의학석사 학위논문

다양한 암종에서 메틸화 변이 정도와 종양 면역원성 사이 상관관계에 관한 연구

Methylation analysis of thirty-three cancer types reveals an inverse correlation between methylation burden and tumor immunogenicity

2020 년 2 월

서울대학교 대학원

의학과 중개의학

박 창 희

다양한 암종에서 메틸화 변이 정도와 종양 면역원성 사이 상관관계에 관한 연구

지도 교수 김 동 완

이 논문을 의학석사 학위논문으로 제출함

2019 년 10 월

서울대학교 대학원

의학과 중개의학

박 창 희

박창희의 의학석사 학위논문을 인준함

2020 년 1 월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

Abstract

Methylation analysis of thirty-three cancer types reveals an inverse correlation between methylation burden and tumor immunogenicity

Changhee Park

Translational Medicine, School of Medicine

The Graduate School

Seoul National University

While genomic alterations in tumors, such as somatic mutations or somatic chromosomal instability, have been reported as reliable biomarkers in immune-oncology, associations between the methylation landscape and tumor immunogenicity is unknown. I sought to find biomarker related to methylation to represent tumor immunogenicity.

I used The Cancer Genome Atlas (TCGA) pan-cancer database (N~8,000) to define methylation burden (MetB) as the number of hypermethylated or hypomethylated

CpG sites to represent the degree of aberrant methylation. I comprehensively investigated the association of MetB with cytolytic activity score, calculated by mean of *GZMA* and *PRF1* expression levels, and other various genomic profiles.

The degree of methylation aberrancy correlated with methylation subtypes defined in previous literatures. It also showed negative correlation expression of molecules coordinating immune recognition of tumors. The pan-cancer analysis showed that MetB was negatively correlated with cytolytic activity score ($\rho = -0.37$, $p < 0.001$), independent of mutation burden and chromosomal instability. The negative correlation was consistent in the external cohort of lung adenocarcinoma and low grade glioma (Spearman $\rho = -0.41$, $p < 0.001$ and $\rho = -0.34$, $p = 0.014$, respectively). MetB also had negative correlation with interferon-gamma signatures and was lower in highly immunogenic subtypes of immune landscape. Furthermore, patients with bottom 20% MetB showed longer progression free survival to ipilimumab in TCGA melanoma patients ($p = 0.029$).

These findings emphasize the importance of methylation aberrancy for tumors to evade immune surveillance and warrant further development of methylation biomarker.

Keyword: The Cancer Genome Atlas, Methylation, Immunotherapy

Student Number: 2018-21621

Contents

List of Figures

List of Tables

Introduction

Immunotherapy with immune checkpoint inhibitors [antibodies for programmed cell death-1 (PD-1) and programmed cell death ligand-1 (PD-L1)] introduced a new era of cancer therapeutics, bringing tumor immunity into play[1]. This led to numerous researches on mechanisms underlying tumor immunogenicity and biomarkers on these mechanisms[2]. For examples, investigations regarding genomic correlations with tumor immunogenicity have been largely explored[2]. Tumor mutation burden, which results in a higher likelihood of generation of tumor neoantigen to stimulate a tumor immune response, is largely explored[3]. Anti-PD-1 antibody provided favorable responses in patients with microsatellite instability-high tumors, and tumor mutation burden enabled successful enrichment of responders in clinical trials[4,5]. Copy number variations have also been found to be associated with immunogenicity[6,7]. Previous studies discovered that gene expression profiles, such as cytolytic activity score (CytAct), interferon gamma gene signature and immune signature, represent immunogenicity and may discriminate responders from non-responders to immunotherapy[7-9]. However, the role of methylation as a biomarker for tumor immunogenicity are not well established compared to other genomic features[2,10].

DNA methylation modulates chromatin remodeling and RNA transcription, which eventually affects the characteristics and behavior of cancer at the cellular level[11]. Aberrant methylation, including promoter hypermethylation and hypomethylation, is involved in tumorigenesis in a variety of cancers[12,13], which is also demonstrated in systemic approach in a recent study[14]. Interestingly, hypermethylation in promoter regions has been described as having an inverse

correlation with gene expression of neoantigens and immune related genes[15,16]. Moreover, epigenetically enhanced events can occur in cancer testis antigens, resulting in highly expressed antigens that are possibly associated with a favorable response to anti-PD-1 antibody[14,17]. On the other hand, DNA methylation loss was found to correlate with the immune escape signatures[18].

These findings raise clinical and investigational questions: would aberrant methylation affect the tumor microenvironment associated with immunogenicity? In addition, would an index that represents overall methylation aberrancy be used as a biomarker for immunotherapy? In this study, I used The Cancer Genome Atlas (TCGA) data to define the methylation burden (MetB) as representative for the degree of overall methylation aberrancy and examined correlation between MetB and tumor immunogenicity.

Results

Defining the degree of methylation aberrancy and its correlation with methylation subtypes

Methylation data of TCGA consists of 396,065 CpG sites expressed as β-scores, which is the ratio of the methylated probe intensity and the overall intensity in the CpG site[19], shown as a numerical value between 0 and 1. To define the degree of methylation aberrancy with MetB, I aimed MetB to represent the number of hypermethylated and hypomethylated CpG sites. While definite cutoffs for hypermethylation and hypomethylation of CpG sites do not exist, distributions of β-scores in each site were largely variable (examples provided in **Figure 1**), which refrained me from setting uniform cutoffs throughout whole CpG sites. Therefore, I rather set cutoff of each CpG sites individually (**Figure 2**). I assumed that in each CpG site, samples with top 5% β-scores among all samples have hypermethylation at the corresponding CpG site and bottom 5% β-scores have hypomethylation at the corresponding CpG site. Then, MetB of a sample was finally defined as the log 2 value of the number of hypermethylated or hypomethylated CpG.

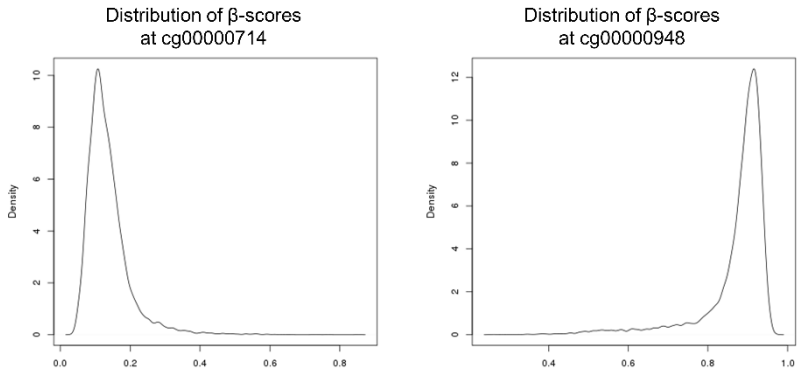

Figure 1. Examples of distribution of β-scores of 2 different CpG sites across pan-cancer samples. Note 'cg00000741' site (left) has right skewed distribution while 'ch00000948' site (right) has left skewed distribution.

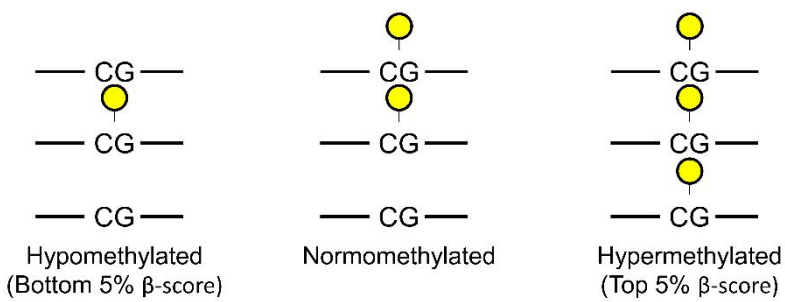

Figure 2. Definition of methylation burden. Schematic figure of definition of hypermethylation and hypomethylation. Majority of samples, 90% in this study, would have average methylation status, or normomethylated status, as depicted in the middle of the figure. Then, there would be 5% of samples with top 5% β-scores as depicted in the right of the figure which are defined as hypermethylated. The remaining 5% of samples with bottom 5% β-scores as depicted in the left of the figure are defined as hypomethylated.

To demonstrate whether my approach in defining hypermethylation and hypomethylation along with MetB is relevant, I evaluated whether my definition correlates with methylation subtypes described in previous literatures of low grade glioma (LGG)[20], lung adenocarcinoma (LUAD)[21], head and neck cancer (HNSC)[22], breast cancer (BRCA)[23] and colon adenocarcinoma (COAD)[24]. As shown in **Figure 3**, methylation subtypes previously known as hypermethylated subtypes ("IDH1/2 mutant" in LGG, "High methylation" in LUAD, "CpG island hyper-methylated" in HNSC, "Cluster 3" in BRCA and "CIMP-H" in COAD) had significantly higher hypermethylation burden. Similarly, previously known hypomethylated subtypes ("Low methylation" in LUAD, and "Hypo-methylated" in HNSC") had significantly higher hypomethylation burden. Notably, MetB was variable in each subtype and although significant differences in MetBs between methylation subtypes are seen, directions of methylation tendency were not always consistent with higher MetB. Specifically, in LGG, LUAD, BRCA and COAD, MetB was generally high in hypermethylated subtypes whereas in HNSC, MetB was the highest in hypomethylated subtype (**Figure 4**). These results suggest that my approach lines with methylation subtypes while providing MetB as a quantitative measure for the degree of methylation aberrancy throughout whole CpG sites which is not available from methylation subtypes.

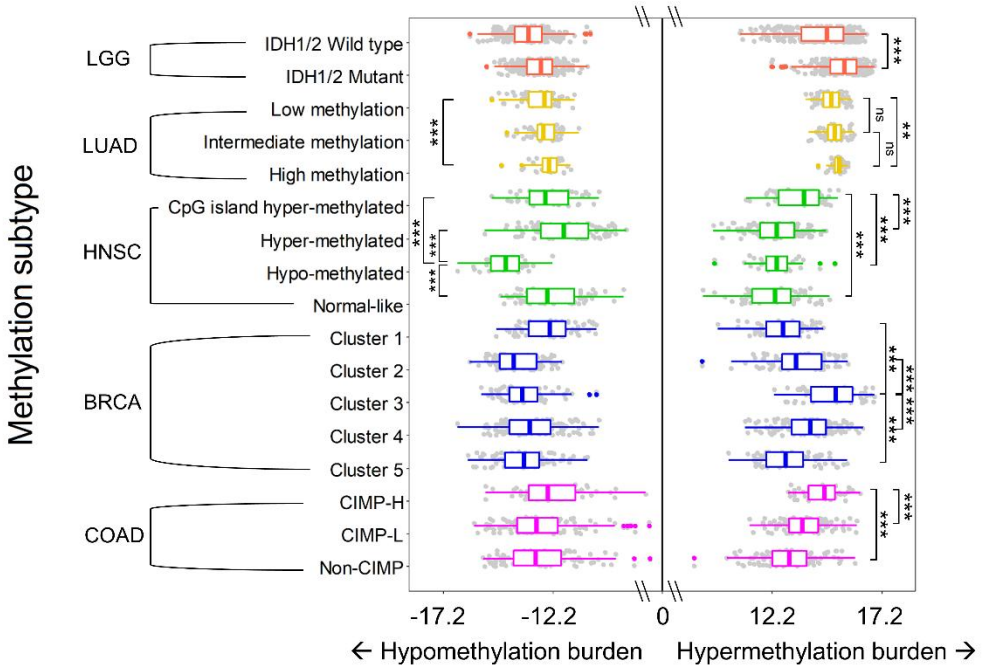

Figure 3. Definition of methylation in correlation with previous methylation subtypes. The boxplots show hypermethylation burden and hypomethylation burden in each methylation subtype of cancer. In the plot, the right-side show hypermethylation burden, higher to the right, and the left-side show hypomethylation burden, higher to the left. Vertical lines in the boxplots represent the upper 25%, median, and lower 25% MetB values. The colors of boxplots are different according to each cancer type; LGG in red; LUAD in yellow; HNSC in green; BRCA in blue; and COAD in magenta. Significant levels are depicted as asterisks; ns, not significant; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. BRCA, breast cancer; CIMP-H, CpG island methylator phenotype-high; CIMP-L, CpG island methylator phenotype-low; COAD, colon adenocarcinoma; HNSC, head and neck cancer; LGG, low grade glioma; LUAD, lung adenocarcinoma.

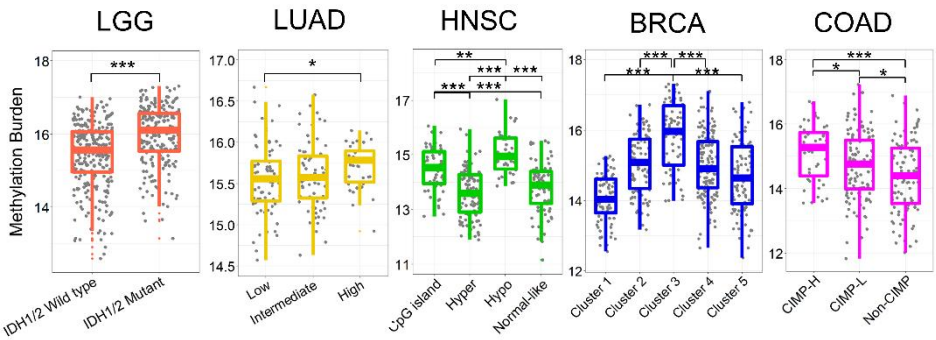

Figure 4. Comparisons of methylation burden between each methylation subtype. The five boxplots show differences of MetB between methylation subtypes of each cancer type. Horizontal lines in the boxplots represent the upper 25%, median, and lower 25% MetB values. The colors of boxplots are different according to each cancer type; LGG in red; LUAD in yellow; HNSC in green; BRCA in blue; and COAD in magenta. Significant levels are depicted as asterisks; ns, not significant; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. BRCA, breast cancer; CIMP-H, CpG island methylator phenotype-high; CIMP-L, CpG island methylator phenotype-low; COAD, colon adenocarcinoma; HNSC, head and neck cancer; LGG, low grade glioma; LUAD, lung adenocarcinoma.

Hypermethylation in promoter of *CD274* (PD-L1) and human leukocyte antigen correlates decreased corresponding RNA expression

Promoter methylation is associated with decreased transcription. By using hypermethylation and hypomethylation of my definition, I examined whether this is also happening in *CD274* gene and human leukocyte antigen genes (*HLA-A*, *HLA-B*, *HLA-C*). Promoter methylation count was calculated by the number of hypermethylated CpG sites subtracted by the number of hypomethylated CpG sites

in each sample. First for *CD274*, there were 5 CpG sites in promoter of *CD274* available for evaluation. Promoter methylation counts of CD274 showed significant negative correlation with *CD274* expression in pan-cancer analysis (**Figure 5**, Spearman rho = -0.12, $p < 0.001$). Furthermore, negative correlations were observed in 25 out of 33 cancer types, statistically significant in 17 cancer types. (**Figure 5, Table 1**). For HLA genes, there were 18, 26 and 44 CpG sites in promoter available for evaluation in each HLA gene, respectively (*HLA-A*, *HLA-B*, *HLA-C*). All three genes showed negative correlations with corresponding gene expressions in pan-cancer analysis (Spearman rho = -0.16, $p < 0.001$ for *HLA-A*; Spearman rho = -0.34, $p < 0.001$ for *HLA-B*; Spearman rho = -0.19, $p < 0.001$ for *HLA-C*) and also in majority of individual cancer types (21, 32, and 29 of 33 cancer types, respectively, **Figure 6, Figure 7, Table 1**). As these three molecules are crucial in antigen recognition by CD8 T cells, I compared the methylation status of 3 HLA genes with CytAct, a representative score of CD8+ T cell activity as well as tumor immunogenicity. The overall methylation status of HLA promoter showed significant negative correlation with CytAct (Spearman rho = -0.21, $p < 0.001$, **Figure 8**). These results suggest that part of tumor immunogenicity that is determined by PD-L1 and HLA molecules is influenced by methylation regulation. In addition, my approach in defining hypermethylation and hypomethylation may represent regulation of methylation in gene expression.

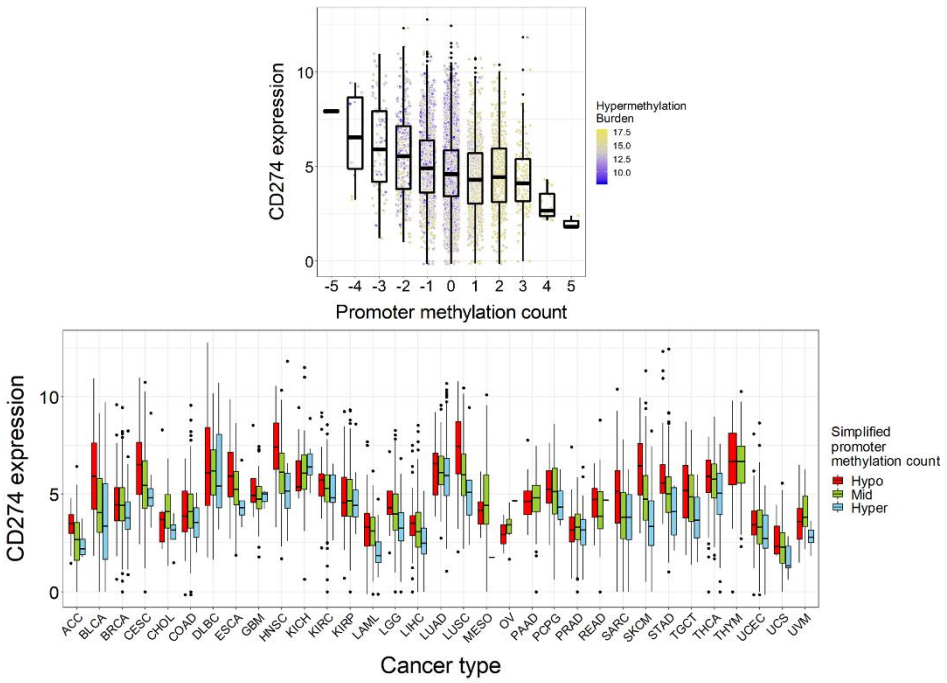

Figure 5. Promoter methylation and expression correlation in *CD274*. Figures show promoter methylation and expression correlation in CD274. The upper plot shows correlation of promoter methylation count with CD274 expression. CD274 expression values in each promoter methylation count group is shown by a boxplot of which horizontal lines represent the upper 25%, median, and lower 25% CD274 expression values. In the plot, samples with promoter methylation counts of more than 5 (5) were combined with that of 5 (-5), positive being hypermethylated and negative being hypomethylated. Each dot is colored by hypermethylation burden in gradient, blue being lowest value and yellow being the highest value. Promoter methylation counts and CD274 expression showed significant negative correlation (Spearman rho = -0.12, $p < 0.001$). The lower plot shows the correlation by each cancer type. The lower plot shows the correlation in each cancer type. Low promoter methylation count stands for the value of less than 0, mid for 0, and high for more than 0. Total of 25 out of 33 cancer types showed negative correlations,

17 of them being statistically significant.

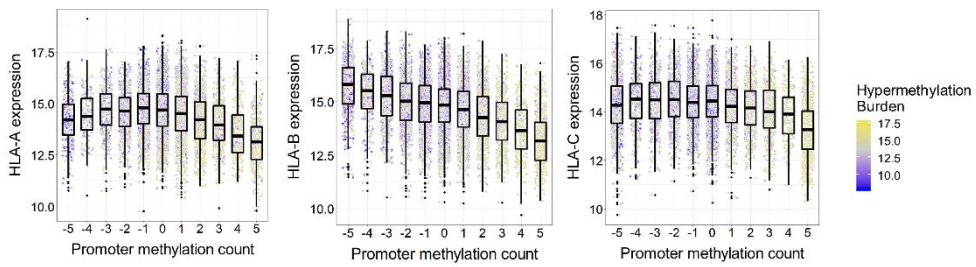

Figure 6. Promoter methylation and expression correlation in *HLA-A, HLA-B and HLA-C*. The three plots show the correlation of promoter methylation count and the corresponding gene, *HLA-A*, *HLA-B* and *HLA-C*, respectively in a same way as depicted in Figure 5 upper plot.

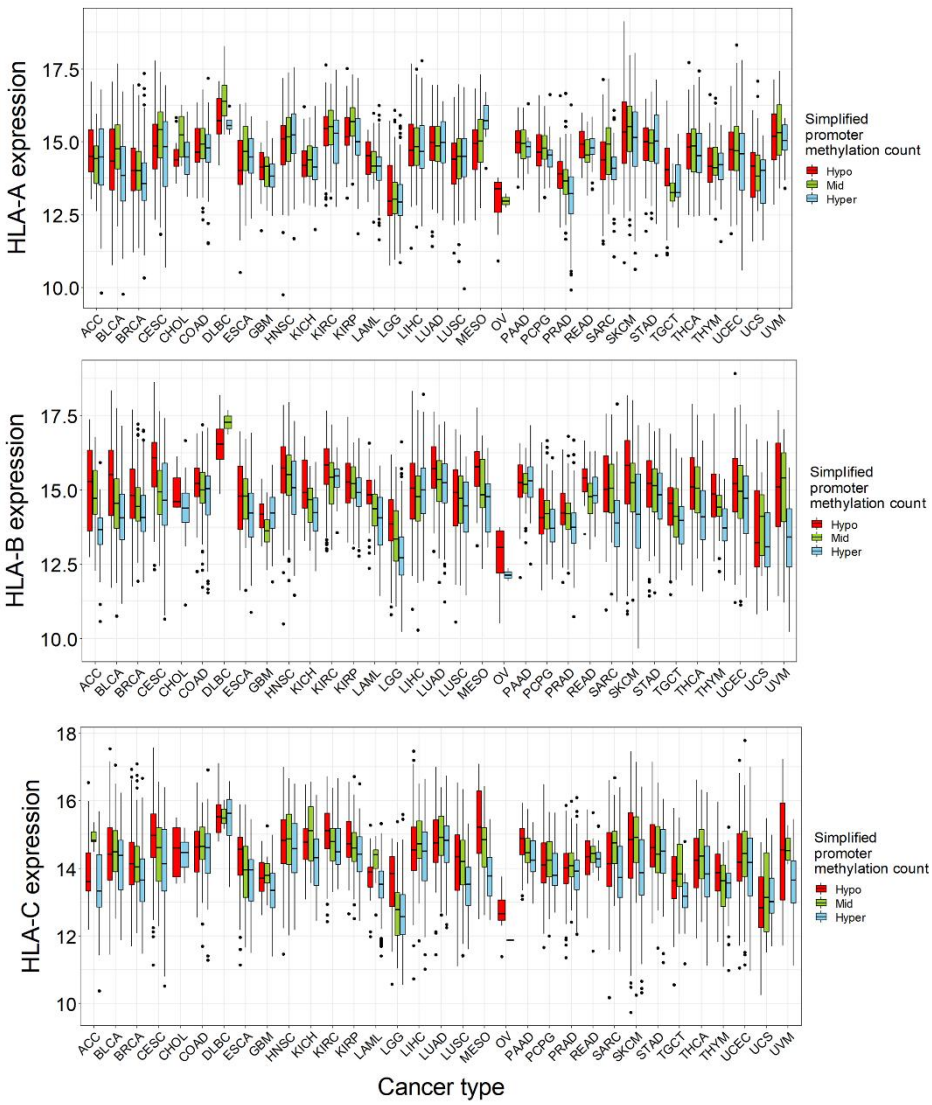

Figure 7. Promoter methylation and expression correlation in *HLA-A, HLA-B and HLA-C* in each cancer type. Each plot shows the correlations of promoter methylation count and expression of corresponding gene (*HLA-A*, *HLA-B*, *HLA-C*) by each cancer type. Low promoter methylation count stands for the value of less than 0, mid for 0, and high for more than 0. Total of 21, 32, and 29 out of 33 cancer types showed negative correlations, respectively for

HLA-A, *HLA-B*, and *HLA-C*. Among them, 6, 25, and 15 were statistically significant, respectively. Boxplot of which horizontal lines represent the upper 25%, median, and lower 25% CD274 expression values.

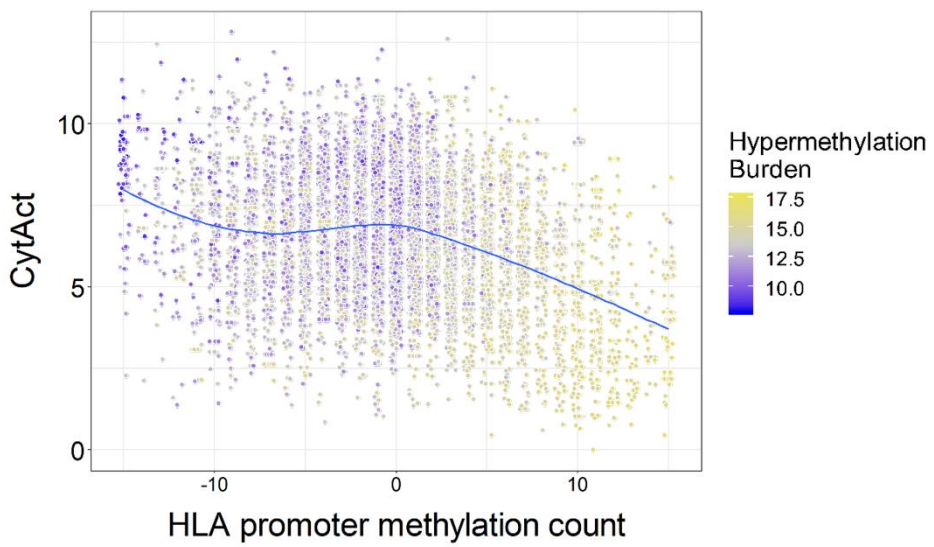

Figure 8. Promoter methylation of HLA associated with decreased CytAct. The lower plot shows the correlation of HLA promoter methylation count and CytAct, which was significantly negatively correlated (Spearman rho = -0.21, $p < 0.001$). Each dot is colored by hypermethylation burden in gradient, blue being lowest value and yellow being the highest value.

Table 1. Correlation of promoter methylation count and corresponding gene expression in each cancer type

Cancer type	*CD274*		*HLA-A*		*HLA-B*		*HLA-C*	
	Spearman rho	p-value	Spearman rho	p-value	Spearman rho	p-value	Spearman rho	p-value
ACC	-0.255	0.024	-0.058	0.617	-0.498	<0.001	-0.220	0.064
BLCA	-0.294	<0.001	-0.057	0.245	-0.411	<0.001	-0.023	0.676
BRCA	-0.112	0.002	-0.210	<0.001	-0.320	<0.001	-0.166	<0.001
CESC	-0.222	<0.001	0.027	0.656	-0.442	<0.001	-0.236	<0.001
CHOL	-0.107	0.552	0.145	0.452	-0.331	0.123	0.228	0.453
COAD	0.016	0.783	-0.060	0.295	-0.188	0.001	-0.075	0.230
DLBC	-0.088	0.552	0.276	0.060	-0.238	0.144	-0.083	0.670
ESCA	-0.220	0.003	0.179	0.017	-0.210	0.005	-0.191	0.019
GBM	-0.096	0.502	-0.156	0.255	-0.030	0.827	-0.182	0.197
HNSC	-0.316	<0.001	0.118	0.006	-0.194	<0.001	-0.032	0.502
KICH	0.105	0.403	-0.062	0.572	-0.375	<0.001	-0.234	0.048
KIRC	-0.102	0.073	-0.074	0.163	-0.274	<0.001	-0.260	<0.001
KIRP	0.005	0.940	-0.049	0.412	-0.183	0.002	-0.260	<0.001
LAML	-0.294	<0.001	-0.160	0.041	-0.475	<0.001	-0.287	0.001
LGG	-0.267	<0.001	-0.108	0.015	-0.342	<0.001	-0.212	<0.001
LIHC	-0.167	0.001	-0.022	0.660	-0.060	0.242	-0.065	0.261
LUAD	-0.119	0.011	0.000	0.999	-0.156	0.001	-0.074	0.143
LUSC	-0.366	<0.001	0.075	0.135	-0.103	0.041	-0.103	0.060
MESO	0.005	0.963	0.269	0.012	-0.376	<0.001	-0.476	<0.001
OV	0.359	0.343	0.046	0.907	0.255	0.507	-0.220	0.569
PAAD	0.036	0.650	-0.070	0.370	-0.021	0.787	-0.244	0.005
PCPG	-0.108	0.151	0.030	0.691	-0.143	0.052	-0.078	0.344
PRAD	0.039	0.391	-0.357	<0.001	-0.346	<0.001	-0.110	0.019
READ	-0.191	0.066	-0.044	0.673	-0.261	0.016	-0.090	0.466
SARC	-0.246	<0.001	-0.018	0.794	-0.337	<0.001	0.008	0.910
SKCM	-0.335	<0.001	-0.176	<0.001	-0.400	<0.001	-0.209	<0.001
STAD	-0.242	<0.001	0.039	0.469	-0.161	0.002	-0.093	0.107
TGCT	-0.325	<0.001	-0.114	0.156	-0.272	0.001	0.015	0.862
THCA	-0.217	<0.001	-0.036	0.403	-0.346	<0.001	-0.174	<0.001
THYM	-0.109	0.236	-0.035	0.708	-0.436	<0.001	-0.196	0.037
UCEC	-0.110	0.025	-0.142	0.003	-0.268	<0.00	-0.052	0.304

						1		
UCS	-0.231	0.084	-0.023	0.877	-0.107	0.476	0.063	0.728
UVM	0.031	0.786	0.005	0.963	-0.455	<0.001	-0.339	0.003

Negative correlation of methylation burden with immunogenicity in the tumor microenvironment

I expanded my observations to evaluate whether overall methylation aberrancy represented by MetB correlates with tumor immunogenicity. First, I used CytAct to evaluate the correlation between the degree of methylation aberrancy and immunogenicity. MetB was significantly negatively correlated with CytAct in pan-cancer-based analysis (Spearman rho = -0.37, $p < 0.001$, **Figure 9**).

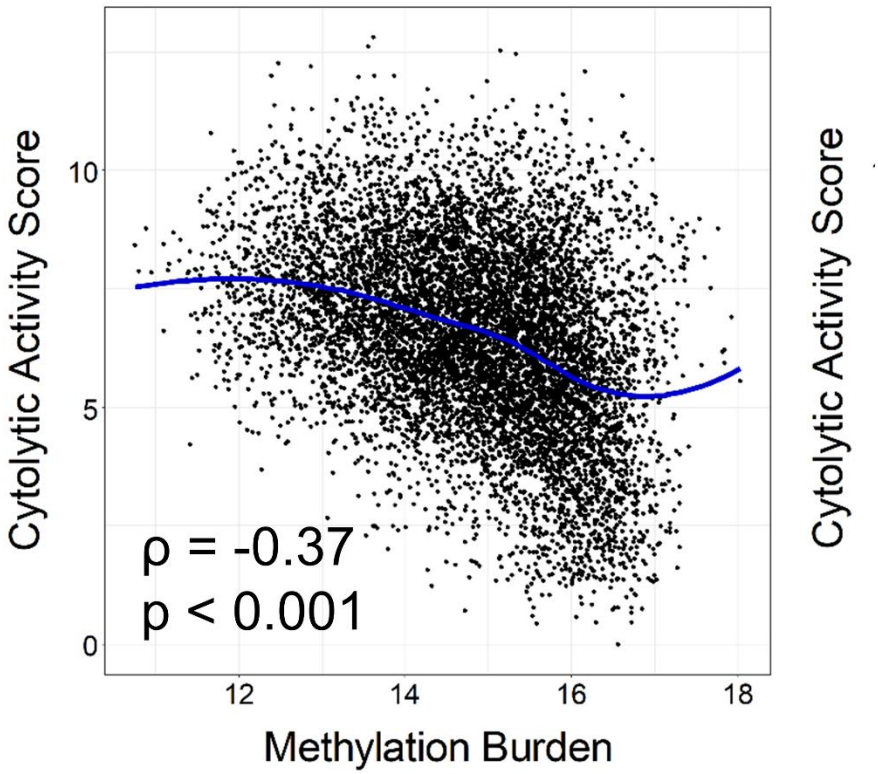

Figure 9. Association between methylation burden and immunogenicity. Dot plots of MetB and CytAct show negative correlation. Each dot represents individual samples from TCGA. Pan-cancer-based analysis is shown. A locally weighted scatterplot smoothing regression line is drawn the plot. MetB showed significant negative correlation with CytAct in pan-cancer analysis (Spearman rho = -0.37, $p < 0.001$).

This significant negative correlation was consistent in majority of cancer types (30 of 33 types) evaluated (Spearman rho value range -0.66 - 0.28, median -0.38, **Figure 10, Table 2**), with an exception of non-significant positive correlation in testicular germ cell tumor (TGCT), which had the widest range of MetB

probably due to their variable methylation status origin[26]. I also noted that cancer types with higher median MetB tended to have higher median CytAct (Spearman rho = -0.52, p = 0.003, **Figure 11, Figure 12**) with an exception of diffuse large B cell lymphoma (DLBC) and acute myeloid leukemia (LAML) which shows consistently high CytAct considering MetB compared to other cancer types.

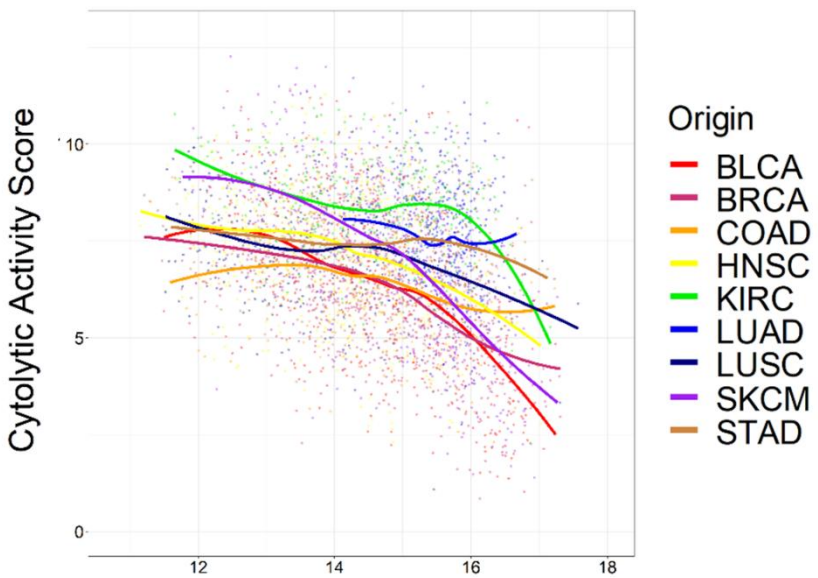

Figure 10. Association between methylation burden and immunogenicity in major cancer types. Dot plots of MetB and CytAct for selected cancer types are shown on the right. A locally weighted scatterplot smoothing regression line is drawn on each plot (Spearman rho = -0.53, $p < 0.001$ for BLCA; Spearman rho = -0.52, $p < 0.001$ for BRCA; Spearman rho = -0.30, $p < 0.001$ for COAD; Spearman rho = -0.27, $p < 0.001$ for HNSC; Spearman rho = -0.19, $p < 0.001$ for KIRC; Spearman rho = -0.11, $p = 0.02$ for LUAD; Spearman rho = -0.22, $p < 0.001$ for LUSC; Spearman rho = -0.66, $p < 0.001$ for SKCM; and Spearman rho = -0.08, $p < 0.12$ for STAD).

Table 2. Correlation of methylation burden and cytolytic activity scores for each cancer type.

Cancer type	Abbreviation	Sample_number	Spearman rho value	p-value
Adrenocortical carcinoma	ACC	78	-0.595	<0.001
Bladder urothelial carcinoma	BLCA	408	-0.533	<0.001
Breast invasive carcinoma	BRCA	778	-0.518	<0.001
Cervical squamous cell carcinoma	CESC	303	-0.315	<0.001
Cholangiocarcinoma	CHOL	36	-0.544	0.001
Colon adenocarcinoma	COAD	288	-0.295	<0.001
Diffuse large B cell lymphoma	DLBC	48	-0.483	0.001
Esophageal carcinoma	ESCA	182	-0.334	<0.001
Glioblastoma multiforme	GBM	52	-0.638	<0.001
Head and neck squamous cell carcinoma	HNSC	515	-0.271	<0.001
Chromophobe renal cell carcinoma	KICH	65	-0.387	0.002
Clear cell renal cell carcinoma	KIRC	311	-0.186	0.001
Papillary renal cell carcinoma	KIRP	270	-0.534	<0.001
Acute myeloid leukemia	LAML	170	-0.347	<0.001
Brain lower grade glioma	LGG	514	-0.364	<0.001
Liver hepatocellular carcinoma	LIHC	368	-0.429	<0.001
Lung adenocarcinoma	LUAD	452	-0.110	0.020
Lung squamous cell carcinoma	LUSC	364	-0.219	<0.001
Mesothelioma	MESO	87	-0.459	<0.001
Ovarian serous cystadenocarcinoma	OV	9	-0.217	0.581
Pancreatic adenocarcinoma	PAAD	177	-0.432	<0.001
Pheochromocytoma and paraganglioma	PCPG	178	-0.400	<0.001
Prostate adenocarcinoma	PRAD	494	-0.447	<0.001
Rectum adenocarcinoma	READ	94	-0.042	0.684
Sarcoma	SARC	255	-0.617	<0.001
Skin cutaneous melanoma	SKCM	469	-0.658	<0.001
Stomach adenocarcinoma	STAD	370	-0.081	0.120
Testicular germ cell tumor	TGCT	149	0.279	0.001
Thyroid cancer	THCA	501	-0.602	<0.001
Thymoma	THYM	120	-0.397	<0.001
Uterine corpus endometrial carcinoma	UCEC	417	-0.456	<0.001
Uterine carcinosarcoma	UCS	57	-0.380	0.004
Uveal melanoma	UVM	80	-0.663	<0.001

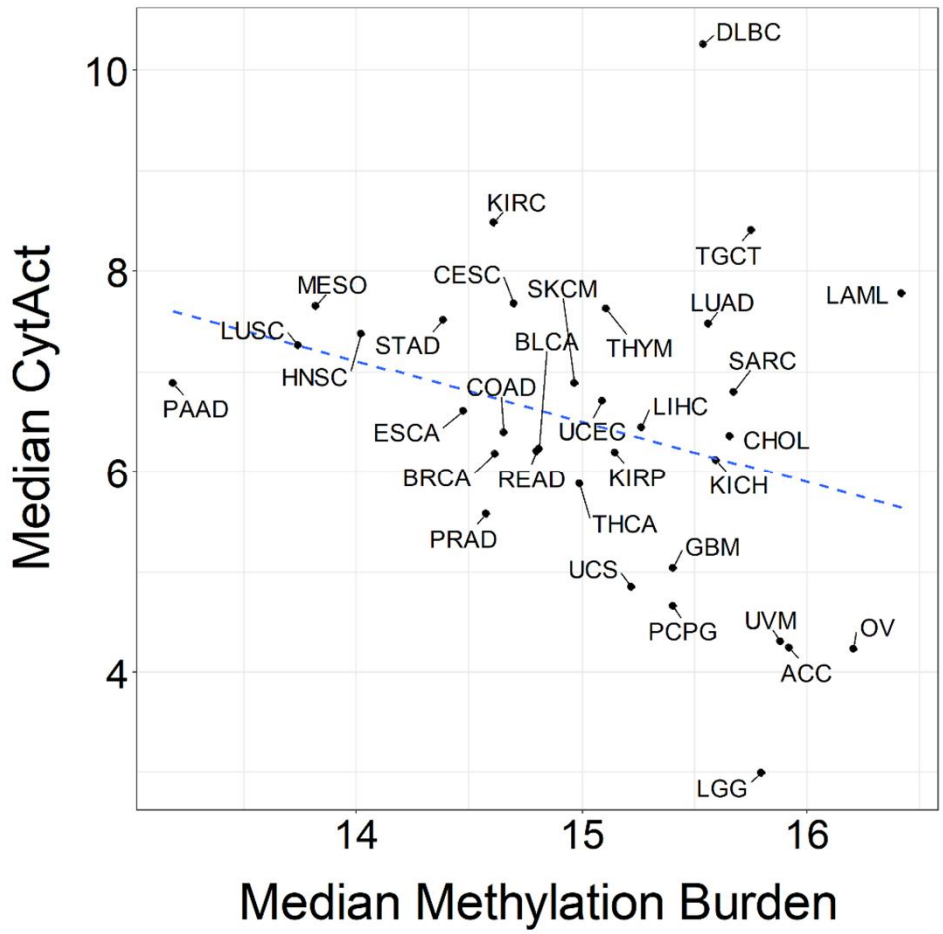

Figure 11. Association between methylation burden and immunogenicity by cancer types. The scatterplot of median MetB and median CytAct of each cancer type. Each dot is annotated with corresponding cancer type. The blue dashed line shows linear regression lines when all cancer types were included for analysis showing tendency of negative correlation (Spearman rho = -0.33, p = 0.06). The analysis of median MetB and median CytAct after excluding DLBC and LAML, both of which have consistently high CytAct compared to other cancer types considering MetB, showed significant negative correlation (Spearman rho = -0.52, p = 0.003).

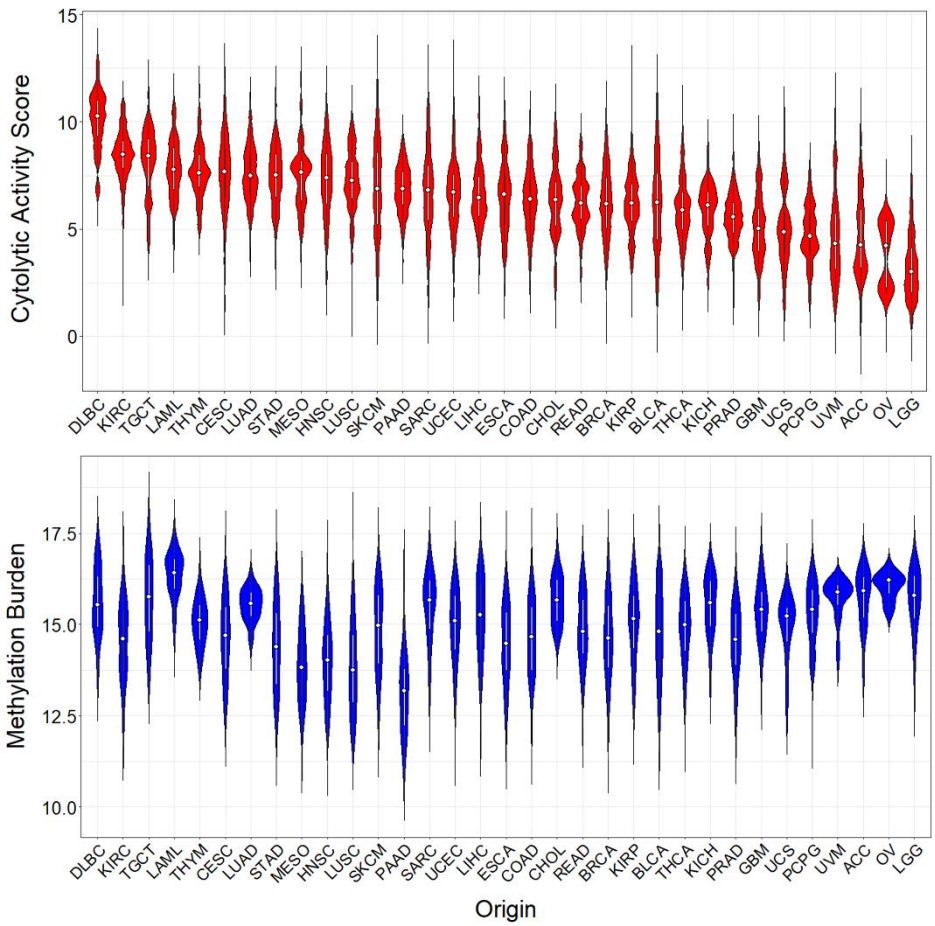

Figure 12. Violin plots of CytAct and MetB of individual cancer type. Violin plot of CytAct (upper) and MetB (lower) of each cancer type. Inside each violin plot, median value is marked with white dot, and values between upper 25% and lower 25% are marked by a white box. The order from left to right starts with the highest mean CytAct value. All cancer types available at TCGA are included in this figure.

The negative correlation was also observed from each of hypermethylation burden and hypomethylation burden with CytAct (**Figure 13**). Still, the regression model that predicts CytAct with the combination of

hypermethylation burden and hypomethylation burden showed significantly more fit than that with either one of hypermethylation burden or hypomethylation burden in pan-cancer analysis ($p < 0.001$ for both comparisons). In addition, the combined model was significantly more fit in 23 of 33 cancer types compared to hypomethylation burden and in 29 of 33 cancer types compared to hypermethylation burden (**Table 3**). These findings indicate that CytAct is associated with aberrancy of both hypermethylation and hypomethylation.

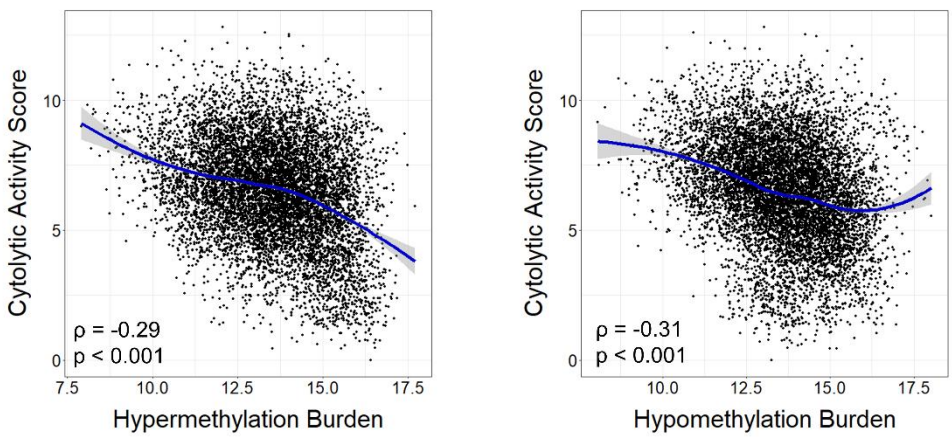

Figure 13. Association between hypermethylation burden/hypomethylation burden and immunogenicity. Association between hypermethylation burden/hypomethylation burden and CytAct. Each dot represents individual sample from TCGA. Locally weighted scatterplot smoothing regression line is drawn on each plot. Both hypermethylation burden and hypomethylation burden showed significant negative correlation with CytAct (Spearman rho = -0.29, $p < 0.001$ and spearman rho -0.31, $p < 0.001$, respectively)

Table 3. Comparison of regression model to predict CytAct with combination of hypomethylation burden and hypermethylation burden compared to that with either one of the two burdens.

Cancer type	Hypomethylation burden + Hypermethylation burden vs. Hypomethylation burden (p-value)	Hypomethylation burden + Hypermethylation burden vs. Hypermethylation burden (p-value)
ACC	<0.001	<0.001
BLCA	<0.001	<0.001
BRCA	<0.001	<0.001
CESC	0.033	<0.001
CHOL	0.173	<0.001
COAD	0.153	<0.001
DLBC	0.005	0.092
ESCA	0.722	<0.001
GBM	0.003	<0.001
HNSC	0.447	<0.001
KICH	0.958	0.009
KIRC	<0.001	0.565
KIRP	<0.001	<0.001
LAML	<0.001	0.008
LGG	<0.001	0.001
LIHC	<0.001	<0.001
LUAD	0.016	<0.001
LUSC	0.449	<0.001
MESO	0.393	0.002
OV	0.059	0.497
PAAD	<0.001	0.012
PCPG	<0.001	<0.001
PRAD	<0.001	<0.001
READ	0.423	0.919
SARC	0.001	<0.001
SKCM	<0.001	<0.001
STAD	<0.001	<0.001
TGCT	<0.001	0.011
THCA	<0.001	<0.001
THYM	<0.001	0.008
UCEC	<0.001	<0.001
UCS	0.189	<0.001
UVM	<0.001	0.023

Abbreviations are provided in **Table 2.**

This finding was consistent when CytAct was substituted with an interferon gamma gene signature score or immune signature score, as defined in previous literatures[7,8] (Spearman rho = -0.42, $p < 0.001$ and spearman rho = -0.34, $p < 0.001$, respectively, **Figure 14**). As the immune subtypes described by Thorsson et al[25] also represent the immune-related tumor microenvironment, I evaluated the association between MetB and immune subtypes (**Figure 15**). C4 and C5 type tumors, which are immune suppressive, had significantly higher MetB compared to C2 and C3 type tumors, deemed as immunogenic.

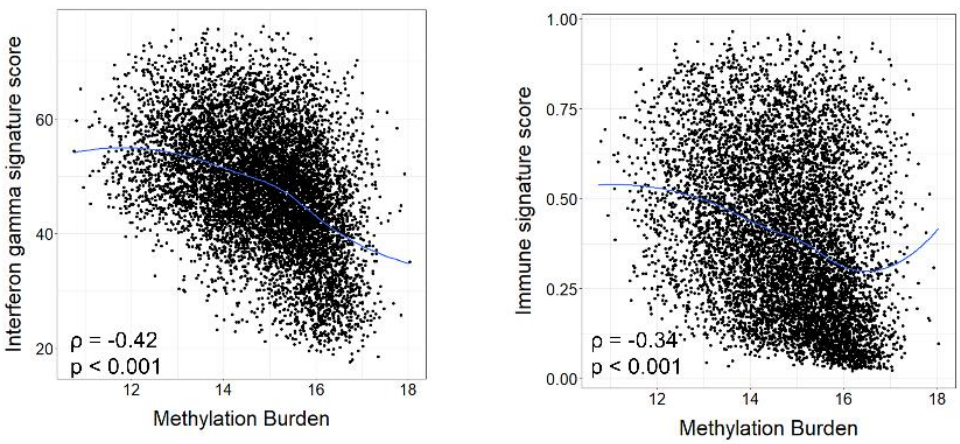

Figure 14. Association of methylation burden and other signatures. Association between methylation burden and immunogenicity; interferon gamma gene signature[1] and immune signature score[2]. Dot plots on the left (right) show correlation of MetB and immune signature score (interferon gamma signature score). Each dot represents individual sample from TCGA. Linear regression line is drawn on each plot. Immune signature scores were imported from supplementary data provided in the previous literature1 and interferon gamma gene signature scores were calculated by geometric mean

of 6 genes (*IDO1*, *CXCL9*, *CXCL10*, *STAT1*, *IFNG*) as described in the literature2. Both interferon gamma signature score and immune signature score showed significant negative correlation with CytAct (Spearman rho = -0.42, $p < 0.001$ and spearman rho -0.34, $p < 0.001$, respectively)

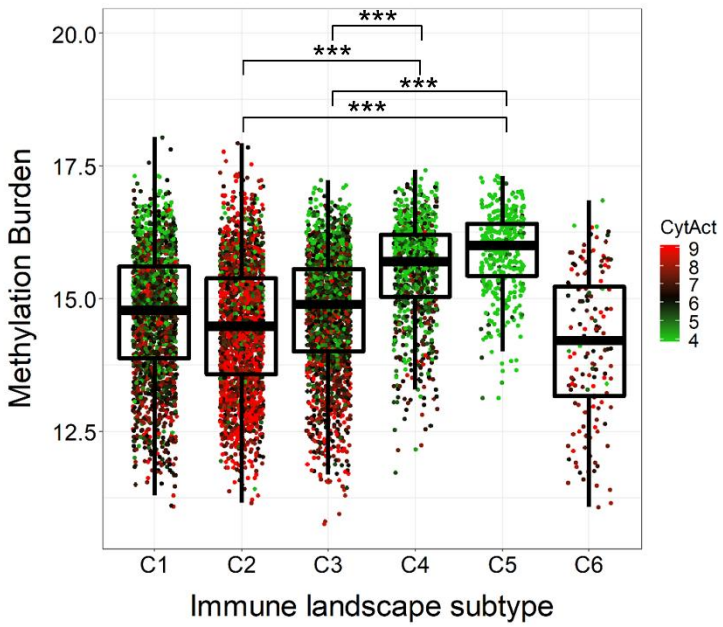

Figure 15. Correlation of MetB and immune landscape subtypes. Boxplots showing MetB for each immune landscape subtype described by Thorsson et al; C1, wound healing; C2, IFN-γ dominant; C3, inflammatory; C4, lymphocyte depleted; C5, immunologically quiet; C6, TGF-β dominant. Each dot represents individual samples and was color coded according to CytAct scores, scaled from green (4 or less) to red (9 or more). Horizontal lines in the box represent the upper 25%, median, and lower 25% MetB values for each immune subtype. On the right table shows names of each immune landscape subtype. Significant levels are depicted as asterisks; ns, not significant; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

I additionally evaluated whether aberrant methylation cutoff other than 5% would show similar correlations with CytAct, and found that all the cutoffs ranging from 1% to 3% showed negative correlation with CytAct, with the Spearman coefficients ranging from -0.30 ~ -0.44. However, when the cutoff was more than 5%, the negative correlation was largely driven by the number of CpG sites that are aberrantly methylated as the same side of the skewedness of each CpG site (**Table 4**). This may implicate that the cutoff more than 5% may result in including normomethylated CpG site. As the correlation coefficient was the highest with the cutoff of 5% compared to 2.5% and 1%, this suggests that cutoff of 5% may be the most appropriate for my aim to develop the MetB.

Table 4. Correlation coefficients of CytAct with methylation burden calculated by various cutoffs

Cutoff (%)	**Whole sites**	**Skewed sides only***	**Opposite to skewed sides only**
1	-0.30	-0.33	-0.33
2.5	-0.34	-0.33	-0.37
5	-0.37	-0.31	-0.41
10	-0.38	-0.25	-0.43
12.5	-0.39	-0.22	-0.44
15	-0.40	-0.18	-0.44
17.5	-0.41	-0.14	-0.44
20	-0.42	-0.09	-0.44
22.5	-0.42	-0.05	-0.43
25	-0.43	-0.01	-0.43
27.5	-0.43	0.04	-0.42
30	-0.44	0.08	-0.42

*: In analysis of skewed sides only, the methylation burden was defined as log 2 value of the number of CpG sites that are aberrantly methylated as the same side of the skewedness of each CpG sites.

Similar to previous findings that the degree of chromosomal instability (CIN) correlated with tumor purity, I hypothesized that MetB would also be correlated with tumor purity[7,27,28]. Pan-cancer analysis showed that higher tumor purity, derived from consensus purity estimates[27], was significantly correlated with higher MetB (Spearman rho = 0.50, $p < 0.001$) in the same manner as the CIN score. To clarify if there was significant correlation between MetB and CytAct according to tumor purity, I classified the samples into 6 groups with similar tumor purity[27]. Analysis showed that the MetB for samples with similar tumor purity, especially those with purity of more than 0.98, had a negative correlation with the CytAct[27] (**Figure 16**).

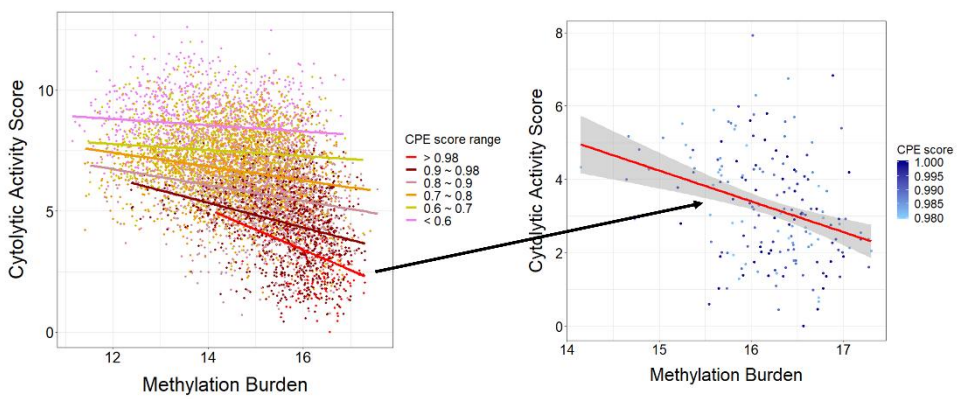

Figure 16. Correlation of MetB and CytAct according to tumor purity. Tumor purity was defined as consensus measurement of purity estimations described by Dvir et al[1]. The tumor purity value was expressed in a value between 0 to 1 which represents estimated proportion of true tumor cells in whole tumor sample. All samples were divided in to 6 groups according to tumor purity; very-high (> 0.98), high (0.9 ~ 0.98), medium-high (0.8 ~ 0.9), medium-low (0.7 ~ 0.8), low (0.6~0.7) and very-low (< 0.6). Correlations of MetB and CytAct were consistently negative

in all 6 purity groups (left plot). Specifically, I note that MetB and CytAct have negative correlation in very-high purity group, suggesting that MetB of tumor cells contribute to immunogenicity (right plot). CPE, consensus measurement of purity estimation.

To validate my findings of the negative correlation using the definition of MetB and cutoffs of my analysis, I used the external cohorts with publicly available datasets for LUAD[29] and LGG[20], which consist of methylation and RNA sequencing data. In both datasets, MetB and CytAct had negative correlations (Spearman rho = -0.41, $p < 0.001$ in the LUAD dataset and Spearman rho = -0.34, $p = 0.014$ in the LGG dataset, **Figure 17**).

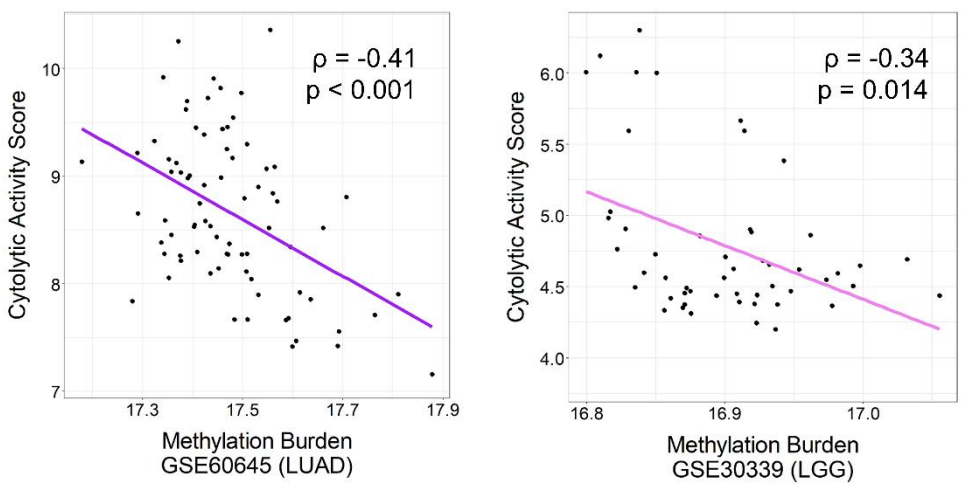

Figure 17. Validation of MetB and CytAct correlation with external datasets. Each dot represents a sample, and a linear regression line is drawn in each plot. The LUAD cohort (left) is from Anna Karlsson et al, and the LGG cohort (right) is from Sevin Turcan et al. The datasets were available and downloaded from

https://www.ncbi.nlm.nih.gov/geo/ (GSE60645 and GSE30339, respectively). In both cohorts, MetB showed significant negative correlation with CytAct (Spearman rho = -0.41, $p < 0.001$ for GSE60645 and spearman rho = -0.34, $p = 0.014$ for GSE30339). LUAD, lung adenocarcinoma; LGG, brain low grade glioma.

I also evaluated the association of MetB with clinical data of immune checkpoint inhibitor outcomes in independent cohort consists of TCGA SKCM patients. I downloaded clinical outcome data of TCGA melanoma patients, which included 22 patients treated with ipilimumab (anti-CTLA4 antibody). Since response rate and 3-year overall survival rate were both around 20% in previous clinical trials[30,31], I used bottom 20% MetB value of whole TCGA SKCM samples as a cutoff to discriminate MetB-high and MetB-low groups. In the 22 patients who were treated with ipilimumab, 4 patients were MetB-low group and 18 patients were MetB-high group. Progression-free survival (PFS) was significantly shorter in MetB-high group than MetB-low group (Median progression free survival 2.3 months vs 10.6 months, hazard ratio 4.92, $p = 0.029$, **Figure 18**). Overall, consistent results of MetB in association with tumor immunogenicity and immunotherapy outcomes suggest that the degree of methylation aberrancy is associated with immune evasion of tumor.

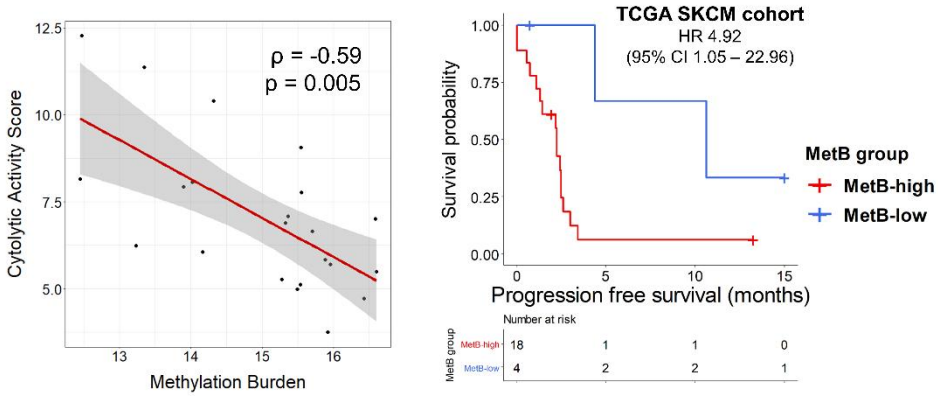

Figure 18. Validation of MetB and immunogenicity correlation with immunotherapy datasets. Kaplan-Meier curve of TCGA SKCM patients treated with ipilimumab shows PFS. Red line represents MetB-high group, defined as samples with MetB higher than bottom 20% value of whole MetB in SKCM samples. Blue line represents MetB-low group, defined as samples with the bottom 20% or lower MetB. Censored data are marked with short vertical segments. MetB-high group had significantly poor PFS compared to MetB-low group (Hazard ratio 4.92, 95% confidence interval 1.05 – 22.96). HR, hazard ratio; CI, confidence interval.

Methylation aberration in addition to mutation burden and copy number alterations predicting immunogenicity

Given that MetB is associated with CytAct, I evaluated whether the epigenetic marker would be an independent variable in determining immunogenicity to mutation burden and CIN score, both of which are previously known to affect immunogenicity[7]. The mutation burden showed positive correlation with CytAct, while the CIN score showed a negative correlation, as previously reported

(Spearman rho = 0.19, $p < 0.001$ for mutation burden and spearman rho = -0.22, $p < 0.001$ for CIN score)[6,7,32]. In 5,241 samples that had all three variables available, multivariate regression analysis showed that MetB significantly and independently determined CytAct along with the mutation burden and CIN score (**Figure 19, Table 5**).

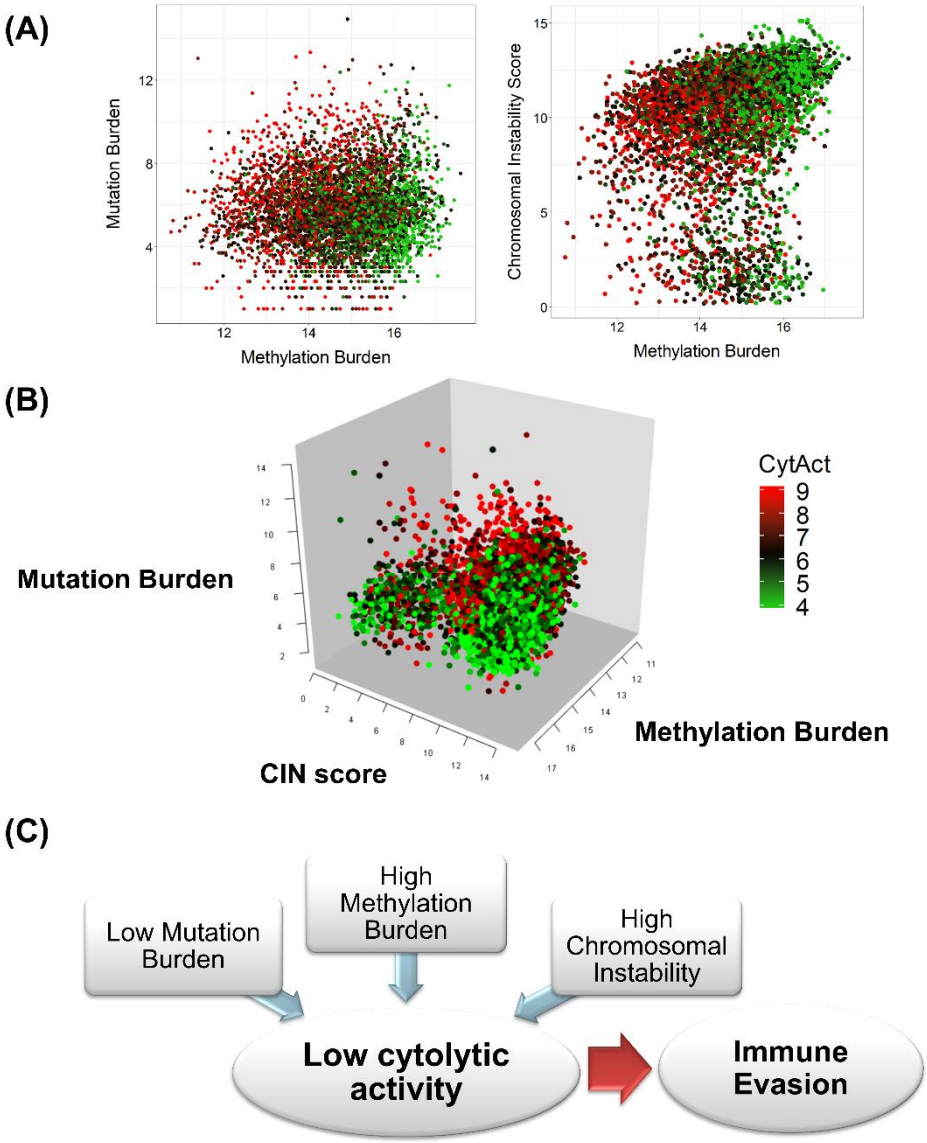

Figure 19. Association between cytolytic activity scores and 3

epigenomic/genomic variables: methylation burden, mutation burden and chromosomal instability score. Each dot represents an individual sample and was color coded according to CytAct scores, scaled from green (4 or less) to red (9 or more). (A) 2-D plots showing association between cytolytic activity and epigenomic/genomic variables. Left and right plots show association between MetB and mutation burden or chromosomal instability (CIN) scores, respectively. (B) 3-D plots showing association between cytolytic activity and 3 epigenomic/genomic variables. DLBC and LAML were not included in the figure since they consistently showed high CytAct. Also, TGCT were also not included in the figure since the MetB in the samples was highly variable. DLBC, diffuse large B cell lymphoma; LAML, acute leukemia; TCGA, testicular germ cell tumor. (C) Schematic figure on how methylation burden, mutation burden and chromosomal instability affect decrease of CytAct resulting in immune evasion. My analysis showed that lower mutation burden, higher MetB, higher CIN scores are associated with lower CytAct, implicating more immune evasion of tumor.

Table 5. Statistical values from multivariate regression analysis of the association between cytolytic activity scores and 3 epigenomic/genomic variables

	Univariate analysis		**Multivariate analysis**	
	Coefficient	*P*-value	Coefficient	*P*-value
Mutation burden	0.14 ± 0.01	$< 2 \times 10^{-16}$	0.20 ± 0.01	$< 2 \times 10^{-16}$
CIN score	-0.08 ± 0.01	$< 2 \times 10^{-16}$	-0.10 ± 0.01	$< 2 \times 10^{-16}$
Methylation burden	-0.49 ± 0.02	$< 2 \times 10^{-16}$	-0.44 ± 0.02	$< 2 \times 10^{-16}$

Selection of CpG sites to assess potential of methylation aberrancy as a biomarker

Finally, I determined CpG sites in which changes in methylation status were associated with significant CytAct change. I found that hypermethylation in 601 CpG sites and hypomethylation in 768 CpG sites were associated with significant CytAct change throughout pan-cancer and most of the major cancer types. Gene ontology (GO) analysis of genes that are related to these CpG sites showed the most significant enrichment in transcriptional regulation networks in embryonic stem cells, signatures that are often shared by cancer cells[33]. In addition, CpG sites also showed enrichment in pathways associated with immunogenicity, including antigen presentation (**Figure 20**). These findings may suggest that aberrant methylation-driven cancers attain distinct immune evasive mechanisms that are accompanied by changes in methylation status in genes associated with immunogenicity.

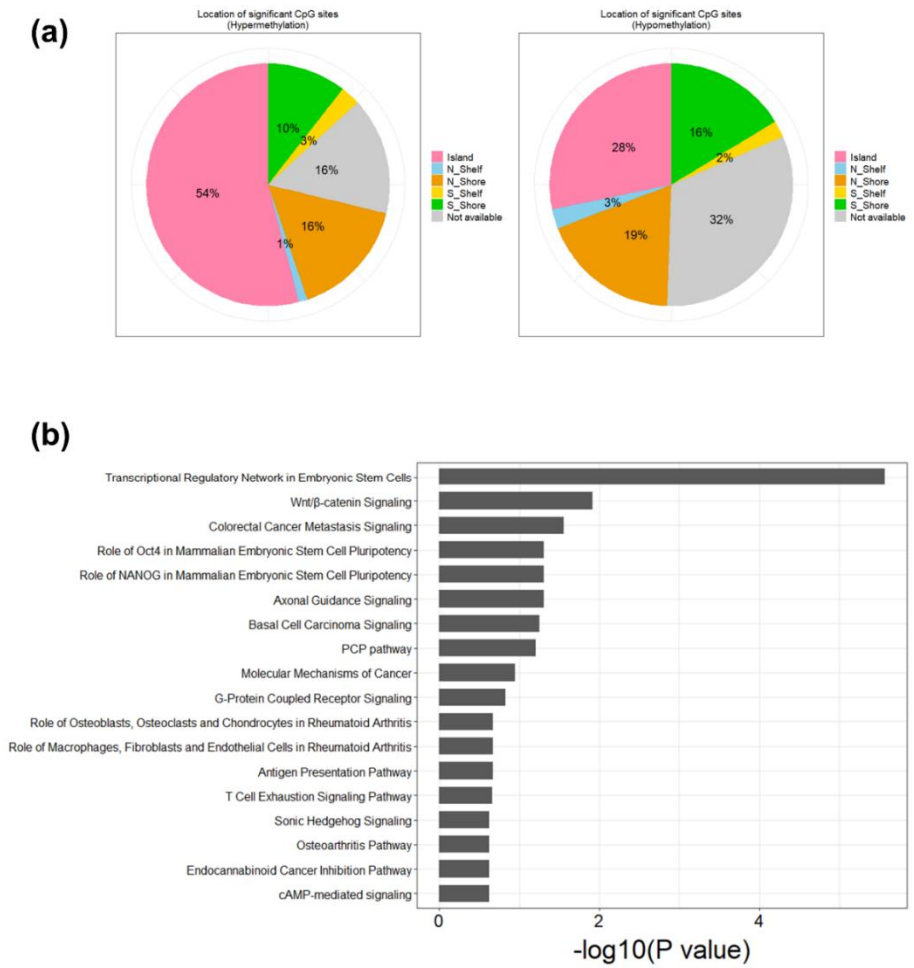

Figure 20. Gene ontology (GO) study of significant CpG sites associated with CytAct. (a) & (b) The location (a) and GO study (b) of CpG sites among all CpG sites in which methylation changes were discovered to be significantly associated with CytAct changes throughout pan-cancer and major cancer types. GO analyses were done using the causal analytic tools, Ingenuity Pathway Analysis (https://www.ingenuity.com). Only top 18 results are shown.

Discussion

In this study, I defined MetB, which reflects the degree of aberrant methylation, and discovered that MetB and CytAct are negatively correlated. This finding was validated in immunotherapy outcome data of TCGA SKCM patients and in external cohorts of LUAD and LGG. Furthermore, MetB predicted CytAct independent of mutation burden and CIN score. These findings implicate that epigenetic modification is involved in generation of distinct immune-related tumor microenvironments.

In this study, methylation aberrancies were determined based not on normal samples but on average pan-cancer methylation status since matched normal samples in TCGA datasets were lacking and confined to a few cell types. Furthermore, recent studies have shown that normal adjacent tissue is not an appropriate control for tumor tissue since it also harbors significant genomic or transcriptomic alterations due to chronologic age[34,35]. Therefore, I assumed that comparison with normal cells in this dataset would lead to bias and setting cutoffs compared to average methylation status would represent a methylation status for tumor samples. Although using the top or bottom 5% value as a cutoff would have specifically reflected a sufficient number of methylated CpG sites, the cutoff would have limited sensitivity in identifying all aberrantly methylated CpG sites.

To evaluate association between MetB and immunogenicity, I used CytAct, a readily applicable score developed by Rooney et al[9] that represents the activity of cytotoxic CD8+ T cells and is associated with a neoantigen profile. CytAct scores have been reported in different cancer types and are correlated with tumor infiltrating lymphocytes[36]. Immune subtypes suggested by Thorsson et al[25]

were identified by comprehensive genomic analyses of immune signatures reflecting various tumor microenvironments. It is notable that the association between MetB and immunogenicity was consistent using both tools. Given that MetB provided additional and seemingly more prediction of CytAct than mutation burden and CIN score regarding CytAct, it is suggested that the methylation aberrancy needs to be evaluated to successfully assess tumor immunogenicity and probability of response to immunotherapy along with the tumor mutation burden.

In fact, biologic rationale for assessing epigenetic changes related to tumor immunogenicity has been described in the previous literatures[37]. Just as epigenetic events can drive cellular growth[38], epigenetic events can also drive immune evasion of tumors[39]. This concept is supported by previous observation on resistance to immunotherapy without any identifiable relevant genomic changes[40]. Consistent with this concept, I showed that promoter hypermethylation is associated with decreased expression of PD-L1, *HLA-A*, *HLA-B* and *HLA-C*. HLA hypermethylation and its implication in immune evasion has been also mentioned in previous study on lung cancer[15]. In addition, a previous report showed that methylation aberration induced by DNA tumor viruses for immune evasion resulted in tumorigenesis[41]. Altogether, the degree of methylation aberrancy might represent increased events of dysregulation of immune-related genes resulting in immune evasion.

Several studies also have reported that methylation changes in specific CpG sites are associated with immunogenicity[10,42]. For example, a set of CpG sites called the EPIMMUNE methylation signature derived from patients treated with anti-PD-1 antibody was investigated and showed an application with methylation data to predict responses to immunotherapy[42]. However, those studies had

limitations in that they either involved only certain types of cancer patients or did not integrate genomic expression profiles represented immune-related tumor microenvironments.

Two recent studies have investigated these issues in pan-cancer analysis of TCGA datasets. One study developed an algorithm called methyl-CIBERSORT, which estimates immune cell profiles in a tumor sample[43]. The other study adapted the RESET algorithm to evaluate differentially expressed transcription profiles and cancer germline antigens according to methylation changes in promoter sites compared to normal tissue[14]. In the present study, I further examined methylation changes throughout all CpG sites. Although the effects of promoter methylation status on RNA transcription and cancer behavior are well known, effects of hypermethylation or hypomethylation on gene transcription in other CpG sites are complex and context-dependent[44]. For example, hypermethylation in the intragenic area is associated with higher expression[45,46], and the methylation status of intra- and intergenic regions have also been described as being associated with cancer[45,47]. Considering these relationships, omitting CpG sites would have resulted in biased assessments on true degree of methylation aberrancy, thereby masking the actual effects of methylation on tumor immunogenicity.

To expand this analysis into clinical applications for various cancer types, further adjustments in β-score cutoffs for CpG sites according to each cancer type would be necessary because methylation status is variable by tissue type, and TCGA datasets tend to include more samples from several major cancer types than other cancer types. Adjustments will be feasible if a large number of samples with methylation array data for each cancer type is collected. In addition, more specific patterns on expression of individual gene by aberrant methylation of each

corresponding CpG site need to be elucidated. Selecting CpG sites as described in the results might be feasible, but more mechanistic approach with validation in multiple cohorts including prospective cohorts is necessary in order to precisely measure biologically and clinically meaningful degree of methylation aberrancy.

Conclusion

In conclusion, I demonstrated that methylation aberrancy drives immune evasion of tumors and assessments on overall methylation aberrancy by MetB can predict the degree of tumor immunogenicity. Based on the findings in this study, further researches and biomarker developments will need to give more focuses on methylation aberrancy to involve whole mechanisms on tumor immunogenicity and immune evasion.

Methods

Dataset acquisition

The TCGA Pan-Cancer Atlas provides normalized, standardized genomic data for 33 TCGA samples[48]. I used human methylation 450K array data composed of 9,664 samples and 396,065 CpG sites and RNA sequencing data, including 20,531 gene expressions from 11,069 samples, from the TCGA Pan-Cancer Atlas. To define and calculate MetB, I used only tumor sample data. If one participant had multiple sample data, only data with a TCGA barcode was used based on alphabetical order. As a result, 8,843 tumor methylation samples remained in downstream analysis. Similarly, I selected only one expression sample per participant and used only data from 10,251 RNA expression samples. Methylation subtypes of TCGA samples of LUAD, HNSC, BRCA and COAD were provided in the previous literatures[21–24], and that of LGG were determined by *IDH1* and *IDH2* mutation status[20].

IlluminaHumanMethylation450kanno.ilmn12.hg19 Bioconductor R package was used to characterize related genes and functional features from 396,065 CpG sites[49].

For external validation, the MetB cutoffs calculated from TCGA Pan-Cancer Atlas data were applied to GEO series GSE30339 and GSE60645 datasets, which contain both methylation 450K array and gene expression array[20,29]. In the GSE30339 dataset, methylation analyses for 81 glioma samples and 53 cell line samples were conducted. Gene expression from 52 glioma samples and 6 cell line samples was analyzed. Among the 81 glioma samples, only 52 had both expression and methylation data[20]. GSE60645 is a dataset with methylation and expression analyses for lung cancer, including 83 adenocarcinomas. Among them, 77

adenocarcinomas had both gene expression and methylation data[29]. I applied the MetB cutoffs obtained from TCGA samples to methylation 450K array data from GSE30339 and GSE60645 and obtained MetB for the samples from each dataset. CytAct scores were obtained from gene expression array data. These datasets were processed using Bioconductor R package GEOquery[50].

Determining β-score cutoffs and definition of methylation burden

I calculated MetB using β values from each of the 8,843 samples and 396,065 CpG sites in the TCGA Pan-Cancer Atlas methylation data. Hypermethylation and hypomethylation were defined as follows:

- Hypomethylated: If the β value of the sample for each CpG site was lower than or equal to the bottom 5% of all samples (8,843), the sample was defined as hypomethylated at that CpG site.
- Hypermethylated: If the β value of the sample for each CpG site was higher than or equal to the top 5% of all samples (8,843), the sample was defined as hypermethylated at that CpG site.

The log 2 value of the number of CpG sites in the sample showing a β value lower than the hypomethylation cutoff was defined as hypomethylation burden, and the log 2 value of the number of CpG sites in the sample showing a β value higher than the hypermethylation cutoff was defined as hypermethylation burden. Finally, the log 2 value of the summation of the number of hypo- and hypermethylation CpG sites was defined as MetB. The percent_rank function of R package dplyr was used to acquire the percentile ranks of β values within each CpG site.

Promoter methylation status and expression determination

I analyzed CpG sites of promoters of *CD274*, *HLA-A*, *HLA-B* and *HLA-C* that have available data. Promoter methylation count a sample was calculated by (the number of hypermethylated CpG sites) – (the number of hypomethylated CpG sites). Expression values of each gene were log 2 normalized.

Genomic profile definition

In TCGA Pan-Cancer Atlas RNA seq data, the following formula was applied to normalized expression values of Granzyme A (*GZMA*) and Periforin-1 (*PRF1*) to measure CytAct[9].

$$CytAct = \frac{\log_2(GZMA + 1) + \log_2(PRF1 + 1)}{2}$$

Immune signature scores were imported from supplementary data provided in previous literature[7], and interferon gamma gene signature scores were calculated based on the geometric mean of 6 genes (*IDO1*, *CXCL9*, *CXCL10*, *STAT1*, *IFNG*), as described in the literature[8].

The mutation burden was defined as the number of nonsynonymous and frameshift mutations, which were curated by Cbioportal with TCGA whole exome sequencing data[51,52]. Definitions and datasets of CIN scores for TCGA samples were imported from previous reports[7]. Both the values of mutation burden and CIN scores were log 2 normalized throughout whole samples. Datasets of tumor purity calculated by genomic and transcriptomic profiles were also obtained from a previous report[27].

Selection of CpG sites

Analysis was conducted to examine immunogenicity variations according to the methylation levels of each CpG site and to identify the CpG sites that interfere with tumor immune responses. Within each CpG site, samples were grouped into 3 categories: hypomethylated (A), hypermethylated (C), and non-extremist normomethylated (B) according to β values.

The mean CytAct of each group was obtained. Fold change in cytolytic activity between hypomethylated and normomethylated groups (A/B) and hypermethylated and normomethylated groups (C/B) were calculated. Wilcoxon rank sum test was conducted for differences between groups to determine the p-value.

To discover significant CpG sites, I first selected CpG sites with fold change of more than 1.25 or less than 0.8, with a Benjamini-Hochberg adjusted p value < 0.05. Of the selected CpG sites, I further selected CpG sites that had same direction of effects to CytAct between pan-cancer and at least 9 of 11 major cancer types (BLCA, BRCA, COAD, HNSC, LGG, LUAD, LUSC, PAAD, READ, SKCM and STAD). GO analyses were performed using the causal analytic tool Ingenuity Pathway Analysis (https://www.ingeuity.com)[53].

Statistical analysis

To evaluation correlation between variables, Spearman's rank correlation coefficient was used. Comparison of continuous values between groups was evaluated with Wilcoxon signed-rank test. Comparison of regression models was analyzed by Analysis of Variance (ANOVA) for regression test. For survival analysis in the TCGA melanoma treatment outcome data, I defined PFS as the time

from start of treatment to the end of the treatment or death. I used Kaplan-Meier method to draw survival curves of PFS and Cox-proportional hazard model to calculate the hazard ratio. P-values less than 0.05 were considered statistically significant. All statistical analyses were performed with R 3.4.3. (https://cran.r-project.org/bin/windows/base/old/3.4.3/)

References

1. D'Errico, G., Machado, H. L. & Sainz, B. A current perspective on cancer immune therapy: step-by-step approach to constructing the magic bullet. *Clin. Transl. Med.* **6**, (2017).
2. Havel, J. J., Chowell, D. & Chan, T. A. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat. Rev. Cancer* **19**, 133–150 (2019).
3. Goodman, A. M. *et al.* Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. *Mol. Cancer Ther.* **16**, molcanther.0386.2017 (2017).
4. Le, D. T. *et al.* PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N. Engl. J. Med.* **372**, 2509–2520 (2015).
5. Hellmann, M. D. *et al.* Nivolumab plus Ipilimumab in Lung Cancer with a High Tumor Mutational Burden. *N. Engl. J. Med.* **378**, 2093–2104 (2018).
6. Davoli, T., Uno, H., Wooten, E. C. & Elledge, S. J. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science (80-.).* **355**, (2017).
7. Ock, C. Y. *et al.* Genomic landscape associated with potential response to anti-CTLA-4 treatment in cancers. *Nat. Commun.* **8**, 1–12 (2017).
8. Ayers, M. *et al.* IFN- γ – related mRNA profile predicts clinical

response to PD-1 blockade. *J. Clin. Invest.* **127**, 2930–2940 (2017).

9. Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61 (2015).

10. Micevic, G., Theodosakis, N. & Bosenberg, M. Aberrant DNA methylation in melanoma: biomarker and therapeutic opportunities. *Clin. Epigenetics* **9**, 1–15 (2017).

11. Robertson, K. D. & Jones, P. A. DNA methylation: past, present and future directions. *Carcinogenesis* **21**, 461–467 (2000).

12. Baylin, S. B. Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. *Hum. Mol. Genet.* **10**, 687–692 (2002).

13. Ehrlich, M. DNA hypomethylation in cancer cells. *Epigenomics* **1**, 239–259 (2010).

14. Saghafinia, S., Mina, M., Riggi, N., Hanahan, D. & Ciriello, G. Pan-Cancer Landscape of Aberrant DNA Methylation across Human Tumors. *Cell Rep.* **25**, 1066-1080.e8 (2018).

15. Rosenthal, R. *et al.* Neoantigen-directed immune escape in lung cancer evolution. *Nature* **567**, 479–485 (2019).

16. Wang, H. *et al.* Genome-wide DNA methylation and transcriptome analyses reveal genes involved in immune responses of pig peripheral blood mononuclear cells to poly I:C. *Sci. Rep.* **7**, 1–11 (2017).

17. Rosenberg, S. A. *et al.* Landscape of immunogenic tumor antigens in

successful immunotherapy of virally induced epithelial cancer. *Science (80-.).* **356**, 200–205 (2017).

18. Jung, H. *et al.* DNA methylation loss promotes immune evasion of tumours with high mutation and copy number load. *Nat. Commun.* **10**, 1–12 (2019).

19. Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11**, (2010).

20. Turcan, S. *et al.* IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature* **483**, 479–483 (2012).

21. Collisson, E. A. *et al.* Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).

22. Lawrence, M. S. *et al.* Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **517**, 576–582 (2015).

23. Koboldt, D. C. *et al.* Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).

24. Liu, Y. *et al.* Comparative Molecular Analysis of Gastrointestinal Adenocarcinomas. *Cancer Cell* **33**, 721-735.e8 (2018).

25. Thorsson, V. V. *et al.* The Immune Landscape of Cancer. *Immunity* **48**, 812-830.e14 (2018).

26. Shen, H. *et al.* Integrated Molecular Characterization of Testicular Germ Cell Tumors. *Cell Rep.* **23**, 3392–3406 (2018).

27. Aran, D., Sirota, M. & Butte, A. J. Systematic pan-cancer analysis of tumour purity. *Nat. Commun.* **6**, 1–11 (2015).

28. Yuan, Y. *et al.* Quantitative Image Analysis of Cellular Heterogeneity in Breast Tumors Complements Genomic Profiling. *Sci. Transl. Med.* **4**, 157ra143-157ra143 (2012).

29. Karlsson, A. *et al.* Genome-wide DNA methylation analysis of lung carcinoma reveals one neuroendocrine and four adenocarcinoma epitypes associated with patient outcome. *Clin. Cancer Res.* **20**, 6127–6140 (2014).

30. Schadendorf, D. *et al.* Pooled analysis of long-term survival data from phase II and phase III trials of ipilimumab in unresectable or metastatic melanoma. *J. Clin. Oncol.* **33**, 1889–1894 (2015).

31. Wolchok, J. D. *et al.* Overall Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma. *N. Engl. J. Med.* **377**, NEJMoa1709684 (2017).

32. Roh, W. *et al.* Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. *Sci. Transl. Med.* **9**, 1 (2017).

33. Kim, J. & Orkin, S. H. Embryonic stem cell-specific signatures in cancer: Insights into genomic regulatory networks and implications for medicine. *Genome Med.* **3**, (2011).

34. Martinocorena, I. *et al.* Somatic mutant clones colonize the human esophagus with age. *Science (80-.).* **362**, 911–917 (2018).

35. Yizhak, K. *et al.* Abstract LB-231: Identifying cancer-related processes in normal tissues via RNA-seq. *Cancer Res.* **77**, LB-231-LB-231 (2017).

36. Roufas, C. *et al.* The Expression and Prognostic Impact of Immune Cytolytic Activity-Related Markers in Human Malignancies: A Comprehensive Meta-analysis. *Front. Oncol.* **8**, 1–18 (2018).

37. Dunn, J. & Rao, S. Epigenetics and immunotherapy: The current state of play. *Mol. Immunol.* **87**, 227–239 (2017).

38. Timp, W. & Feinberg, A. P. Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. *Nat. Rev. Cancer* **13**, 497–510 (2013).

39. Peng, D. *et al.* Epigenetic silencing of TH1-type chemokines shapes tumour immunity and immunotherapy. *Nature* **527**, 249–253 (2015).

40. Zaretsky, J. M. *et al.* Mutations Associated with Acquired Resistance to PD-1 Blockade in Melanoma. *N. Engl. J. Med.* **375**, 819–829 (2016).

41. Kuss-Duerkop, S. K., Pyeon, D. & Westrich, J. A. DNA tumor virus regulation of host dna methylation and its implications for immune evasion and oncogenesis. *Viruses* **10**, 1–24 (2018).

42. Duruisseaux, M. *et al.* Epigenetic prediction of response to anti-PD-1 treatment in non-small-cell lung cancer: a multicentre, retrospective analysis. *Lancet Respir. Med.* **6**, 771–781 (2018).

43. Chakravarthy, A. *et al.* Pan-cancer deconvolution of tumour

composition using DNA methylation. *Nat. Commun.* **9**, (2018).

44. Hunt, B. G., Glastad, K. M., Yi, S. V. & Goodisman, M. A. D. The function of intragenic DNA methylation: Insights from insect epigenomes. *Integr. Comp. Biol.* **53**, 319–328 (2013).

45. Shenker, N. & Flanagan, J. M. Intragenic DNA methylation: Implications of this epigenetic mechanism for cancer research. *Br. J. Cancer* **106**, 248–253 (2012).

46. Rauscher, G. H. *et al.* Exploring DNA methylation changes in promoter, intragenic, and intergenic regions as early and late events in breast cancer formation. *BMC Cancer* **15**, 1–15 (2015).

47. McCabe, M. T., Brandes, J. C. & Vertino, P. M. Cancer DNA methylation: Molecular mechanisms and clinical implications. *Clin. Cancer Res.* **15**, 3927–3937 (2009).

48. Network, T. C. G. A. R. *et al.* The Cancer Genome Atlas Pan-cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).

49. Hansen KD. IlluminaHumanMethylation450kanno.ilmn12.hg19: Annotation for Illumina's 450k methylation arrays. (2016).

50. Sean, D. & Meltzer, P. S. GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**, 1846–1847 (2007).

51. Sinha, R. *et al.* Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci. Signal.* **6**, pl1–pl1 (2013).

52. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–4 (2012).

53. Krämer, A., Green, J., Pollard, J. & Tugendreich, S. Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* **30**, 523–530 (2014).

초록

다양한 암종에서 메틸화 변이 정도와 종양 면역원성 사이 상관관계에 관한 연구

서울대학교 대학원

의학과 중개의학

박 창 희

체세포 돌연변이 또는 체세포 염색체 불안정성과 같은 종양에서의 유전자 변이가 종양면역치료에서 신뢰할 수 있는 바이오마커로 보고되어 있지만, 유전자 메틸화의 변화와 종양 면역원성 사이의 연관성은 알려져 있지 않다. 본 연구에서는 종양 면역원성을 나타내는 메틸화와 관련된 바이오마커를 찾고자 한다.

TCGA (The Cancer Genome Atlas)의 약 8000여개에 달하는 전체 암종 데이터베이스를 활용하여, 메틸화 변이 정도를 검체가 가지고 있는 과메틸화 혹은 저 메틸화 된 CpG 부위의 수로 정의하였다. 이렇게 정의된 메틸화 변이 정도와 *GZMA*와 *PRF1* 유전자의 발현의 평균값인 세포용해활성도를 비롯하여 다양한 종양 면역원성과의 연관성을 포괄적으로

조사하였다.

메틸화 변이 정도는 기존의 문헌에 정의되었던 메틸화 분류와 관련성이 있었다. 또한, 종양 면역 인식에 관여하는 분자의 발현과 음의 상관관계가 있는 것도 확인하였다. 전체 암종 분석을 하였을 때, 메틸화 변이 정도는 세포 용해활성도와 음의 상관 관계(ρ = -0.37, p <0.001)가 있음을 확인하였고, 이는 체세포 돌연변이 개수 및 염색체 불안정성과 독립적으로 연관이 있음을 확인하였다. 이러한 음의 상관 관계는 TCGA 이외에 폐선암종 및 저등급 뇌교종으로 각각 구성된 외부 코호트에서도 확인되었다 (ρ = -0.41, p <0.001 및 ρ = -0.34, p = 0.014). 메틸화 변이 정도는 또한 인터페론 감마 유전자 시그니처와 음의 상관관계를 보였으며, 면역원성이 높은 아형에서 더 낮은 것이 확인되었다. 또한, TCGA 흑색종 환자 중 이필리무맙 면역치료를 받은 환자들을 대상으로 보았을 때, 하위 20%의 메틸화 변이를 갖는 환자들에서 더 긴 무질병진행생존을 보였다 (p = 0.029). 그리고 면역 반응과 관련된 유전자 프로모터의 과 메틸화 및 저 메틸화는 세포 용해활성점수에 유의하게 영향을 미쳤다.

이러한 결과는 종양의 메틸화 변이 정도가 종양의 면역 회피에 중요하다는 것을 강조하며, 차후 이와 관련된 바이오마커의 개발이 필요함을 시사한다.

주요어: 종양 메틸화, 종양면역치료

학번: 2018-21621