이학석사 학위논문

유전자 특이적 DNA 메틸화의 가이드로서 microRNA 의 가능성 탐구

Exploring the potential of miRNAs as guider for gene-specific DNA methylation.

울산대학교대학원

의 과 학 과

전 세 연

유전자 특이적 DNA 메틸화의 가이드로서 microRNA 의 가능성 탐구

지도교수 천 성 민, 성 창 옥

이 논문을 이학석사 학위 논문으로 제출함

2023 년 8 월

울산대학교대학원
의 과 학 과
전 세 연

전세연의 이학석사학위 논문을 인준함

심사위원　　김 지 훈　(인)

심사위원　　천 성 민　(인)

심사위원　　김 지 선　(인)

울 산 대 학 교 대 학 원

2023 년 8 월

Abstract

Objective: DNA methylation is known as an important factor in epigenetic regulation of gene expression and has been reported to be closely associated with the development of many types of cancer. Aberrant regulation of gene expression due to changes in DNA methylation patterns of specific genes has frequently been reported in tumor cells. However, a clear explanation of the specific mechanisms by which DNA methylation is induced in certain genes has not yet been established. Therefore, in this study, we targeted microRNAs that can specifically recognize DNA sequences and match various genes to investigate the correlation between microRNA and promoter methylation. We aim to explore statistically significant target microRNAs and genes as potential guides for promoter methylation of specific genes.

Method: A total of 734 microRNAs and the methylation levels of 54,531 gene promoter clusters were analyzed using Spearman's rank correlation in publicly available databases. After selecting target pairs (miRNA-gene cluster) with higher relative microRNA expression and lower false discovery rate (FDR), the dependent relationships of these identified targets were further validated through linear regression analysis.

Result: Through simple linear regression analysis, 25 target genes whose promoter methylation is induced by the expression levels of four target microRNAs (hsa-miR-200a, hsa-miR-200b, hsa-miR-200c and hsa-miR-141) were identified (Positive correlation: $\beta > 0$, $p < 0.05$; Negative correlation: $\beta < 0$, $p < 0.05$), and statistically significant correlations were found between the promoter methylation levels of these genes and their mRNA expression levels regardless of the positive or negative correlation. In the comparison of the cancer group divided into eight categories according to miRNA expression patterns, the correlation between the level of the four miRNAs expression and methylation of each target gene was best explained in cancer group 1, which includes colorectal, gastric, lung, and ovarian cancer. Genes such as *ST14*, a tumor suppressor gene, and *OVOL1* and *EPCAM*, which are associated with tumor invasion and metastasis, were identified as target genes in this study.

Conclusion: Using bioinformatics-based screening analysis, we discovered target gene pairs

that exhibit statistically significant changes in promoter methylation patterns due to specific miRNA expression, utilizing large-scale open data. We were able to confirm the possibility that the expression of genes related to tumor formation, invasion, and metastasis can be regulated by promoter methylation induced by microRNAs. This expands our understanding of the mechanism underlying tumor development through methylation and provides a new perspective on the utilization of microRNAs in the field of cancer treatment.

Keywords: DNA methylation, Linear regression analysis, microRNA, Spearman's rank correlation analysis

Contents

List of Tables and Figures

Introduction

DNA methylation is a phenomenon in which a methyl group is transferred by the DNA methyltransferase (DNMT) enzyme and is considered one of the most important mechanisms for regulating gene expression in epigenetics. It mainly occurs at CpG sites, where cytosine and guanine nucleotides are positioned next to each other and connected by a phosphate bond. Specifically, a methyl group is added to the 5th carbon of the cytosine residue. CpG sites are primarily distributed in regions of the genome called CpG islands, which are densely packed with CpG sites and are generally located at the 5' end of gene regulatory regions. Methylation of CpG islands can lead to the loss of gene expression, so normal cells have intrinsic protective mechanisms to protect against methylation. Therefore, most CpG islands exist in an unmethylated state.[1)] DNA methylation exhibits tissue-specific patterns [1,2)] which are well-maintained and transmitted to the next generation of cells. These patterns are maintained or established by DNMT enzymes, including DNMT3A, DNMT3B, and DNMT1.[3)]

The phenomenon of DNA methylation has been reported to be closely associated with many cancer development processes. In tumor cells, the overall level of 5-methylcytosine is reduced by 5-10% compared to normal cells, but local hypermethylation has been found in many CpG islands.[4,5)] Promoter hypermethylation of *RB* (retinoblastoma), a gene that regulates the cell cycle, occurs in about 10% of sporadic unilateral retinoblastoma, leading to reduced expression of the *RB* gene.[6,7)] In the case of DNA mismatch repair-related *MLH1*, promoter methylation of *MLH1* and reduced expression of *MLH1* were observed in sporadic colon tumors, and treatment with the demethylating agent 5-aza-2'-deoxycytidine was found to decrease *MLH1* suppression.[8,9)] Thus, hypermethylation of tumor suppressor genes in cancer cells can lead to their decreased expression, resulting in the promotion of cancer. Although frequent abnormal regulation of gene expression through specific DNA methylation patterns in tumor cells has been reported,[10)] a clear explanation for the

mechanism by which specific gene-specific DNA methylation occurs has not yet been established.

To specifically regulate DNA methylation of tens of thousands of genes, a regulator is required to precisely induce DNMT to each gene. This regulator should possess the ability to recognize the nucleotide sequence of the target gene and consist of multiple groups capable of matching various genes. Additionally, it should be a substance that has been reported to show differential expression patterns in tumors. From this perspective, we have considered the possibility of microRNAs, which have already been reported in the thousands and are known to bind to the 3' untranslated regions of multiple target mRNAs, thereby inhibiting protein translation of the target genes [11,12], as potential guides for DNMT regulation.

MicroRNAs are small non-coding RNAs composed of 21-25 nucleotides, known to function primarily as post-transcriptional regulators in eukaryotic gene expression.[13-15] Unlike mRNA, which acts as an intermediary in protein expression, microRNAs function based on their ability to recognize and bind to their own nucleotide sequence. The hairpin-shaped pre-miRNA, about 70 nucleotides in length, generated in the nucleus is transported to the cytoplasm and cleaved into a 21-25 nucleotide duplex by Dicer. One strand of the resulting miRNA duplex, along with several proteins such as Dicer1, transactivation-responsive RNA-binding protein (TRBP), and AGO (Argonaute), forms the microRNA-induced silencing complex (miRISC).[16] This complex binds to the target mRNA 3' end with a complementary nucleotide sequence to the microRNA, leading to the degradation of the mRNA or inhibition of its translation.[17]

In addition to its post-transcriptional regulation function, small RNAs have been shown to mediate DNA methylation in plants through a phenomenon known as the RNA-directed DNA methylation (RdDM) pathway.[18] The RdDM pathway has been observed in some yeasts [19] and several studies reporting RdDM-like mechanisms have been reported in human cells.[20-22] Various studies have reported that the sequence recognition ability of microRNAs

is not limited to the mRNA 3' UTR, and that microRNAs function in the nucleus as well as in the cytoplasm.[23-26] Turunen et al. reported that changes occur in the distribution of microRNAs in the nucleus and cytoplasm under hypoxic stress [27], and various studies have reported on other regulatory mechanisms of microRNAs, including transcriptional gene silencing (TGS) and transcriptional gene activating (TGA) mechanisms that contribute to gene silencing or activation within the nucleus, in addition to their post-transcriptional regulatory function in the cytoplasm.[26, 28-29] Kim et al. reported that miR-230, encoded in the antisense direction within the promoter region of the human cell cycle gene *POLR3D*, inhibits the transcription of the host gene.[26] In addition, Younger et al. predicted microRNA binding sites within DNA promoter regions in the nucleus in silico analysis.[30]

So far, research on the interaction between microRNA and DNMT has mainly focused on the regulation of DNMT mRNA expression by microRNA or the methylation of DNA regions encoding microRNA by DNMT, which leads to a decrease in the expression of the microRNA. However, the direct interaction between DNMTs and microRNAs has yet to be elucidated. Additionally, similar processes to RNA-directed DNA methylation (RdDM) have been reported in human cells, but mostly involve self-feedback mechanisms, such as methylation of siRNA-derived repetitive sequences or de novo methylation of CpG islands by promoter antisense-derived small RNAs.[31] However, the molecular mechanism by which endogenous microRNAs in human tumor cells recognize specific target genes outside of microRNA-derived DNA sequences and induce methylation has not been elucidated.

Gene-specific DNA methylation is an important mechanism in carcinogenesis, yet our understanding of this process remains incomplete, and many questions remain unanswered. In this study, we aimed to elucidate how DNA methylation occurs specifically in certain tissues (or genes) and how patterns of methylation change in specific disease states. To investigate the relationship between microRNA and methylation of gene promoter CpG islands, we used public databases to examine various human cancer cell lines. As a result of

our study, we confirmed the possibility that specific microRNAs can recognize DNA and regulate the methylation of gene promoters in a gene-specific manner, expanding the theory of gene-specific methylation and providing a new perspective on the role of microRNAs in cancer development. Through further in-depth research, we expect that the regulation of methylation by microRNAs may contribute to cancer treatment.

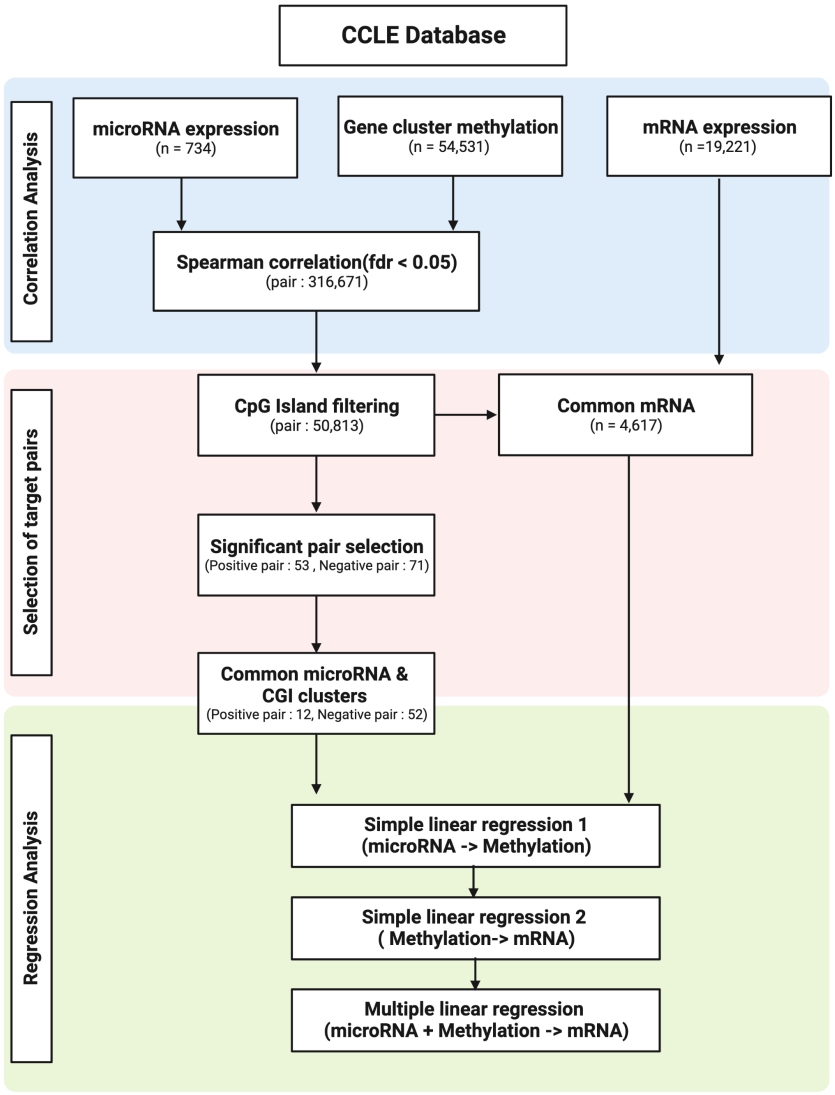

Figure 1. Overview.

An Overview of the Target Screening and Validation Process. CCLE: Cancer Cell Line Encyclopedia. Pair: Correlated pairs of microRNAs and Gene clusters. Common mRNA: Only genes with mRNA expression data were selected from genes in gene clusters estimated to be CpG islands from the results of Spearman correlation analysis.

Materials and Methods

Dataset

This study was conducted based on the data collected from the CCLE (Cancer Cell Lines Encyclopedia) database (https://sites.broadinstitute.org/ccle), which provides large-scale genomic data of cancer cell lines. We obtained expression data for 734 microRNAs and methylation data for 54,531 promoter CpG clusters (corresponding to 20,587 genes). To perform linear regression analysis on the target genes selected through Spearman's rank correlation analysis of microRNA expression and gene cluster methylation, mRNA expression data was obtained from CCLE as additional data. The analysis was conducted on a common set of 813 cell lines that were identified in all three datasets (**Figure 1**).

Statistical analysis

Spearman's rank correlation analysis between the miRNA expression and specific gene cluster methylation levels in each gene was performed for all cases, obtaining correlation coefficients, *p*-values, and the false discovery rate (FDR) for each pair. A correlation with an FDR value below 0.05 was considered significant. For the regression analysis of the final selected target pairs, we sequentially performed simple linear regression and multiple linear regression for the continuous dependent variable. One-way analysis of variance (ANOVA) was conducted to compare differences among continuous variables. Post-hoc tests were performed to verify differences between all groups. A *p*-value less than 0.05 was considered statistically significant in all statistical analyses. All statistical analyses and visualizations were conducted using software version 4.0.3.

Filtering of promoter CpG island clusters

We performed a selection process to identify gene clusters with potential promoter CpG Islands, excluding non-coding RNA (RefSeq id: NM) gene clusters. Our selection criteria were based on a self-generated standard, taking into account the form and characteristics of methylation data. Specifically, we defined CpG Island clusters as those with a minimum of 20 CpG sites within a cluster and a ratio of CpG sites within the entire cluster of 10% or more. Ultimately, we identified 5,070 gene promoter CpG Island clusters through this process.

Cancer type grouping

A total of 26 cancer types were identified in a set of 813 cell lines that shared microRNA expression data sets, gene methylation data sets, and mRNA expression data sets in common. We performed clustering of the entire microRNA expression values using the NMF consensus tool in Gene Pattern (https://www.genepattern.org/). This analysis was conducted under conditions ranging from k=1 to k=12. The microRNA expression patterns of cell lines were clustered based on their similarity, and the k-value was set at the point where the cophenetic coefficient showed a sharp drop and clear clustering. we employed classification into a single group for cancer types only when they demonstrated similar microRNA expression patterns in two or more instances.

Gene ontology analysis

To investigate the interactions and biological characteristics of the selected genes, gene ontology analysis was performed using the Enrichr (https://maayanlab.cloud/Enrichr/) database. Adjusted *p*-values of 0.05 or higher were considered statistically significant in the Biological Process, Molecular Function, and Cellular Component categories for the 25 selected genes.

Sequence homology

To verify the sequence homology between the selected microRNAs and gene clusters, we utilized BLASTn (version: 2.2.31+) from the Basic Local Alignment Search Tool. Since microRNAs are relatively short in size, approximately 22 bp, and we specifically focused on assessing homology within the promoter CpG island of specific genes, we employed the blastn-short program, which is suitable for identifying sequence homology of less than 30 bp. The nucleotide sequence of the target gene cluster was defined as the region between the first CpG site and the last CpG site. To convert these sequences into a database, the Makblastdb package was employed. Subsequently, the homology between the database and query sequence was checked by setting the microRNA as the query sequence. In the known mechanism of microRNA targeting in the mRNA 3' UTR, the seed sequence of 2-8 bp is known to play a crucial role. However, our study explains a different mechanism from the targeting of existing microRNAs. Therefore, we considered the entire mature microRNA sequence, rather than limiting it to the seed sequence. However, we set the word size option, which represents the minimum matching length, to 7bp.

Result

Selection of pairs with confirmed correlations between miRNA and gene promoter CpG islands

Total of 382,152 microRNA-gene cluster pairs were identified with a significant correlation (FDR < 0.05) in the Spearman's rank correlation analysis, among which 50,813 pairs were gene clusters corresponding to the promoter CpG Island (CGI). From these pairs, 929 positive correlated pairs with a correlation coefficient (r) of +0.3 or higher (**Figure 2A**) and 777 negative correlated pairs with a correlation coefficient (r) of less than -0.4 (**Figure 2B**) were selected. To balance the number of positive and negative pairs for analysis, we applied a different correlation coefficient threshold, as there were significantly more negative correlation pairs than positive ones in the overall correlation analysis results. As a result, we performed an additional selection process to find more significant target pairs based on the correlation analysis results. We divided the positive and negative correlation pairs into four intervals each, based on the microRNA expression and correlation FDR values (log2 transformed). We then selected pairs with relatively high mean microRNA expression and low FDR values across all cell lines (**Figure 2A** and **Figure 2B**). Consequently, we identified 53 positive correlated pairs (**Figure 2C**) and 71 negative correlated pairs (**Figure 2D**) within the selected intervals [Positive: log2(miRNA expression mean) > 9.18, log2(FDR) < -76.42; Negative: log2(miRNA expression mean) > 8.48, log2(FDR) < -211.82]. We selected pairs with relatively higher microRNA expression mean for future experimental validation. To select target microRNAs for further investigation, we focused on the four microRNAs (hsa-miR-200a, hsa-miR-200b, hsa-miR-200c, and hsa-miR-141) that were commonly found in both the positively and negatively correlated pairs. A total of 12 gene clusters were positively regulated by these microRNAs based on their expression, while 52 gene clusters were negatively regulated. The final number of target microRNAs and gene clusters are presented in **Table 1**.

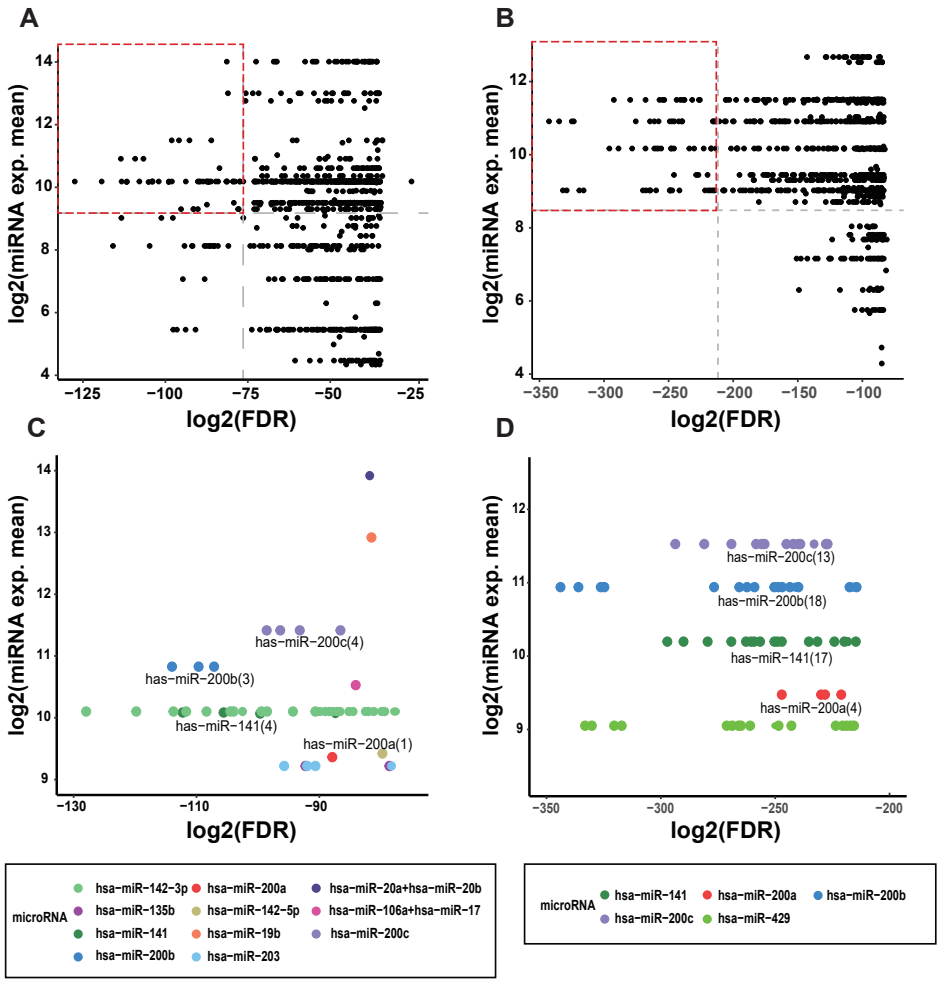

Figure 2. Selection of target microRNA and gene cluster pairs. (A) Scatter plot of positive correlation pairs with Spearman's correlation coefficient (r) +0.3 or higher. The X-axis represents log2(FDR) values and the Y-axis represents the log2-transformed average expression of microRNAs across all cell lines. The red box in the upper left corner indicates a region with relatively high microRNA expression and low FDR values. The middle lines of the plot correspond to x = -76.42 and y = 9.18. (B) Scatter plot of negative correlation pairs with correlation coefficient (r) of -0.4 or less, where the X-axis represents log2(FDR) and the Y-axis represents the log2-transformed average microRNA expression across all cell lines. The red box in the upper-left corner represents the region where the microRNA expression is relatively high, and the FDR value is low. The middle line of the plot is x = -211.82 and y = 8.48. (C) shows an enlarged view of the red box area in the positive correlation plot (A)

[log2(microRNA expression mean) > 9.18 and log2(FDR) < -76.42] (D) shows an enlarged view of the red box area in the negative correlation plot (B). [log2(microRNA expression mean) > 8.48 and log2(FDR) < -211.82] The color of the dots represents the targeted microRNA in each plot.

Table 1. Summary of target microRNAs and gene clusters.

microRNA	Positive correlation		Negative correlation	
	Gene cluster	Gene	Gene cluster	Gene
hsa-miR-200a	1	1	4	4
hsa-miR-200b	3	2	18	17
hsa-miR-200c	4	3	13	12
hsa-miR-141	4	3	17	16

Gene cluster: A gene cluster is a grouping of CpG sites with similar methylation changes around the transcription start site (TSS); Positive correlation: A pair in which the microRNA expression and Gene cluster methylation are positively correlated; Negative correlation: A pair in which the microRNA expression and Gene cluster methylation are negatively correlated.

Regulation of methylation by microRNA expression in target pairs

We conducted three regression analyses to verify whether the target pairs that showed strong correlations actually have a dependent relationship. First, a simple linear regression (first SLR) analysis was performed on the gene promoter methylation influenced by microRNA expression. The results showed significant positive correlations for all 12 target pairs with $\beta > 0$ and p-value < 0.05. Bonferroni correction confirmed statistically significant results for all 12 positive target pairs (**Figure 3A**). Additionally, all 52 pairs with negative correlations showed statistically significant results with an $\beta < 0$ and p-value < 0.05. However, two pairs, has-miR-200b & *DSP*_2 and has-miR-200b & *FAM83H_3*, did not show significant results in Bonferroni correction. Excluding these two pairs, significant results were observed for 50 pairs (**Figure 3B**).

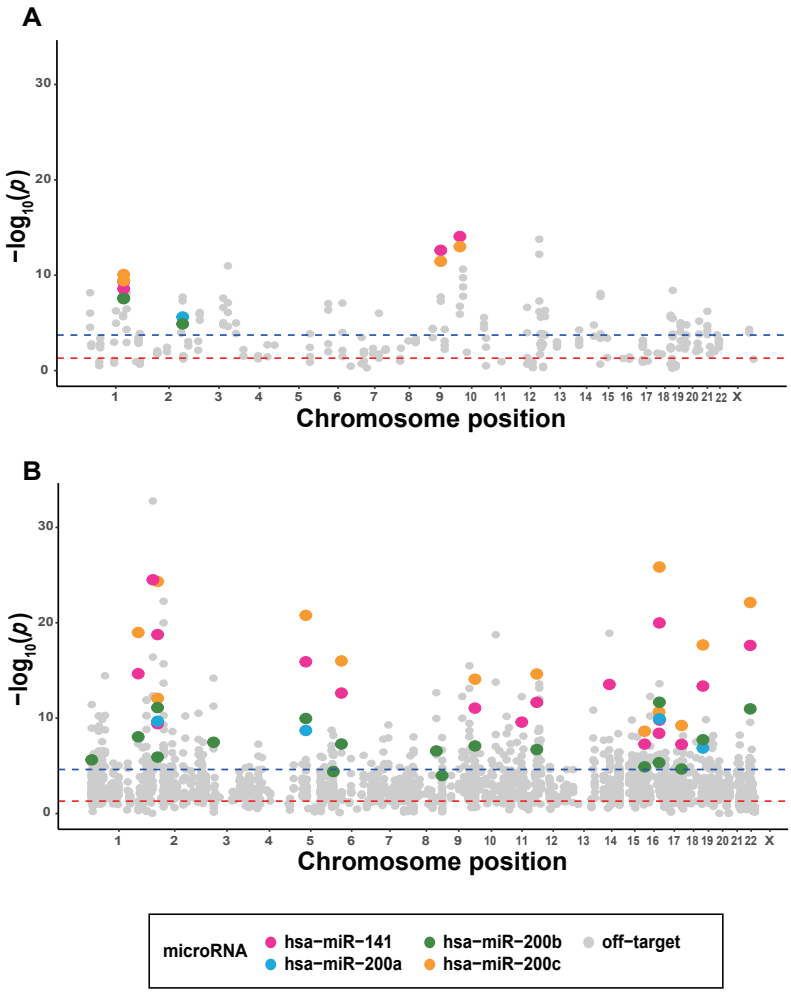

Figure 3. Simple linear regression analysis of gene cluster methylation regulated by target microRNA expression. For Manhattan plots, the x-axis represents target gene positions across the entire genome by chromosome and the y-axis is the negative logarithm *p*-value: -log10(*p*) of each simple linear regression *p*-values. (A) Positive correlation pairs, (B) Negative correlation pairs. The red dashed line indicates the *p*-value of 0.05, while the blue dashed line represents the *p*-value threshold of the Bonferroni correction. Dots in color indicate the target microRNA.

Validation of mRNA expression suppression by promoter CpG island methylation in selected target pairs

To investigate whether the mechanism of mRNA expression regulation by DNA promoter methylation, a well-known theory, can be verified in our data, we performed a second simple linear regression analysis (second SLR) to examine the expression of mRNA regulated by gene promoter methylation. We separated the gene clusters that showed positive and negative correlations with the target microRNAs into positive and negative groups, respectively, and examined mRNA expression for each microRNA (**Figure 4A** and **Figure 4B**). As a result, we found a significant negative correlation between promoter methylation and mRNA expression in both groups. In the SLR2 results for each target pair, excluding hsa-miR-141 & *TJP2_3* and hsa-miR-200c & *TJP2_3* in positive pairs, 10 pairs showed significant results with $\beta < 0$ and *p*-value < 0.05, while all negative pairs showed significant results with $\beta < 0$ and *p*-value < 0.05 (**Supplementary Table 1**).

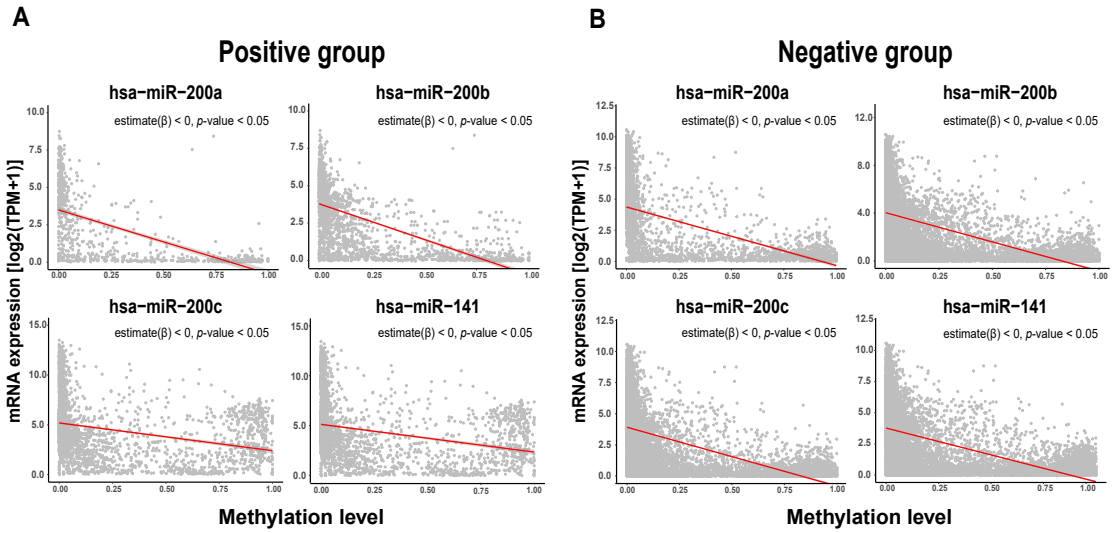

Figure 4. Simple linear regression analysis of mRNA expression regulated by gene promoter methylation. (A) Gene group positively correlated with target microRNA, (B) Gene group negatively correlated with target miRNA. (A) and (B) are the results of simple linear regression analysis for gene cluster methylation and mRNA expression, and $p < 0.05$ was considered statistically significant.

Confirmation of mRNA expression regulation potential by promoter CpG island methylation and microRNA expression in selected target pairs

Finally, to investigate whether microRNA-mediated methylation regulation affects mRNA expression, we performed Multiple linear regression (MLR) analysis. Among the 12 target pairs with positive correlations, hsa-miR-141 & *TJP2*_3 and hsa-miR-200c & *TJP2_3* did not show $\beta < 0$ in both microRNA-mRNA expression and methylation-mRNA expression, contrary to the hypothesis we aimed to test. Excluding these two pairs, $\beta < 0$ was observed in both microRNA-mRNA expression and methylation-mRNA expression for 10 pairs, and 8 pairs were significant with *p*-value < 0.05. For negative correlations, all pairs showed $\beta > 0$ in microRNA-mRNA expression and $\beta < 0$ in methylation-mRNA expression, consistent with our hypothesis, and 51 pairs, except for one pair with *p*-value > 0.05, were statistically significant (**Table 2**).

Table 2. Summary of multiple regression analysis of mRNA expression regulation by target microRNA expression and promoter methylation.

12 Positive correlation pairs

miRNA	Gene	CHR	Region	CpG sites(n)	Cluster No.	miRNA Estimate(β) [95% CI]	miRNA *p*-value	Methylation Estimate(β) [95% CI]	Methylation *p*-value	Hypothesis-matching status
hsa-miR-200c	*VIM*	10	17271051-17271669	35	3	-0.000154[-0.000182,-0.000126]	9.5×10^{-25}	-6.884424[-7.429931,-6.338917]	3.6×10^{-101}	Matched
hsa-miR-141	*VIM*	10	17271051-17271669	35	3	-0.000264[-0.000327,-0.000201]	6.8×10^{-16}	-7.012845[-7.573847,-6.451842]	9.7×10^{-100}	Matched
hsa-miR-200c	*TJP2*	9	71788717-71789619	49	3	0.000039[0.000022,0.000055]	5.7×10^{-06}	0.425321[0.156017,0.694624]	2.0×10^{-03}	Unmatched
hsa-miR-141	*TJP2*	9	71788717-71789619	49	3	0.000071[0.000035,0.000107]	1.2×10^{-04}	0.441101[0.169942,0.712261]	1.5×10^{-03}	Unmatched
hsa-miR-200c	*ATP8B2*	1	154297991-154298672	45	5	-0.000031[-0.000047,-0.000015]	1.4×10^{-04}	-4.691735[-5.120700,-4.262770]	2.9×10^{-81}	Matched
hsa-miR-200c	*ATP8B2*	1	154297991-154298889	49	1	-0.000029[-0.000045,-0.000014]	3.0×10^{-04}	-5.016112[-5.472699,-4.559525]	7.7×10^{-82}	Matched
hsa-miR-200b	*ITGA4*	2	182321830-182322486	40	3	-0.000036[-0.000057,-0.000015]	8.5×10^{-04}	-4.177790[-4.536945,-3.818635]	2.2×10^{-89}	Matched
hsa-miR-141	*ATP8B2*	1	154297991-154298672	45	5	-0.000056[-0.000090,-0.000022]	1.3×10^{-03}	-4.727194[-5.156244,-4.298144]	3.3×10^{-82}	Matched
hsa-miR-141	*ATP8B2*	1	154297991-154298889	49	1	-0.000052[-0.000087,-0.000018]	2.7×10^{-03}	-5.055129[-5.511927,-4.598332]	8.8×10^{-83}	Matched
hsa-miR-200a	*ITGA4*	2	182321830-182322486	40	3	-0.000092[-0.000160,-0.000025]	7.2×10^{-03}	-4.189632[-4.550371,-3.828892]	3.6×10^{-89}	Matched
hsa-miR-200b	*ATP8B2*	1	154297991-154298672	45	5	-0.000007[-0.000024,0.000009]	3.9×10^{-01}	-4.837212[-5.267588,-4.406837]	8.2×10^{-85}	Not significant
hsa-miR-200b	*ATP8B2*	1	154297991-154298889	49	1	-0.000007[-0.000024,0.000010]	4.1×10^{-01}	-5.169473[-5.626440,-4.712507]	1.1×10^{-85}	Not significant

52 Negative correlation pairs

miRNA	Gene	CHR	Region	CpG sites(n)	Cluster No.	miRNA Estimate(β) [95% CI]	miRNA *p*-value	Methylation Estimate(β) [95% CI]	Methylation *p*-value	Hypothesis-matching status
hsa-miR-141	*ENTPD2*	9	139947960-139948871	57	3	0.000181[0.000153,0.000208]	5.9×10^{-35}	-1.637635[-1.929845,-1.345426]	2.6×10^{-26}	Matched
hsa-miR-200c	*EPCAM*	2	47596144-47596799	42	1	0.000183[0.000154,0.000212]	2.7×10^{-32}	-6.709177[-7.415748,-6.002607]	9.6×10^{-65}	Matched
hsa-miR-200c	*ENTPD2*	9	139947960-139948871	57	3	0.000081[0.000068,0.000094]	3.0×10^{-32}	-1.592006[-1.889012,-1.295001]	2.4×10^{-24}	Matched
hsa-miR-200c	*LLGL2*	17	73521001-73522598	103	2	0.000104[0.000086,0.000123]	1.2×10^{-26}	-4.787559[-5.391603,-4.183515]	6.1×10^{-48}	Matched
hsa-miR-141	*LLGL2*	17	73521001-73522598	103	2	0.000222[0.000182,0.000261]	2.1×10^{-26}	-4.882432[-5.483583,-4.281282]	5.8×10^{-50}	Matched
hsa-miR-200c	*CRB3*	19	6464244-6464725	26	2	0.000108[0.000088,0.000128]	2.8×10^{-25}	-3.822919[-4.241026,-3.404812]	8.1×10^{-61}	Matched
hsa-miR-141	*EPCAM*	2	47596144-47596799	42	1	0.000346[0.000282,0.000409]	4.4×10^{-25}	-6.963381[-7.679199,-6.247563]	2.4×10^{-67}	Matched
hsa-miR-141	*OVOL1*	11	65553837-65555703	113	2	0.000157[0.000128,0.000187]	1.7×10^{-24}	-2.507981[-2.850340,-2.165623]	6.4×10^{-42}	Matched

Table 2. (continue)

miRNA	Gene	CHR	Region	CpG sites(n)	Cluster No.	miRNA Estimate(β) [95% CI]	miRNA *p*-value	Methylation Estimate(β) [95% CI]	Methylation *p*-value	Hypothesis-matching status
hsa-miR-200c	*MARVELD2*	5	68710818-68711559	60	1	0.000050[0.000041,0.000060]	7.5×10^{-23}	-2.881747[-3.035745,-2.727749]	1.8×10^{-174}	Matched
hsa-miR-200c	*ST14*	11	130029475-130030104	57	3	0.000129[0.000103,0.000154]	5.1×10^{-22}	-5.611508[-6.143475,-5.079542]	1.0×10^{-76}	Matched
hsa-miR-141	*CRB3*	19	6464244-6464725	26	2	0.000206[0.000163,0.000248]	2.7×10^{-20}	-3.983615[-4.402544,-3.564686]	7.4×10^{-65}	Matched
hsa-miR-200c	*EPCAM*	2	47596829-47597242	42	2	0.000110[0.000087,0.000133]	4.0×10^{-20}	-6.682829[-7.064472,-6.301186]	4.7×10^{-160}	Matched
hsa-miR-200c	*LAD1*	1	201368323-201369008	53	1	0.000107[0.000085,0.000130]	1.1×10^{-19}	-4.466859[-4.901262,-4.032456]	1.1×10^{-73}	Matched
hsa-miR-200c	*FAM83F*	22	40390536-40391052	37	1	0.000054[0.000042,0.000065]	2.3×10^{-19}	-1.664256[-1.871155,-1.457358]	4.2×10^{-49}	Matched
hsa-miR-200b	*ENTPD2*	9	139947960-139948871	57	3	0.000065[0.000051,0.000078]	3.6×10^{-19}	-1.830098[-2.132388,-1.527807]	4.1×10^{-30}	Matched
hsa-miR-141	*MARVELD2*	5	68710818-68711559	60	1	0.000098[0.000077,0.000118]	4.7×10^{-19}	-2.937100[-3.090639,-2.783561]	2.4×10^{-179}	Matched
hsa-miR-141	*TMEM30B*	14	61747174-61748250	73	3	0.000164[0.000128,0.000200]	3.5×10^{-18}	-3.727758[-4.018869,-3.436646]	1.6×10^{-103}	Matched
hsa-miR-200b	*FAM83F*	22	40390536-40391052	37	1	0.000052[0.000040,0.000064]	1.2×10^{-17}	-1.779800[-1.981235,-1.578365]	1.9×10^{-57}	Matched
hsa-miR-200c	*PPL*	16	4986344-4987429	63	3	0.000090[0.000070,0.000110]	1.4×10^{-17}	-4.458818[-4.981708,-3.935928]	3.2×10^{-54}	Matched
hsa-miR-141	*ESRP2*	16	68269687-68270512	44	2	0.000136[0.000105,0.000167]	2.5×10^{-17}	-4.240235[-4.459201,-4.021268]	3.6×10^{-182}	Matched
hsa-miR-200b	*EPCAM*	2	47596144-47596799	42	1	0.000137[0.000105,0.000168]	6.0×10^{-17}	-7.276979[-8.002392,-6.551566]	8.9×10^{-71}	Matched
hsa-miR-141	*EPCAM*	2	47596829-47597242	42	2	0.000206[0.000157,0.000255]	8.1×10^{-16}	-6.835578[-7.215859,-6.455297]	1.5×10^{-165}	Matched
hsa-miR-141	*ST14*	11	130029475-130030104	57	3	0.000230[0.000174,0.000285]	1.3×10^{-15}	-5.804306[-6.341478,-5.267133]	1.1×10^{-79}	Matched
hsa-miR-200c	*ESRP2*	16	68269687-68270512	44	2	0.000061[0.000046,0.000075]	1.6×10^{-15}	-4.213368[-4.437123,-3.989613]	6.3×10^{-176}	Matched
hsa-miR-200b	*LLGL2*	17	73521001-73522598	103	2	0.000078[0.000058,0.000098]	3.4×10^{-14}	-5.160014[-5.777810,-4.542217]	2.3×10^{-52}	Matched
hsa-miR-141	*LAD1*	1	201368323-201369008	53	1	0.000183[0.000134,0.000232]	4.8×10^{-13}	-4.661607[-5.098878,-4.224337]	4.9×10^{-78}	Matched
hsa-miR-141	*FAM83F*	22	40390536-40391052	37	1	0.000091[0.000066,0.000115]	1.2 x 10-^{12}	-1.761512[-1.969698,-1.553327]	1.9×10^{-53}	Matched
hsa-miR-200b	*MARVELD2*	5	68710818-68711559	60	1	0.000037[0.000027,0.000047]	2.3×10^{-12}	-3.015929[-3.169800,-2.862058]	7.9×10^{-185}	Matched
hsa-miR-200c	*C6orf132*	6	42109887-42110586	46	2	0.000063[0.000045,0.000080]	2.4×10^{-12}	-3.800784[-4.155739,-3.445828]	1.5×10^{-78}	Matched
hsa-miR-200b	*EPCAM*	2	47596829-47597242	42	2	0.000085[0.000061,0.000109]	4.5×10^{-12}	-7.009746[-7.385854,-6.633638]	2.3×10^{-173}	Matched

Table 2. (continue)

miRNA	Gene	CHR	Region	CpG sites(n)	Cluster No.	miRNA Estimate(β) [95% CI]	miRNA *p*-value	Methylation Estimate(β) [95% CI]	Methylation *p*-value	Hypothesis-matching status
hsa-miR-200c	*CMTM4*	16	66729833-66731248	93	3	0.000041[0.000029,0.000053]	3.1×10^{-11}	-7.808731[-8.657953,-6.959509]	1.8×10^{-61}	Matched
hsa-miR-200a	*EPCAM*	2	47596829-47597242	42	2	0.000251[0.000176,0.000327]	1.3×10^{-10}	-7.053892[-7.430059,-6.677724]	1.1×10^{-174}	Matched
hsa-miR-141	*C6orf132*	6	42109887-42110586	46	2	0.000122[0.000085,0.000159]	1.6×10^{-10}	-3.873295[-4.226709,-3.519881]	1.7×10^{-81}	Matched
hsa-miR-200b	*ST14*	11	130029475-130030104	57	3	0.000089[0.000062,0.000116]	2.0×10^{-10}	-6.027261[-6.564860,-5.489662]	1.9×10^{-84}	Matched
hsa-miR-200a	*MARVELD2*	5	68710818-68711559	60	1	0.000106[0.000073,0.000138]	4.4×10^{-10}	-3.037455[-3.191768,-2.883142]	8.6×10^{-186}	Matched
hsa-miR-200b	*DSP*	6	7541549-7543453	117	2	0.000079[0.000054,0.000104]	5.2×10^{-10}	-6.120234[-6.745222,-5.495246]	4.0×10^{-68}	Matched
hsa-miR-141	*CMTM4*	16	66729833-66731248	93	3	0.000082[0.000056,0.000108]	7.1×10^{-10}	-7.932765[-8.779915,-7.085616]	2.5×10^{-63}	Matched
hsa-miR-200b	*MAL2*	8	120220032-120221120	63	2	0.000093[0.000063,0.000122]	8.7×10^{-10}	-5.692206[-6.232472,-5.151940]	1.5×10^{-76}	Matched
hsa-miR-141	*PPL*	16	4986344-4987429	63	3	0.000140[0.000095,0.000184]	1.0×10^{-09}	-4.624059[-5.156584,-4.091534]	6.8×10^{-56}	Matched
hsa-miR-200b	*CMTM4*	16	66729833-66731248	93	3	0.000039[0.000027,0.000052]	1.4×10^{-09}	-8.064568[-8.905299,-7.223837]	7.1×10^{-66}	Matched
hsa-miR-200b	*FAM83H*	8	144815479-144816443	85	3	0.000048[0.000031,0.000066]	6.0×10^{-08}	-5.596152[-6.035813,-5.156490]	1.4×10^{-102}	Matched
hsa-miR-200b	*CRB3*	19	6464244-6464725	26	2	0.000057[0.000036,0.000078]	1.5×10^{-07}	-4.286738[-4.714195,-3.859280]	1.1×10^{-70}	Matched
hsa-miR-200b	*LAD1*	1	201368323-201369008	53	1	0.000063[0.000040,0.000087]	2.2×10^{-07}	-4.877211[-5.313298,-4.441123]	3.8×10^{-84}	Matched
hsa-miR-200b	*PPL*	16	4986344-4987429	63	3	0.000057[0.000035,0.000078]	3.9×10^{-07}	-4.730075[-5.262929,-4.197222]	5.6×10^{-58}	Matched
hsa-miR-141	*CDH3*	16	68678698-68680184	89	3	0.000139[0.000083,0.000194]	1.1×10^{-06}	-4.834867[-5.421913,-4.247821]	3.8×10^{-51}	Matched
hsa-miR-200a	*CRB3*	19	6464244-6464725	26	2	0.000168[0.000100,0.000236]	1.3×10^{-06}	-4.317734[-4.745264,-3.890203]	1.7×10^{-71}	Matched
hsa-miR-200b	*PRKCZ*	1	1981414-1982852	102	2	0.000033[0.000018,0.000048]	2.3×10^{-05}	-3.634686[-4.017558,-3.251814]	9.1×10^{-65}	Matched
hsa-miR-141	*KRTCAP3*	2	27665142-27665651	31	3	0.000053[0.000027,0.000080]	7.5×10^{-05}	-4.674550[-4.871200,-4.477901]	1.3×10^{-231}	Matched
hsa-miR-200a	*ESRP2*	16	68269687-68270512	44	2	0.000096[0.000048,0.000144]	1.0×10^{-04}	-4.451007[-4.671486,-4.230527]	1.1×10^{-191}	Matched
hsa-miR-200b	*ESRP2*	16	68269687-68270512	44	2	0.000029[0.000014,0.000044]	2.1×10^{-04}	-4.446705[-4.668444,-4.224967]	3.7×10^{-190}	Matched
hsa-miR-200b	*C6orf132*	6	42109887-42110586	46	2	0.000029[0.000011,0.000047]	1.8×10^{-03}	-4.061693[-4.416678,-3.706708]	3.7×10^{-87}	Matched
hsa-miR-200b	*CDCP1*	3	45186965-45188085	76	1	0.000003[-0.000016,0.000022]	7.7×10^{-01}	-4.894937[-5.255606,-4.534269]	1.0×10^{-112}	Not significant

β, Beta coefficient; CHR, Chromosome; CI, Confidence interval; Cluster No., Cluster number obtained by grouping CpG sites with similar methylation changes around the transcription start site of a gene; CpG sites (n), Number of CpG sites in the region; Hypothesis matching status, indicates whether the β value from the simple regression analysis aligns with the study hypothesis. In cases labeled as "Not significant", it signifies that the *p*-value associated with the hypothesis is not statistically significant; Region, The range between the first and last CpG sites within the cluster.

Classification of cancer types based on target microRNAs expression

We selected a target that could explain all 26 types of cancer and performed NMF clustering and K-means clustering to investigate whether the expression of the target microRNA exhibits different characteristics across cancer types. However, we were unable to identify a range of cophenetic coefficients and clear clustering results from k=2 to k=12 in the NMF clustering (**Figure 5A**). Subsequently, we used the k-means function in R and cut the hierarchical clustering tree into three clusters. As a result, we found that most of the expression was concentrated in cluster 3, and only a small number of cell lines (n = 4) were identified in cluster 1 (**Figure 5B**). Therefore, cancer-type clustering was not observed in the NMF clustering. On the other hand, we confirmed that the expression levels of four target microRNAs were similar in one cell line.

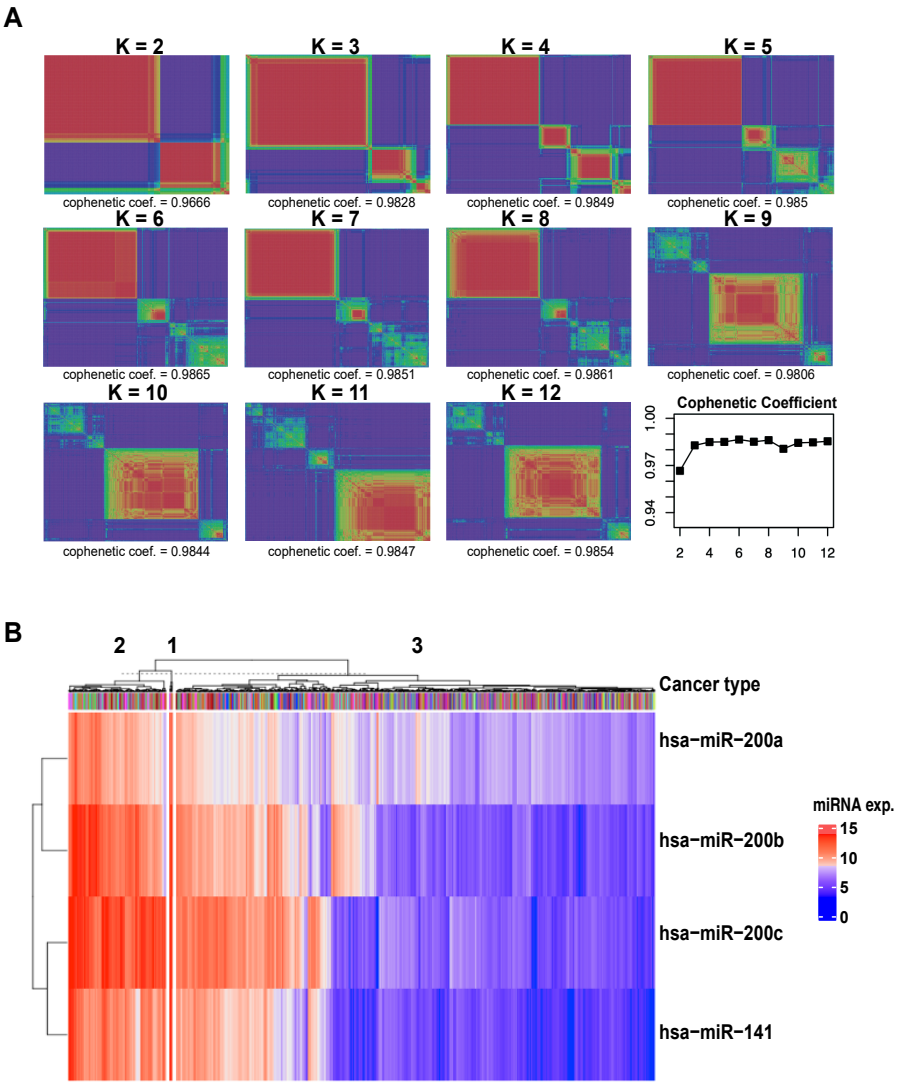

Figure 5. Clustering for target microRNAs expression. (A) NMF consensus clusters of target microRNA expression. Maximum cophenetic coefficients for k = 2 to 12 clusters and the consensus matrices for k = 2 to 12 are shown. The right bottom plot shows a comparison of cophenetic coefficients among k clusters. The cophenetic coefficient plot serves as an indicator of the overall clustering strength across the consensus matrix, with lower values suggesting weaker partitions and higher values indicating stronger partitions. Based on the level of clustering strength observed in the plot, the optimal number of subgroups in the data can be determined by identifying the peak cophenetic coefficient value, which represents the most robust partition. (B) shows the k-means heatmap of target microRNA expression (k=3).

Comparison of the regression analysis results for each cancer group

Although clustering based solely on the selected target microRNA did not result in cancer-type grouping (**Figure 5**), we further performed NMF clustering on the expression of all microRNAs (n=734) in the dataset to explore the significance of target pairs across cancer types. We observed a rebound in the Cophenetic coefficient at K=5 and K=6, but we proceeded with grouping using K=5, which exhibited the most distinct clustering (**Figure 6A**). Among the 26 types of cancer, only those with similar clusters containing two or more types of cancer were grouped together. Consequently, the cancer types were categorized into seven groups. Additionally, to include Fibroblast, which is expected to exhibit patterns similar to normal cells, we included it as the eighth group, resulting in a total of eight groups (**Figure 6B**).

We observed that the expression of target microRNAs was relatively low in groups 3, 6, 7, and 8 compared to other cancer-type groups (p < 2.2e-16 by one-way ANOVA, **Figure 6C**). Subsequently, we performed simple linear regression (SLR1) analysis of target microRNA expression and gene cluster methylation in each cancer type group using the same method as for the overall cancer types. As a result, group 1 (Colon cancer, Gastric cancer, Lung cancer, Ovarian cancer), which included the largest number of cancer types, showed the most similar SLR1 results to those of the overall cancer types, with 7 positive pairs and 45 negative pairs (81.2% of target pairs, **Figure 6D**). On the other hand, SLR1 results of group 2 (Breast cancer, Endometrial/uterine cancer, Pancreatic cancer) and group 3 (Brain cancer, Liver cancer, Skin cancer), which include only three types of cancer with a similar number of cell lines, showed a different trend (**Figure 6D**). Positive target pairs showed significant results in two pairs (2/12) for both group 2 and group 3, while negative target pairs showed significance in 35 pairs (35/52) for group 2 but only one pair (1/52) for group 3. The SLR1 results for target pairs in each cancer type group revealed that hsa-miR-141 & *ATP8B2_1*

and hsa-miR-141 & *ATP8B2_5* were the most frequent positive target pairs, exhibiting significant results in three groups (i.e., group 1, group 3, and group 5). On the other hand, hsa-miR-200b & *ESRP2_2* was the most frequent negative target pair, demonstrating significant results in all four groups (i.e., group 1, group 2, group 3, and group 5).

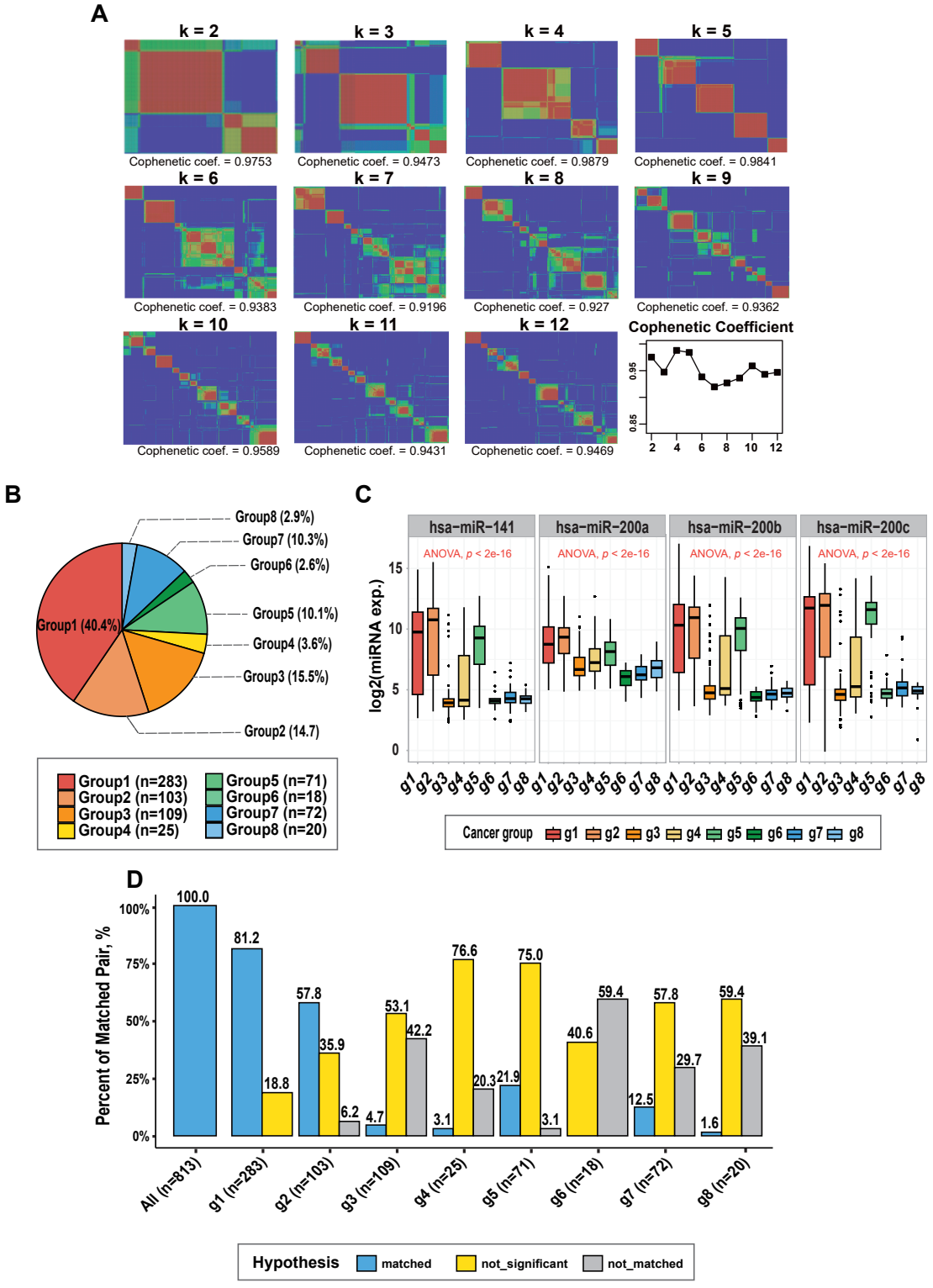

Figure 6. Cancer grouping and comparison of regression analysis result across cancer groups. (A) NMF consensus clusters of total microRNA expression data. Maximum cophenetic coefficients for k = 2 to 12 clusters and the consensus matrices for k = 2 to 12 are

shown. The right bottom plot shows a comparison of cophenetic coefficients among k clusters. (B) Cancer groups, Group 1 includes Colon cancer, Gastric cancer, Lung cancer, and Ovarian cancer (cell line: 283, accounting for 40.4% of the grouped cancer types, 283/701). Group 2 includes Breast cancer, Endometrial cancer, and Pancreatic cancer (cell line: 103, accounting for 14.7% of the grouped cancer types, 103/701). Group 3 includes Brain cancer, Liver cancer, and Skin cancer (cell line: 109, accounting for 15.5% of the grouped cancer types, 109/701). Group 4 includes Bile duct cancer, Sarcoma, and Thyroid cancer (cell line: 25, accounting for 3.6% of the grouped cancer types, 25/701). Group 5 includes Bladder cancer, Esophageal cancer, and Head and Neck cancer (cell line: 71, accounting for 10.1% of the grouped cancer types, 71/701). Group 6 includes Neuroblastoma and Rhabdoid (cell line: 18, accounting for 2.6% of the grouped cancer types, 18/701). Group 7 includes Lymphoma and Myeloma (cell line: 72, accounting for 10.3% of the grouped cancer types, 72/701). Group 8 includes Fibroblast (cell line: 20, accounting for 2.9% of the grouped cancer types, 20/701). Only cancer types with two or more classifications were designated as groups, and as an exception, Fibroblast, which is expected to have a similar pattern to normal cells, was added to Group 8. (C) Boxplots of target microRNA expression across cancer groups (p < 2.2e-16 by one-way ANOVA). (D) Bar plots of SLR1 results across cancer groups. SLR1 refers to simple linear regression analysis of the association between microRNA expression and gene cluster methylation. The blue bars ("matched") in the bar plot of (D) indicate the proportion of statistically significant pairs (p-value < 0.05) with correlation that match the study hypothesis (Positive correlation $\beta > 0$, Negative correlation $\beta < 0$). The yellow bars ("not significant") represent the proportion of pairs with correlation that match the study hypothesis but are not statistically significant (p-value > 0.05). The gray bars ("not matched") indicate the proportion of pairs with correlation that do not match the study hypothesis (Positive correlation $\beta < 0$, Negative correlation $\beta > 0$).

Functional analysis of target genes

We performed Gene Ontology analysis on 25 target genes and identified significant terms with a *p*-value < 0.05 in the Biological Process (BP), Molecular Function (MF), and Cellular Component (CC) categories (**Figure 7**). *DSP, CDH3, MARVELD2*, and *TJP2* were found to be significantly associated with the Cell-cell junction organization (GO:0045216) with the lowest *p*-value (7.03E-04) in biological process (BP). In Molecular function (MF), the Keratin filament binding (GO:1990254) of *FAM83H* and *VIM* showed the most significant association with a *p*-value of 2.24E-05. In Cellular component (CC), the Intermediate filament (GO:0005882) was significantly associated with *FAM83H, DSP, VIM*, and *PPL*, with the lowest *p*-value (4.21E-07). Notably, the genes *MARVELD2, EPCAM*, and *TJP2* showed functional roles in Bicellular tight junction (GO:0005923), Tight junction (GO:0070160), and Apical junction complex (GO:0043296) in the Cellular component (CC).

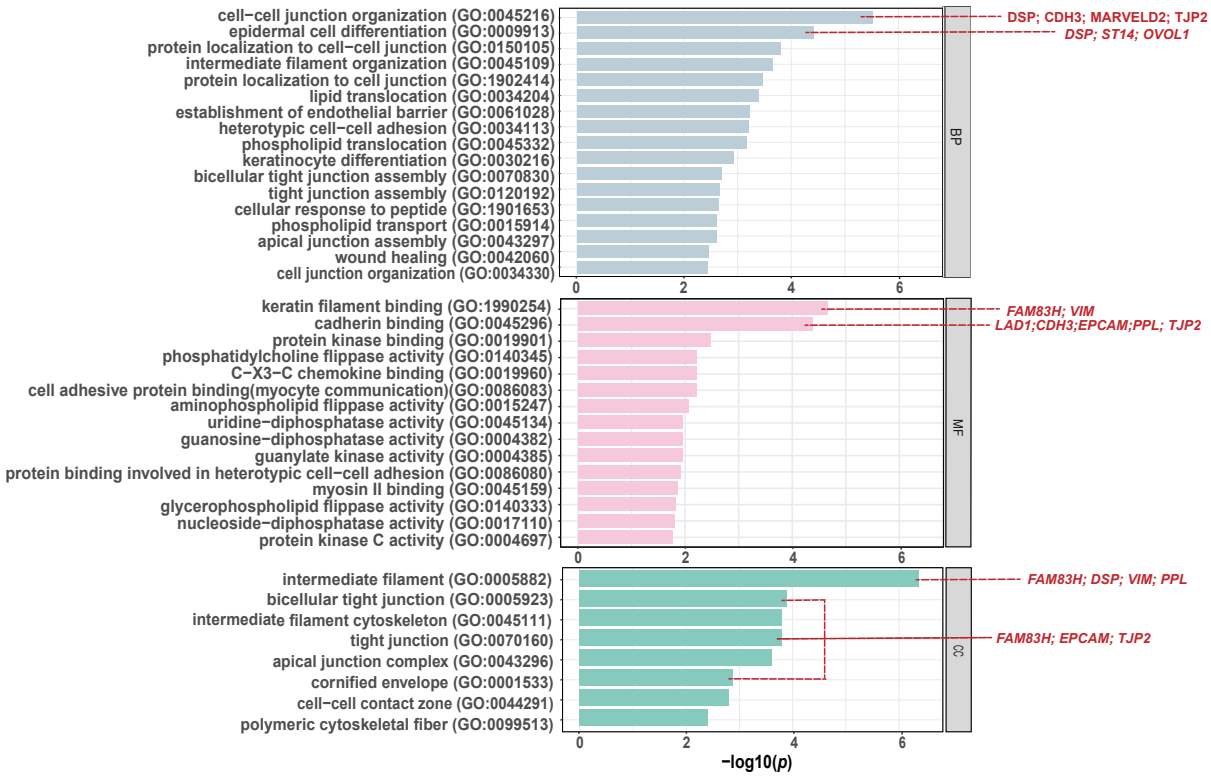

Figure 7. Gene ontology analysis of target genes. A barplot representing the results of the Gene Ontology (GO) analysis. The dark blue bar represents Biological Process (BP), the pink bar represents Molecular Function (MF), and the green bar represents Cellular Component (CC). The y-axis of all bar plots represents the GO term, while the x-axis represents -log10 (adjusted *p*-value). The results of the GO analysis are considered significant when the adjusted *p*-value is less than 0.05.

Discussion

In this study, we investigated the possibility that microRNA regulates gene promoter methylation using data collected from open-source databases. We examined the correlation between 734 microRNAs and various genes using Spearman rank correlation analysis, identifying significant target pairs (**Figure 2**). In addition, we further validated the dependency between the identified pairs through linear regression analysis. This confirmed that the selected targets were indeed regulated by microRNA-mediated methylation of specific genes (**Figure 3**). In the results of the Simple Linear Regression (SLR2) analysis on gene promoter methylation and mRNA expression, we found that, regardless of the directionality and correlation between microRNA and gene clusters, all target pairs showed a significant negative relationship between methylation and mRNA expression, as reported in previous studies. This was further confirmed by our data (**Figure 4** and **Supplementary Table 1**). The results of multiple linear regression showed that, except for some pairs, gene promoter methylation by microRNA could also affect mRNA expression (**Table 2**). However, mRNA expression is a complex mechanism that can be regulated by various factors other than methylation alone, so promoter methylation alone is not sufficient to fully explain the expression.

In the analysis of target microRNA and gene cluster pairs, it was found that the selected targets could best explain the group 1 cancers (Colon cancer, Gastric cancer, Lung cancer, and Ovarian cancer). Group 2 (Breast cancer, Endometrial/Uterine cancer, Pancreatic cancer) and group 3 (Brain cancer, Liver cancer, Skin cancer), which had similar cell numbers, showed significant differences in SLR1 results for target pairs. It is likely that this was due to differences in the expression of target microRNAs across cancer types, particularly between Group 2 and Group 3 (**Figure 6C**). Furthermore, we observed a lower proportion of statistically significant target pairs in the SLR1 results for Groups 3, 4, 6, 7, and 8, which had

overall lower levels of target microRNA expression (**Figure 6D**).

Previous studies have reported on the nuclear functions of microRNAs.[24-26] In this study, we were able to confirm that microRNAs indeed function in the nucleus, specifically at the promoter CpG islands of genes. Promoter methylation of tumor suppressor genes is known to be an important mechanism in tumorigenesis.[32] According to Kim et al., the expression of Suppressor of tumorigenicity 14 (*ST14*) or Serine protease 14 (*Prss14*) genes is associated with poor prognosis and increased invasiveness and metastatic potential in breast cancer patients.[33] *ST14* was identified as a target gene in this study and was found to be negatively regulated by promoter methylation mediated by hsa-miR-200b, hsa-miR-200c, and hs-miR-141 (SLR1). These results suggest that microRNA-mediated regulation of *ST14* gene methylation at the 5' end, and subsequent changes in mRNA expression, may function as a regulatory factor in tumor invasiveness and metastasis. Another target gene, *OVOL1*, is a transcription factor that functions in human cancer cells and induces the reversal process of EMT (M-EMT, mesenchymal to epithelial transition), thereby inhibiting tumor invasion and metastasis, and affecting the growth, apoptosis, and mobility of cancer cells.[34] The target gene *EPCAM* (Epithelial Cell Adhesion Molecule), frequently reported in the Gene Ontology analysis, is an important marker gene for cancer diagnosis and treatment, as it is overexpressed in various types of cancer cells including colon and breast cancer.[35,36] The possibility that the promoter methylation of *EPCAM* can be regulated by microRNA suggests a new perspective in cancer diagnosis and treatment.

The clustering of the selected target microRNAs (hsa-miR-200a, hsa-miR-200b, hsa-miR-200c, hsa-miR-141) did not group the cancer types (**Figure 5A** and **Figure 5B**). Nevertheless, to investigate the significance of target pairs across cancer types, we performed additional NMF clustering on the expression of all 734 microRNAs in the microRNA dataset. This result provides evidence that the microRNAs are not specific to

particular cancer group but rather have a broad impact across all cancer types. In the k-means clustering results, the expression patterns of the four target microRNAs were not distinguished by cancer type, but a similar pattern was observed in one cell line (**Figure 5B**). This suggests that the target microRNAs may have similar or complementary functions because they are all members of the miR-200 family. In fact, it has been revealed that the microRNA 200 family acts as a suppressor in the EMT process, a major mechanism of cancer metastasis.[37)]

The results of the data analysis so far have limitations in proving the hypothesis that microRNA guides methylation. To more clearly confirm the hypothesis, direct evidence of interaction between microRNA and DNMT is needed. To this end, further experimental studies, such as microRNA knockdown using siRNA, microRNA transfection, and ChIP-seq (Chromatin Immunoprecipitation Sequencing), will be conducted for experimental validation. Before conducting experimental validation, we randomly verified the sequence homology between target gene clusters and microRNAs using BLASTn. The results are shown in **Supplementary Table 2** and **Supplementary Table 3**, and we were able to confirm sequence homology in some target pairs.

Conclusion

Through screening using bioinformatics, we efficiently explored a large dataset to identify target pairs with high correlation. Additionally, we validated the expression of mRNA for the target pairs using an mRNA database. Furthermore, our research revealed the possibility that the expression of various genes related to tumorigenesis, invasion, and metastasis could be regulated by promoter methylation via microRNA. This expands our understanding of the mechanism of tumorigenesis through methylation and provides a new perspective on the use of microRNA in cancer treatment.

Reference

1. Bird, Adrian. DNA methylation patterns and epigenetic memory. *Genes & development*. **16**(1), 6-21 (2002).

2. Rakyan, Vardhman K., et al. DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS biology.* **2**(12), e405 (2004).

3. Okano, M., Bell, D. W., Haber, D. A., & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*. **99**(3), 247-257 (1999).

4. Feinberg, A. P., Gehrke, C. W., Kuo, K. C., & Ehrlich, M. Reduced genomic 5-methylcytosine content in human colonic neoplasia. *Cancer research*. **48**(5), 1159-1161 (1988).

5. Gama-Sosa, Miguel A., et al. The 5-methylcytosine content of DNA from human tumors. *Nucleic acids research*. **11**(19), 6883-6894 (1983).

6. Greger V, Debus N, Lohmann D, Hopping W, Passarge E, Horsthemke B, Debus N, Lohmann D, Hopping W, Passarge E, Horsthemke B: Frequency and parental origin of hypermethylated RB1 alleles in retinoblastoma. *Hum Genet*. **94**(5), 491–496 (1994).

7. Ohtani-Fujita, N., et al. "CpG methylation inactivates the promoter activity of the human retinoblastoma tumor-suppressor gene." *Oncogene.* **8**(4), 1036-1067 (1993).

8. Kane, Michael F., et al. "Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines." *Cancer research.* **57**(5), 808-811 (1997).

9. Herman, James G., et al. Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proceedings of the National Academy of Sciences.* **95**(12), 6870-6875 (1998).

10. Schinke, Carolina, et al. Aberrant DNA methylation in malignant melanoma. *Melanoma research.* **20**(4), 253-265 (2010).

11. Bousquet, M., Nguyen, D., Chen, C., Shields, L., & Lodish, H. F. MicroRNA-125b transforms myeloid cell lines by repressing multiple mRNA. *Haematologica*. **97**(11), 1713 (2012).

12. Ritchie, William, John EJ Rasko, and Stéphane Flamant. MicroRNA target prediction and validation. *MicroRNA Cancer Regulation: Advanced Concepts, Bioinformatics and Systems Biology Tools*. 39-53 (2013).

13. Khvorova, A., Reynolds, A., & Jayasena, S. D. Functional siRNAs and miRNAs exhibit strand bias. *Cell*. **115**(2), 209-216 (2003).

14. Chendrimada, Thimmaiah P., et al. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature.* **436**(7051), 740-744 (2005).

15. Ameres, Stefan L., and Phillip D. Zamore. Diversifying microRNA sequence and function. *Nature reviews Molecular cell biology.* **14**(8), 475-488 (2013).

16. Bernstein, E., Caudy, A. A., Hammond, S. M., & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*. **409**(6818), 363-366 (2001).

17. van den Berg, A., Mols, J., & Han, J. RISC-target interaction: cleavage and translational suppression. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*. **1779**(11), 668-677 (2008).

18. Wassenegger, Michael. RNA-directed DNA methylation. *Plant molecular biology.* **43**, 203-220 (2000).

19. Verdel, André, et al. RNAi-mediated targeting of heterochromatin by the RITS complex. *Science*. **303**(5658), 672-676 (2004).

20. Castanotto, Daniela, et al. Short hairpin RNA-directed cytosine (CpG) methylation of the RASSF1A gene promoter in HeLa cells. *Molecular Therapy.* **12**(1). 179-183 (2005).

21. Kawasaki, H., & Taira, K. Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. *Nature*. **431**(7005), 211-217 (2004).

22. Morris, K. V., Chan, S. W. L., Jacobsen, S. E., & Looney, D. J. Small interfering RNA-induced transcriptional gene silencing in human cells. *Science*. **305**(5688), 1289-1292 (2004).

23. Meister, Gunter, et al. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Molecular cell.* **15**(2), 185-197 (2004).

24. Hwang, H. W., Wentzel, E. A., & Mendell, J. T. A hexanucleotide element directs microRNA nuclear import. *Science.* **315**(5808), 97-100 (2007).

25. Gagnon, K. T., Li, L., Janowski, B. A., & Corey, D. R. Analysis of nuclear RNA interference in human cells by subcellular fractionation and Argonaute loading. *Nature protocols*. **9**(9), 2045-2060 (2014).

26. Kim, D. H., Sætrom, P., Snøve Jr, O., & Rossi, J. J. MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proceedings of the National Academy of Sciences*. **105**(42), 16230-16235 (2008).

27. Turunen, Tiia A., et al. Changes in nuclear and cytoplasmic microRNA distribution in response to hypoxic stress. *Scientific reports.* **9**(1), 10332 (2019).

28. Miao, Lingling, et al. A dual inhibition: microRNA-552 suppresses both transcription and translation of cytochrome P450 2E1. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms.* **1859**(4), 650-662 (2016).

29. Place, Robert F., et al. MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proceedings of the National Academy of Sciences.* **105**(5), 1608-1613 (2008).

30. Younger, S. T., Pertsemlidis, A., & Corey, D. R. Predicting potential miRNA target sites within gene promoters. *Bioorganic & medicinal chemistry letters*, **19**(14), 3791-3794 (2009).

31. Tufarelli, Cristina, et al. Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nature genetics.* **34**(2), 157-165 (2003).

32. Battagli, Cristina, et al. Promoter hypermethylation of tumor suppressor genes in urine from kidney cancer patients. *Cancer research.* **63**(24), 8695-8699 (2003).

33. Kim, S., Yang, J. W., Kim, C., & Kim, M. G. Impact of suppression of tumorigenicity 14 (ST14)/serine protease 14 (Prss14) expression analysis on the prognosis and management of estrogen receptor negative breast cancer. *Oncotarget*. **7**(23), 34643 (2016).

34. Roca, Hernan, et al. Transcription factors OVOL1 and OVOL2 induce the mesenchymal to epithelial transition in human cancer. *PloS one.* **8**(10). e76773 (2013).

35. Cimino, Ashley, et al. Epithelial cell adhesion molecule (EpCAM) is overexpressed in breast cancer metastases. *Breast cancer research and treatment.* **123**, 701-708 (2010).

36. Spizzo, Gilbert, et al. EpCAM expression in primary tumour tissues and metastases: an immunohistochemical analysis. *Journal of clinical pathology.* **64**(5), 415-420 (2011).

37. Babaei, Ghader, et al. The emerging role of miR-200 family in metastasis: focus on EMT, CSCs, angiogenesis, and anoikis. *Molecular Biology Reports.* 1-13 (2021).

Supplementary Tables

Supplementary Table 1. Summary of simple regression analysis of mRNA expression regulation by target gene promoter methylation.

12 Positive correlation pairs							
miRNA	**Gene**	**CHR**	**Region**	**CpG sites(n)**	**Cluster No.**	**Estimate(β) [95% CI]**	***p*-value**
hsa-miR-141	*VIM*	10	17271051-17271669	35	3	-7.644[-8.207, -7.082]	6.4×10^{-113}
hsa-miR-200c	*VIM*	10	17271051-17271669	35	3	-7.644 [-8.207, -7.082]	6.4×10^{-113}
hsa-miR-200a	*ITGA4*	2	182321830-182322486	40	3	-4.271 [-4.629, -3.914]	2.8×10^{-93}
hsa-miR-200b	*ITGA4*	2	182321830-182322486	40	3	-4.271 [-4.629, -3.914]	2.8×10^{-93}
hsa-miR-141	*ATP8B2*	1	154297991-154298889	49	1	-5.206 [-5.655, -4.758]	2.7×10^{-89}
hsa-miR-200b	*ATP8B2*	1	154297991-154298889	49	1	-5.206 [-5.655, -4.758]	2.7×10^{-89}
hsa-miR-200c	*ATP8B2*	1	154297991-154298889	49	1	-5.206 [-5.655, -4.758]	2.7×10^{-89}
hsa-miR-200b	*ATP8B2*	1	154297991-154298672	45	5	-4.873 [-5.296, -4.451]	2.0×10^{-88}
hsa-miR-141	*ATP8B2*	1	154297991-154298672	45	5	-4.873 [-5.296, -4.451]	2.0×10^{-88}
hsa-miR-200c	*ATP8B2*	1	154297991-154298672	45	5	-4.873 [-5.296, -4.451]	2.0×10^{-88}
hsa-miR-141	*TJP2*	9	71788717-71789619	49	3	0.577 [0.312, 0.841]	2.1×10^{-05}
hsa-miR-200c	*TJP2*	9	71788717-71789619	49	3	0.577 [0.312, 0.841]	2.1×10^{-05}
52 Negative correlation pairs							
miRNA	**Gene**	**CHR**	**Region**	**CpG sites(n)**	**Cluster No.**	**Estimate(β) [95% CI]**	***p*-value**
hsa-miR-141	*KRTCAP3*	2	27665142-27665651	31	3	-4.815 [-5.001, -4.630]	1.3×10^{-254}
hsa-miR-200b	*ESRP2*	16	68269687-68270512	44	2	-4.549 [-4.766, -4.332]	6.9×10^{-201}
hsa-miR-200c	*ESRP2*	16	68269687-68270512	44	2	-4.549 [-4.766, -4.332]	6.9×10^{-201}
hsa-miR-141	*ESRP2*	16	68269687-68270512	44	2	-4.549 [-4.766, -4.332]	6.9×10^{-201}
hsa-miR-200a	*ESRP2*	16	68269687-68270512	44	2	-4.549 [-4.766, -4.332]	6.9×10^{-201}
hsa-miR-141	*MARVELD2*	5	68710818-68711559	60	1	-3.141 [-3.296, -2.987]	2.4×10^{-193}

Supplementary Table 1. (continue)

miRNA	Gene	CHR	Region	CpG sites(n)	Cluster No.	Estimate(β) [95% CI]	*p*-value
hsa-miR-200c	*MARVELD2*	5	68710818-68711559	60	1	-3.141 [-3.296, -2.987]	2.4×10^{-193}
hsa-miR-200a	*MARVELD2*	5	68710818-68711559	60	1	-3.141 [-3.296, -2.987]	2.4×10^{-193}
hsa-miR-200b	*MARVELD2*	5	68710818-68711559	60	1	-3.141 [-3.296, -2.987]	2.4×10^{-193}
hsa-miR-200b	*EPCAM*	2	47596829-47597242	42	2	-7.330 [-7.706, -6.954]	2.5×10^{-183}
hsa-miR-200c	*EPCAM*	2	47596829-47597242	42	2	-7.330 [-7.706, -6.954]	2.5×10^{-183}
hsa-miR-141	*EPCAM*	2	47596829-47597242	42	2	-7.330 [-7.706, -6.954]	2.5×10^{-183}
hsa-miR-200a	*EPCAM*	2	47596829-47597242	42	2	-7.330 [-7.706, -6.954]	2.5×10^{-183}
hsa-miR-200b	*CDCP1*	3	45186965-45188085	76	1	-4.905 [-5.259, -4.552]	2.4×10^{-116}
hsa-miR-141	*TMEM30B*	14	61747174-61748250	73	3	-4.074 [-4.368, -3.780]	3.4×10^{-116}
hsa-miR-200b	*FAM83H*	8	144815479-144816443	85	3	-5.762 [-6.206, -5.319]	6.9×10^{-106}
hsa-miR-141	*C6orf132*	6	42109887-42110586	46	2	-4.169 [-4.519, -3.819]	1.4×10^{-92}
hsa-miR-200c	*C6orf132*	6	42109887-42110586	46	2	-4.169 [-4.519, -3.819]	1.4×10^{-92}
hsa-miR-200b	*C6orf132*	6	42109887-42110586	46	2	-4.169 [-4.519, -3.819]	1.4×10^{-92}
hsa-miR-200b	*LAD1*	1	201368323-201369008	53	1	-5.110 [-5.544, -4.675]	4.5×10^{-91}
hsa-miR-200c	*LAD1*	1	201368323-201369008	53	1	-5.110 [-5.544, -4.675]	4.5×10^{-91}
hsa-miR-141	*LAD1*	1	201368323-201369008	53	1	-5.110 [-5.544, -4.675]	4.5×10^{-91}
hsa-miR-200c	*ST14*	11	130029475-130030104	57	3	-6.348 [-6.889, -5.806]	1.9×10^{-90}
hsa-miR-200b	*ST14*	11	130029475-130030104	57	3	-6.348 [-6.889, -5.806]	1.9×10^{-90}
hsa-miR-141	*ST14*	11	130029475-130030104	57	3	-6.348 [-6.889, -5.806]	1.9×10^{-90}
hsa-miR-200b	*MAL2*	8	120220032-120221120	63	2	-5.999 [-6.543, -5.455]	2.4×10^{-82}
hsa-miR-141	*CRB3*	19	6464244-6464725	26	2	-4.513 [-4.940, -4.087]	3.6×10^{-77}
hsa-miR-200b	*CRB3*	19	6464244-6464725	26	2	-4.513 [-4.940, -4.087]	3.6×10^{-77}
hsa-miR-200a	*CRB3*	19	6464244-6464725	26	2	-4.513 [-4.940, -4.087]	3.6×10^{-77}

Supplementary Table 1. (continue)

miRNA	Gene	CHR	Region	CpG sites(n)	Cluster No.	Estimate(β) [95% CI]	*p*-value
hsa-miR-200c	*CRB3*	19	6464244-6464725	26	2	-4.513 [-4.940, -4.087]	3.6×10^{-77}
hsa-miR-141	*EPCAM*	2	47596144-47596799	42	1	-7.814 [-8.560, -7.067]	7.9×10^{-76}
hsa-miR-200b	*EPCAM*	2	47596144-47596799	42	1	-7.814 [-8.560, -7.067]	7.9×10^{-76}
hsa-miR-200c	*EPCAM*	2	47596144-47596799	42	1	-7.814 [-8.560, -7.067]	7.9×10^{-76}
hsa-miR-200b	*DSP*	6	7541549-7543453	117	2	-6.408 [-7.041, -5.775]	7.0×10^{-72}
hsa-miR-141	*CMTM4*	16	66729833-66731248	93	3	-8.484 [-9.332, -7.635]	1.7×10^{-70}
hsa-miR-200b	*CMTM4*	16	66729833-66731248	93	3	-8.484 [-9.332, -7.635]	1.7×10^{-70}
hsa-miR-200c	*CMTM4*	16	66729833-66731248	93	3	-8.484 [-9.332, -7.635]	1.7×10^{-70}
hsa-miR-200b	*PRKCZ*	1	1981414-1982852	102	2	-3.772 [-4.153, -3.390]	3.7×10^{-69}
hsa-miR-141	*FAM83F*	22	40390536-40391052	37	1	-1.993 [-2.197, -1.788]	2.5×10^{-67}
hsa-miR-200c	*FAM83F*	22	40390536-40391052	37	1	-1.993 [-2.197, -1.788]	2.5×10^{-67}
hsa-miR-200b	*FAM83F*	22	40390536-40391052	37	1	-1.993 [-2.197, -1.788]	2.5×10^{-67}
hsa-miR-200c	*PPL*	16	4986344-4987429	63	3	-4.942 [-5.477, -4.407]	5.6×10^{-62}
hsa-miR-141	*PPL*	16	4986344-4987429	63	3	-4.942 [-5.477, -4.407]	5.6×10^{-62}
hsa-miR-200b	*PPL*	16	4986344-4987429	63	3	-4.942 [-5.477, -4.407]	5.6×10^{-62}
hsa-miR-141	*CDH3*	16	68678698-68680184	89	3	-5.161 [-5.742, -4.581]	3.9×10^{-58}
hsa-miR-141	*LLGL2*	17	73521001-73522598	103	2	-5.521 [-6.154, -4.888]	2.3×10^{-56}
hsa-miR-200b	*LLGL2*	17	73521001-73522598	103	2	-5.521 [-6.154, -4.888]	2.3×10^{-56}
hsa-miR-200c	*LLGL2*	17	73521001-73522598	103	2	-5.521 [-6.154, -4.888]	2.3×10^{-56}
hsa-miR-141	*OVOL1*	11	65553837-65555703	113	2	-2.912 [-3.268, -2.556]	1.5×10^{-50}
hsa-miR-141	*ENTPD2*	9	139947960-139948871	57	3	-2.096 [-2.407, -1.784]	4.4×10^{-36}
hsa-miR-200c	*ENTPD2*	9	139947960-139948871	57	3	-2.096 [-2.407, -1.784]	4.4×10^{-36}
hsa-miR-200b	*ENTPD2*	9	139947960-139948871	57	3	-2.096 [-2.407, -1.784]	4.4×10^{-36}

β, Beta coefficient; CHR, Chromosome; CI, Confidence interval; Cluster No., Cluster number obtained by grouping CpG sites with similar methylation changes around the transcription start site of a gene; CpG sites (n), Number of CpG sites in the region; Region, The range between the first and last CpG sites within the cluster.

Supplementary Table 2. The number of target gene clusters with confirmed sequence homology.

		hsa-miR-141	hsa-miR-200a	hsa-miR-200b	hsa-miR-200c
Positive gene clusters	**Homology**	1	-	3	3
	Non-homology	3	1	-	1
Negative gene clusters	**Homology**	11	2	10	7
	Non-homology	6	2	8	6

Homology, In case of sequence homology between microRNA and gene promoter cluster; Non-homology, In the absence of sequence homology between microRNA and gene promoter cluster.

Supplementary Table 3. Summary of sequence homology of target pairs.

Positive correlation pairs						
miRNA	**Gene**	**Cluster No.**	**Pident**	**Length**	**Evalue**	**Bitscore**
hsa-miR-141	*TJP2*	3	100	9	13	18.3
hsa-miR-200c	*TJP2*	3	100	8	55	16.4
hsa-miR-141	*TJP2*	3	100	7	200	14.4
hsa-miR-200b	*ATP8B2*	1	100	7	200	14.4
hsa-miR-200b	*ATP8B2*	5	100	7	200	14.4
hsa-miR-200b	*ITGA4*	3	100	7	200	14.4
hsa-miR-200c	*ATP8B2*	1	100	7	218	14.4
hsa-miR-200c	*ATP8B2*	5	100	7	218	14.4
Negative correlation pairs						
miRNA	**Gene**	**Cluster No.**	**Pident**	**Length**	**Evalue**	**Bitscore**
hsa-miR-141	*CMTM4*	3	100	8	30	16.4
hsa-miR-141	*MARVELD2*	1	100	8	30	16.4
hsa-miR-200a	*MARVELD2*	1	100	8	30	16.4
hsa-miR-200b	*C6orf132*	2	100	8	30	16.4
hsa-miR-200b	*CDCP1*	1	100	8	30	16.4
hsa-miR-200b	*MAL2*	2	100	8	30	16.4
hsa-miR-200c	*LAD1*	1	100	8	30	16.4
hsa-miR-200c	*LLGL2*	2	100	8	30	16.4
hsa-miR-141	*CDH3*	3	100	7	119	14.4
hsa-miR-141	*CRB3*	2	100	7	119	14.4
hsa-miR-141	*EPCAM*	2	100	7	119	14.4
hsa-miR-141	*ESRP2*	2	100	7	119	14.4

Supplementary Table 3. (continue)

miRNA	Gene	Cluster No.	Pident	Length	Evalue	Bitscore
hsa-miR-141	*FAM83F*	1	100	7	119	14.4
hsa-miR-141	*LAD1*	1	100	7	119	14.4
hsa-miR-141	*LLGL2*	2	100	7	119	14.4
hsa-miR-141	*OVOL1*	2	100	7	119	14.4
hsa-miR-141	*TMEM30B*	3	100	7	119	14.4
hsa-miR-200a	*ESRP2*	2	100	7	119	14.4
hsa-miR-200b	*CDCP1*	1	100	7	119	14.4
hsa-miR-200b	*CMTM4*	3	100	7	119	14.4
hsa-miR-200b	*DSP*	2	100	7	119	14.4
hsa-miR-200b	*EPCAM*	2	100	7	119	14.4
hsa-miR-200b	*ESRP2*	2	100	7	119	14.4
hsa-miR-200b	*FAM83H*	3	100	7	119	14.4
hsa-miR-200b	*LLGL2*	2	100	7	119	14.4
hsa-miR-200b	*MAL2*	2	100	7	119	14.4
hsa-miR-200b	*PRKCZ*	2	100	7	119	14.4
hsa-miR-200c	*CMTM4*	3	100	7	119	14.4
hsa-miR-200c	*ENTPD2*	3	100	7	119	14.4
hsa-miR-200c	*ESRP2*	2	100	7	119	14.4
hsa-miR-200c	*LLGL2*	2	100	7	119	14.4
hsa-miR-200c	*PPL*	3	100	7	119	14.4
hsa-miR-200c	*ST14*	3	100	7	119	14.4

Evalue, Expect value; Length, Alignment length; Pident, Percentage of identical matches.

국문 요약

연구 배경: DNA 메틸화는 유전자의 발현을 조절하는 후성유전학 현상 중 하나로 다양한 암 발생 과정과 밀접하게 연관되어 있다. 특히 종양세포에서는 특정 유전자의 메틸화 패턴의 변화로 인한 비정상적인 유전자 발현이 빈번하게 보고되고 있지만, 어떠한 기작에 의해 유전자 특이적 메틸화 현상이 유도되는지에 대한 명확한 설명은 아직 이루어지지 않았다. 따라서 본 연구에서는 DNA 서열을 특이적으로 인지하며 다양한 유전자별 매칭이 가능하도록 다수의 군으로 이루어진 microRNA 를 표적으로 하여 microRNA 와 유전자 프로모터 메틸화와 사이의 상관관계를 확인하였다. 이를 통해 microRNA 가 DNA 메틸화의 가이드로 기능할 수 있는 가능성을 확인하고, 통계적으로 유의한 표적 microRNA 와 유전자를 탐색하고자 하였다.

방법: 734 개의 microRNA 와 54,531 개의 유전자 클러스터에 대한 공개 데이터를 이용하여, microRNA 와 유전자 프로모터 메틸화 수준에 대한 스피어만 순위 상관분석을 수행하였다. 이후 상대적으로 낮은 FDR (False Discovery Rate)과 높은 microRNA 발현을 가진 microRNA-유전자 클러스터 쌍(pair)들을 선별한 후, 선형 회귀 분석을 통해 이들 간의 종속 관계를 검증하였다.

결과: 첫 번째로, 선형회귀분석을 통해 4 종의 표적 microRNA (hsa-miR-200a, hsa-miR-200b, hsa-miR-200c, hsa-miR-141)의 발현에 따라 프로모터 메틸화가 조절되는 25 개의 표적 유전자의 유의성을 확인하였다(양의 상관관계의 표적 12 쌍 : $\beta > 0$, $p < 0.05$, 음의 상관관계의 표적 52 쌍 : $\beta < 0$, $p < 0.05$). 또한, 유전자 프로모터 메틸화 수준과 mRNA 발현에 대한 선형 회귀분석 결과에서는 양과 음의 상관관계 구별 없이 모든 표적 쌍에서 기존에 알려진 메틸화로 인한 mRNA 발현 억제 기작과 동일한 결과 확인하였다($\beta < 0$, $p < 0.05$). 마지막으로, microRNA 와 프로모터 메틸화로 인한 mRNA 의 관계를 확인하기 위한 다중 선형 회귀분석을 수행하여 전체 64 개의 표적 쌍 중 59 개의 쌍에서 유의성을 확인하였다(양의 상관관계의 표적 8 쌍 : $\beta > 0$, $p < 0.05$, 음의 상관관계의 표적 51 쌍 : $\beta < 0$, $p < 0.05$). 추가로 microRNA 의 발현을 기준으로 그룹화한 암종 그룹 중 대장암, 위암, 폐암, 난소암을 포함하는 그룹 1 에서 표적 쌍의 관계가 가장 잘 설명되었다(81.2%). 또한 상대적으로 표적 microRNA 의 발현이 낮은 그룹에 비해 높은 그룹에서 표적 쌍의 설명력이 더 높았다.

본 연구에서는 종양 억제 유전자인 *ST14*가 표적 유전자로 확인되었으며, hsa-miR-200b, hsa-miR-200c, hsa-miR-141 과의 프로모터 메틸화 수준 간에 음의 상관관계를 확인하였다 ($\beta < 0$, $p < 0.05$). 또한, 암 전이와 관련된 *OVOL1* 및 암에서 과발현되는 *EPCAM* 유전자도 표적 유전자로 확인되었다.

결론: 본 연구에서는 생물정보학 기반의 스크리닝 과정을 통해 대용량의 데이터에서 통계적으로 유의한 상관관계를 갖는 표적 microRNA 와 유전자 쌍을 선별하였다. 더불어, 종양형성, 침습, 전이와 관련된 다양한 유전자들의 발현이 microRNA 에 의한 프로모터 메틸화를 통해 조절될 수 있는 가능성을 확인하였다.

핵심어 : DNA 메틸화; microRNA; 상관분석; 선형회귀분석; 종양억제유전자