ghislat, C. Puri, A. Ashkenazi, M. Vicinanza, F. M. Menzies, and D. 
Biochemistry* 85: 685-713.

217. He, C., and D. J. Klionsky. 2009. Regulation Mechanisms and Signaling Pathways of 

218. Saxton, R. A., and D. M. Sabatini. mTOR Signaling in Growth, Metabolism, and Disease. *Cell* 
168: 960-976.
Lyons, J. Ma, Y. Ma, Y. Ma, Z. Ma, J. Machado, G. M. Machado-Santelli, F. Macian, G. C.
Malorni, A. Maloyan, F. Mami-Chouaib, N. Man, J. D. Mancias, E. M. Mandelkow, M. A.
Marambaud, A. M. Marconi, Z. Marelja, G. Marfe, M. Margeta, E. Margittai, M. Mari, F. V.
Mariani, C. Marin, S. Marinelli, G. Marino, I. Markovic, R. Marquez, A. M. Martelli, S.
Martinez-Velazquez, M. Martinez-Vicente, W. K. Martins, H. Mashima, J. A. Mastrianni, G.
Matarese, P. Matarrese, R. Mateo, S. Matoba, N. Matsumoto, T. Matsushita, A. Matsuura,
T. Matsuzawa, M. P. Mattson, S. Matus, N. Maugeri, C. Mauvezin, A. Mayer, D. Maysinger,
McKenna, J. J. McMahon, I. A. McNeish, F. Mechta-Grigoriou, J. P. Medema, D. L. Medina,
K. Megyeri, M. Mehrpour, J. L. Mehta, Y. Mei, U. C. Meier, A. J. Meijer, A. Melendez, G.
Melino, S. Melino, E. J. de Melo, M. A. Mena, M. D. Meneghini, J. A. Menendez, R. Menezes,
L. Meng, L. H. Meng, S. Meng, R. Menghini, A. S. Menko, R. F. Menna-Barreto, M. B. Menon,
Rosa, A. Rosello, K. V. Rosen, P. Rosenstiel, M. Rost-Roszkowska, K. A. Roth, G. Roue, M.
Rouis, K. M. Rouschop, D. T. Ruan, D. Ruano, D. C. Rubinsztein, E. B. Rucker, 3rd, A. Rudich,
Salazar-Roa, P. Salomoni, A. K. Saluja, P. M. Salvaterra, R. Salvioli, A. Samali, A. M. Sanchez,
J. A. Sanchez-Alcazar, R. Sanchez-Prieto, M. Sandri, M. A. Sanjuan, S. Santaguida, L.
Santambrogio, G. Santoni, C. N. Dos Santos, S. Saran, M. Sardiello, G. Sargent, P. Sarkar, S.
E. Schneider, E. M. Schneider, J. L. Schneider, E. A. Schon, M. J. Schonenberger, A. H.
Schonthal, D. F. Schorderet, B. Schroder, S. Schuck, R. J. Schulze, M. Schwarten, T. L.
Schwarz, S. Sciarretta, K. Scotto, A. I. Scovazzi, R. A. Screaton, M. Screen, H. Seca, S. Sedej,
Setoguchi, C. Settembre, J. J. Shacka, A. N. Shajahan-Haq, I. M. Shapiro, S. Sharma, H. She,
Shepherd, J. Shi, Q. Shi, Q. Shi, Y. Shi, S. Shibutani, K. Shibuya, Y. Shidoji, J. J. Shieh, C. M.
Shirabe, R. Shiri-Sverдов, O. Shirihai, G. C. Shore, C. W. Shu, D. Shukla, A. A. Sibirny, V. Sica,
C. J. Sigurdson, E. M. Sigurdsson, P. S. Sijwali, B. Sikorska, W. A. Silveira, S. Silvente-Poirot,
Sirohi, B. J. Sishi, A. Sittler, P. M. Siu, E. Sivridis, A. Skwarska, R. Slack, I. Slaninova, N. Slavov,
S. S. Smaili, K. S. Smalley, D. R. Smith, S. J. Soenen, S. A. Soleimanpour, A. Solhaug, K.
Somasundaram, J. H. Son, A. Sonawane, C. Song, F. Song, H. K. Song, J. X. Song, W. Song, K.
Y. Soo, A. K. Sood, T. W. Soong, V. Soontornniyomkij, M. Sorin, F. Sot gia, D. R. Soto-
Pantoja, A. Sothibundhu, M. J. Sousa, H. P. Spain, P. N. Span, A. Spang, J. D. Sparks, P. G.
Speck, S. A. Spector, C. D. Spies, W. Springer, D. S. Clair, A. Stacchiotti, B. Staels, M. T. Stang,
Stoka, Z. Storchova, B. Stork, V. Stratoulis, D. J. Stravopodis, P. Strnad, A. M. Strohecker, A.
L. Strom, P. Stromhaug, J. Stulik, Y. X. Su, Z. Su, C. S. Subauste, S. Subramaniam, C. M. Sue,
S. W. Suh, X. Sui, S. Sukser ee, D. Sulzer, F. L. Sun, J. Sun, J. Sun, S. Y. Sun, Y. Sun, Y. Sun, Y.
Vellai, Y. Takahashi, S. Takats, G. Takemura, N. Takigawa, N. J. Talbot, E. Tamagno, J.
Tch etina, A. R. Tee, F. Teixeira-Clerc, S. Telang, T. Tencomnao, B. B. Teng, R. J. Teng, F.
Tripathi, D. Trisciuglio, R. Troncoso, I. P. Trougakos, A. C. Truttmann, K. J. Tsai, M. P.
Tumbarello, B. Turk, V. Turk, R. F. Turner, A. A. Tveita, S. C. Tyagi, M. Ubukata, Y. Uchiyama,
A. Udelnow, T. Ueno, M. Umekawa, R. Umemiyashi, B. R. Underwood, C. Ungermann,
Vachova, H. Vakifahmetoglu-Norberg, R. Valdor, E. M. Valente, F. Vallette, A. M. Valverde,
G. Van den Berghe, L. Van Den Bosch, G. R. van den Brink, F. G. van der Goot, I. J. van der
Klei, L. J. van der Laan, W. G. van Doorn, M. van Egmond, K. L. van Golen, L. Van Kaer, M.
van Lookeren Campagne, P. Vandenabeele, W. Vandenberghe, I. Vanhorebeek, I. Varela-Nieto,
M. H. Vasconcelos, R. Vasko, D. G. Vavvas, I. Vega-Naredo, G. Velasco, A. D.
Velentzas, P. D. Velentzas, T. Vellai, E. Vellengha, M. H. Vendelbo, K. Venkatachalam, N.
Ventura, S. Ventura, P. S. Veras, M. Verdier, B. G. Vertessy, A. Viale, M. Vidal, H. L. Vieira, R.
Schwarzenberg, D. E. Voth, V. Vouret-Craviari, K. Vuori, J. M. Vyaz, C. Waeber, C. L. Walker,
Wang, T. F. Wang, X. Wang, X. J. Wang, X. W. Wang, X. Wang, X. Wang, Y. Wang, Y. Wang, Y.
Wang, Y. J. Wang, Y. Wang, Y. Wang, Y. T. Wang, Y. Wang, Z. N. Wang, P. Wappner, C.
Weekes, J. Wei, T. Weide, C. C. Weihi, G. Weindl, S. N. Weis, L. Wen, X. Wen, Y. Wen, B.
Westermann, C. M. Weyand, A. R. White, E. White, J. L. Whitton, A. J. Whitworth, J. Wiels,


Vita

Michael McGuire was born and raised in Lemoore, California. He attended college at the University of California, Los Angeles, where he majored in Microbiology, Immunology, and Molecular Genetics. After college he worked in a prostate cancer lab at Geffen School of Medicine, before being accepted into the PhD program at UT MD Anderson GSBS. There, he worked in the lab of Dr. Anil Sood, where he focused on the functional and pathological consequences of changes to site-specific DNA methylation in cancer, among other topics. After graduating, Michael will begin work at ClearView Healthcare Partners, a life sciences consulting firm located in Boston.