Molecular Therapy
Nucleic Acids
Original Article

Prediction of the Prognosis Based on Chromosomal Instability-Related DNA Methylation Patterns of *ELOVL2* and *UBAC2* in PTCs

Jun Han,[1,6,7] Meijun Chen,[2,6] Qingxiao Fang,[3,6] Yanqing Zhang,[4] Yihan Wang,[5] Jamaspishvili Esma,[2] and Hong Qiao[2]

[1]Department of Endocrinology and Metabolism, The Second Affiliated Hospital, Harbin Medical University, The Fourth Affiliated Hospital, Harbin Medical University, Harbin 150001, China; [2]Department of Endocrinology and Metabolism, The Second Affiliated Hospital, Harbin Medical University, Harbin 150001, China; [3]Surgical Oncology, The Second Affiliated Hospital, Harbin Medical University, Harbin 150001, China; [4]Hematological Department, The Second Affiliated Hospital, Harbin Medical University, Harbin 150001, China; [5]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150001, China

Papillary thyroid carcinoma (PTC) is the most common malignant tumor of endocrine systems. Chromosomal instability (CIN) is crucial to the clinical prognoses of tumor patients. DNA methylation plays an important role in the regulation of gene expression and CIN. Based on PTC samples from The Cancer Genome Atlas database, we used multiple regression analyses to identify methylation patterns of CpG sites with the strongest correlation with gene expression. A total of 4,997 genes were obtained through combining the CpG sites, which were represented as featured DNA methylation patterns. In order to identify CIN-related epigenetic markers of PTC survival, we developed a method to characterize CIN based on DNA methylation patterns of genes using the Student's *t* statistics. We found that 1,239 genes were highly associated with CIN. With the use of the log-rank test, univariate Cox regression analyses, and the Kaplan-Meier method, DNA methylation patterns of *UBAC2* and *ELOVL2*, highly correlated with CIN, provided potential prognostic values for PTC. The higher these two genes, risk scores were correlated with worse PTC patient prognoses. Moreover, the *ELOVL2* risk score was significantly different in the four stages of PTC, suggesting that it was related to the progress of PTC. The DNA methylation pattern associated with CIN may therefore be a good predictor of PTC survival.

INTRODUCTION

Papillary thyroid carcinoma (PTC) is derived from the thyroid follicular epithelium. PTC is the most common type of endocrine cancer, and its incidence has increased rapidly over the past several decades.[1] It accounts for 85% of thyroid cancer, 60% of adult thyroid cancer,[2] and 100% of child thyroid cancer.[3] The vast majority of patients are diagnosed with differentiated thyroid carcinoma, especially with PTC.[4] This causes difficulty in planning the therapy, because some patients are overtreated, whereas in other patients, the same therapy does not result in the eradication of the neoplastic foci and inhibition of the natural course of the disease. PTCs are usually curable with a 5-year survival of over 95%;[5] however, occasionally, they dedifferentiate into more aggressive and lethal thyroid cancers.[6] For this reason, it is important to identify effective prognostic markers to evaluate the prognoses of PTC patients.

Commonly used prognostic markers presently include proteins, microRNAs (miRNAs), mRNAs, and DNA methylations. Ma and Yu[7] suggested that TBL1XR1 overexpression was an unfavorable prognostic factor for epithelial ovarian cancer, and Wang et al.[8] suggested that DHX32 overexpression was an unfavorable prognostic biomarker for breast cancer. The signature of chromosomal instability (CIN), inferred from gene-expression levels, can predict clinical outcomes in multiple human cancers.[9] CIN describes a dynamic state in which cells continuously gain or lose whole chromosomes or parts of chromosomes at an elevated rate and is therefore a principal mediator of aneuploidy and intra-tumor heterogeneity.[10–13] Because aneuploidy is a consequence of CIN, genes with expression levels are consistently associated with aneuploidy, so gene-expression signatures provide a means to estimate levels of CIN.[9] Carter et al.[9] developed a computational method to characterize CIN based on gene-expression levels using the Student's *t* statistics. They mapped the genes to chromosomal sub-bands, with CIN describing the net deviation in expression of genes contained in each chromosomal region relative to the remainder of the sampled transcriptome. Patients with a higher CIN score had worse clinical prognoses. They suggested that gene-expression signatures that had high correlations with CIN could therefore predict the clinical prognoses of tumor patients.

Genomic DNA hypomethylation is another important factor associated with CIN.[14–17] Methylation of the carbon-5 position of cytosine, mostly in the context of CpG dinucleotides, is the main epigenetic

Received 14 February 2019; accepted 17 September 2019;
https://doi.org/10.1016/j.omtn.2019.09.027.

[6]These authors contributed equally to this work.

[7]Present address: Department of Endocrinology and Metabolism, The Fourth Affiliated Hospital, Harbin Medical University, Harbin 150001, China

Correspondence: Hong Qiao, Department of Endocrinology and Metabolism, The Second Affiliated Hospital, Harbin Medical University, Harbin 150001, China.
E-mail: qiaoh0823@sina.com

Table 1. Clinical Information on PTC Patients

Characteristics	Training Set (n = 49)	Testing Set (n = 442)
State		
Living	45	431
Dead	4	11
Survival (years)		
Mean ± SD	4.22 ± 3.52	2.46 ± 2.37
Gender		
Male	14	117
Female	35	325
Age (years)		
Mean ± SD	45.29 ± 17.22	47.57 ± 15.72
Stage		
I	30	246
II	5	47
III	11	99
IV	3	50
Histological Type		
Thyroid papillary carcinoma		
Classical/usual	42	313
Tall cell (≥50% tall cell features)	3	33
Follicular (≥99% follicular patterned)	4	96

modification of DNA and is essential for a properly functioning genome, including maintenance of chromosome stability and transcriptional repression.[18–21] Recently, DNA methylation biomarkers for the diagnoses, molecular typing, and prognoses of cancers were identified.[22–24] Lu et al.[22] suggested that hypermethylation of *hMLH1* in PTC was significantly correlated with age, size, and the number of primary lesions, local invasion, T stage, and lymph node metastases. Shou et al.[23] reported that aberrant methylation of the *RASSF1A* promoter was more frequently detected in thyroid cancer than in noncancerous controls. Wang et al.[24] reported that hypermethylation of *RUNX3* significantly increased the risk of PTC recurrence by using appropriate site-specific cut-off values.

Genomic DNA hypomethylation has been associated with increased CIN, which plays a central role in tumorigenesis.[14–17] Kawano et al.[14] suggested that whole genome hypomethylation initiated carcinogenesis of esophageal squamous cells through CIN. Nishida et al.[15] concluded that DNA hypomethylation is an important cause of CIN in the earliest phase of human hepatocellular carcinoma, especially in the background of noncirrhotic livers. Rodriguez et al.[16] reported that CIN was correlated with genome-wide DNA demethylation in human primary colorectal cancers, and Suzuki et al.[17] reported that global DNA demethylation in gastrointestinal cancer was correlated with increased genomic damage. However, few reports have shown that hypermethylation is associated with CIN.

Gene-expression levels can be affected by a number of factors, including the environment, gene mutations, and DNA methylations.[25–28] Hypermethylated promoters lead to an "off" state of expression, whereas less methylation may lead to an "on" state.[19] Methylation is an acquired epigenetic phenomenon but can be faithfully reproduced in the progeny of affected cells, and the methylation will then be propagated during clonal selection during the development of tumors.[29] DNA methylations are therefore more stable than gene expressions. Although several methylation biomarkers have been identified to predict cancer survival, they are usually limited to average methylation levels of several genes based on experimental data. However, there is a weak correlation between the average DNA methylation levels of gene promoters and the levels of gene expression.[30] This report prompted us to speculate that methylated CpGs might not have equivalent regulatory effects on gene expression, which results in the maximum regulatory effect of DNA methylation on gene expression.[31] We then identified the DNA methylation patterns that had high correlations with CIN as prognostic markers of PTC.

In the following study, based on The Cancer Genome Atlas (TCGA) database, we identified differentially methylated CpG sites between PTC and normal samples. Multiple regression analyses were then used to obtain the methylation patterns of CpGs with the highest correlations with gene expression. We obtained specified genes by combining CpG sites, which were represented as specific DNA methylation patterns. In order to identify CIN-related epigenetic markers of PTC survival, a method was developed to characterize CIN based on the DNA methylation patterns of genes using the Student's *t* statistics. Pearson's correlation coefficient (PCC) was used to evaluate the correlations between DNA methylation patterns and the CIN of each gene. With the use of PCC and a permutation test, we verified that the featured genes were highly associated with CIN. With the use of the log-rank test, univariate Cox regression analyses, and the Kaplan-Meier method, we conducted prognostic analyses. The DNA methylation patterns of *UBAC2* and *ELOVL2* that had high correlations with CIN provided good prognostic values for PTC. Moreover, *UBAC2* and *ELOVL2* were hypomethylation phenotypes. The DNA methylation patterns associated with CIN may therefore be a good predictor of PTC survival.

RESULTS

Identifying Differentially Methylated CpGs Associated with Gene Expression

With the use of the Illumina Infinium HumanMethylation450 BeadChip assay (Illumina, San Diego, CA), raw data (level 3 data), raw UNC RNAseqV2 level 3 expression data, and the clinical prognostic information for PTC were collected from TCGA. The DNA methylation data and the gene-expression data both contained 562 samples, comprised of 49 matched normal samples, 494 PTC samples, 11 metastatic thyroid carcinoma samples, and eight samples of other types of thyroid cancers. We eliminated batch effects between these 562 samples. In total, 49 PTC samples and 49 matched normal samples comprised the training set, with the remaining 445 PTC samples used as the testing set. Three samples were excluded because they did not

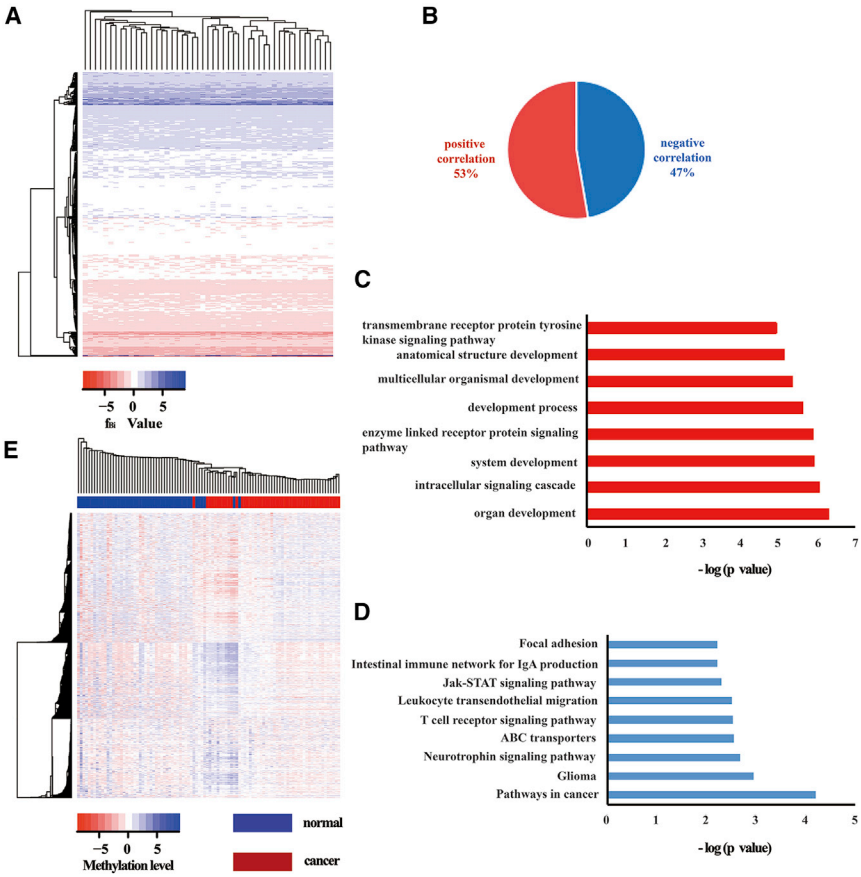

Figure 1. Analyses of Papillary Thyroid Carcinoma Data

(A) Two-way hierarchical clustering of the f_{Bi} value in papillary thyroid carcinoma tissue samples. The blue area was >0, and the red area was <0. (B) In total, DNA methylation patterns of 1,239 genes were highly related to chromosomal instability (CIN), and 653 (53% of the total) were positively correlated genes. In total, 586 (47% of the total) were negatively correlated genes. (C) Gene ontology (GO) functional enrichment analyses for DNA methylation patterns of 1,239 genes were highly related to CIN using DAVID. (D) The Kyoto Encyclopedia of Genes and Genomes pathway enrichment analyses for DNA methylation patterns of 1,239 genes were highly related to CIN using DAVID. (E) Two-way hierarchical clustering of DNA methylation patterns of 1,239 genes were highly related to CIN in papillary thyroid carcinoma tissue samples and normal tissue samples. The blue area was >0, and the red area was <0.

contain survival information. The final study included 442 PTC patients in the testing set (Table 1). Eleven metastatic thyroid carcinoma samples and eight samples of other types of thyroid cancers were excluded.

All of the CpG sites were from the Illumina Infinium HumanMethylation450 BeadChip assay. Raw data (level 3 data) contained all CpG sites in the gene sequence and all CpG sites in the promoter of the analyzed gene. For a specific CpG site, we calculated the correlations with expression of the nearest gene. In total, 203,015 differentially methylated CpG sites were identified from the training set, and 7,541 differentially methylated CpG sites were significantly related to gene expression (false discovery rate [FDR]-corrected p value < 0.05) and included 4,997 genes. There were 3,673 hypermethylated sites and 3,868 hypomethylated sites. More than 50% of the hypomethylated sites were negatively related to gene expression (Figure S1A), and more than 50% of the hypermethylated sites were negatively related to gene expression (Figure S1B). A total of 2,035 genes with the proportion of the hypomethylated sites associated with gene expression were greater than 9.60%, accounting for more than 50%, and the expression of seven genes was influenced by all of their hypomethylated sites (Figure S1C). A total of 2,082 genes with the proportion of the hypermethylated sites associated with gene expression were greater than 9.22%, accounting for more than 50%, and the expression of three genes was influenced by all of their hypermethylated sites (Figure S1D). The hypomethylated sites showed significant regulatory effects on gene expression (Figure S1E), and the hypermethylated sites also showed significant regulatory effects on gene expression (Figure S1F).

Identification of the Genes Related to CIN

The methylation pattern score (score value) was used to describe the maximal regulatory effect of DNA methylation on gene expression. The distribution of the score value was consistent with the normal distribution, which was the same as the gene expression (Figure S2). We therefore characterized the CIN based on the score value using Student's *t* statistics. The f_{Bi} was the net deviation in the score value contained in each chromosomal region relative to the remainder of the sampled score value. The results of clustering analyses showed that most of the bands had CIN (Figure 1A). As a measure of overall CIN, the MFA of a sample was defined as the sum of the magnitudes of its f_{Bi} features. In total, the DNA methylation patterns of 1,239 genes from preselected 4,997 genes were significantly related to the MFA (Table S1), and 53% of them showed a positive correlation, indicating the higher the gene score value, the higher the CIN (Figure 1B). A total of 572 of them were hypermethylation phenotypes (Table S2), and 667 of them were hypomethylation phenotypes (Table S3).

Gene ontology (GO) functional enrichment analyses (Figure 1C) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses (Figure 1D) showed that the 1,239 genes were significantly related to organ development and were enriched in pathways that were related to cancer. The DNA methylation patterns of these genes were significantly related to the MFA and significantly differentiated between normal samples and cancer samples (Figure 1E).

Table 2. Prognosis Markers of PTC Patients

Variable	HR (95% CI)	Regression Coefficient	p Value
TMEM18	1.951 × 10-9(2.351e−17-0.1619)	−20.05	0.0311
UBAC2	7.909 × 10-19(3.658e−33-0.000171)	−41.68	0.0133
ELOVL2	0.002013(5.646e−06-0.7177)	−6.208070	0.0384
ALMS1P	1.7 × 109(1.075-2.687e+18)	21.25	0.0492

Differentially Methylated Markers Associated with PTC Clinical Prognoses

We used the log-rank test to analyze the DNA methylation patterns of 1,239 genes that had high correlations with CIN, with 28 genes being left. The DNA methylation patterns of thirteen genes that had high correlations with CIN were significant ($p < 0.05$) using univariate Cox regression analyses. The hazard ratios (HRs) and 95% confidence intervals (CIs) of the clinical parameters for mortality were calculated using univariate Cox proportional hazard model analyses. The Kaplan-Meier method was used to estimate the overall survival times of patients. The DNA methylation patterns of four genes, *TMEM18*, *UBAC2*, *ELOVL2*, and *ALMS1P*, which had high correlations with CIN, were finally identified as prognosis markers of PTC. We developed four distinct risk scores, each of them based on the methylation pattern of one of four genes. The risk score formula for each patient was calculated as follows: risk score = (−20.05 × *TMEM18*), (−41.68 × *UBAC2*), (−6.20807 × *ELOVL2*), and (21.25 × *ALMS1P*). If the regression coefficient estimated by the univariate Cox proportional hazards model > 0, then the hypermethylation of the risk gene was bad for the survival time. We subdivided the PTC patients into high-risk and low-risk groups by using the median of the risk scores. The HRs (95% CI) of *TMEM18*, *UBAC2*, *ELOVL2*, and *ALMS1P* were 1.951×10^{-9} (2.351e−17-0.1619), 7.909×10^{-19} (3.658e−33-0.000171), 0.002013 (5.646e−06-0.7177), and 1.7×10^{9} (1.075-2.687e+18), respectively (Table 2).

The DNA methylation pattern of *UBAC2* that had a high correlation with CIN significantly predicted the survival of PTC patients in the training set (Figure 2A). The 5-year survival percentage of the high-risk score patients was 68.2% ± 13.6% and was less than that of the low-risk score patients (100%, $p = 0.038$). From the low-risk group to the high-risk group, the methylation levels of the *UBAC2* cg16941122 site showed a significant upward trend and had a strong linear relationship with the risk index, although the expression levels of *UBAC2* and the average methylation levels had no obvious trend (Figure 2B). Furthermore, a higher *UBAC2* risk score was associated with a worse PTC patient prognosis.

The DNA methylation pattern of *TMEM18* that had a high correlation with the CIN significantly predicted the survival of PTC patients in the training set. The 5-year survival percentage of the high-risk score patients was 59.8% ± 16.3% and was significantly lower than that of the low-risk score patients (100%, $p = 0.01$) (Figure S3). The DNA methylation pattern of *ALMS1P* that had a high correlation with CIN significantly predicted the survival of PTC patients in the training set. The 5-year survival percentage of the high-risk score patients was 67.6% ± 14.2% and was significantly lower than that of the low-risk score patients (100%, $p = 0.041$) (Figure S4). Therefore, a higher gene risk score predicted a worse PTC patient prognosis.

The DNA methylation pattern of *ELOVL2* that had a high correlation with CIN significantly predicted the survival of PTC patients in the training set (Figure 3A). The 5-year survival percentage of the high-risk score patients was 65.8% ± 14.6% and was significantly lower than that of the low-risk score group (100%, $p = 0.029$). From the low-risk group to the high-risk group, the methylation levels of the *ELOVL2* cg24724428 site showed a significant upward trend and had a strong linear relationship with the risk index, although the expression levels of *ELOVL2* and the average methylation levels had no obvious trend (Figure 3B). Moreover, the *ELOVL2* risk score was significantly different in the four stages of PTC (Kruskal-Wallis test, $p = 0.001527$), suggesting that the DNA methylation pattern of *ELOVL2* that had a high correlation with CIN was related to the progress of PTC (Figure 3C). A higher *ELOVL2* risk score correlated with a worse PTC patient prognosis, further suggesting that the *ELOVL2* score value that had a high correlation with CIN significantly influenced the patient's clinical condition, progression of the disease, and survival time.

We performed a time-dependent receiver-operating characteristic (ROC) curve analysis to compare the sensitivity and specificity for survival predictions among the DNA methylation patterns of these four genes. The area under the ROC curve (AUC) value was obtained from the ROC analyses and was compared among the DNA methylation patterns of these four genes. The AUC values of *TMEM18*, *UBAC2*, *ELOVL2*, and *ALMS1P* were 0.95, 0.886, 0.764, and 0.854, respectively (Figure 4).

Verification of the Testing Set

The testing set was used to evaluate the reproducibility and availability of these four genes in a prognostic model. The DNA methylation patterns of *ELOVL2* and *UBAC2* were obtained using independent cancer samples that not only associated with CIN but could also be used to predict the prognosis of PTC. Moreover, *UBAC2* and *ELOVL2* were hypomethylation phenotypes. The DNA methylation pattern of *UBAC2* that had a high correlation with CIN predicted the survival of PTC patients in the testing set (Figure 5A). The 5-year survival percentage of the high-risk score patients was 83.6% ± 6.2%, which was significantly lower than that of the low-risk score group (97.1% ± 2%, $p = 0.024$). From the low-risk group to the high-risk group, the methylation levels of the *UBAC2* cg16941122 site showed a significant upward trend and had a strong linear relationship with the risk index, although the expression levels of *UBAC2* and the average methylation levels had no obvious trend (Figure 5B). The samples from the testing set were subgrouped based on the tumor stage, and the survival times of patients from the high-risk score

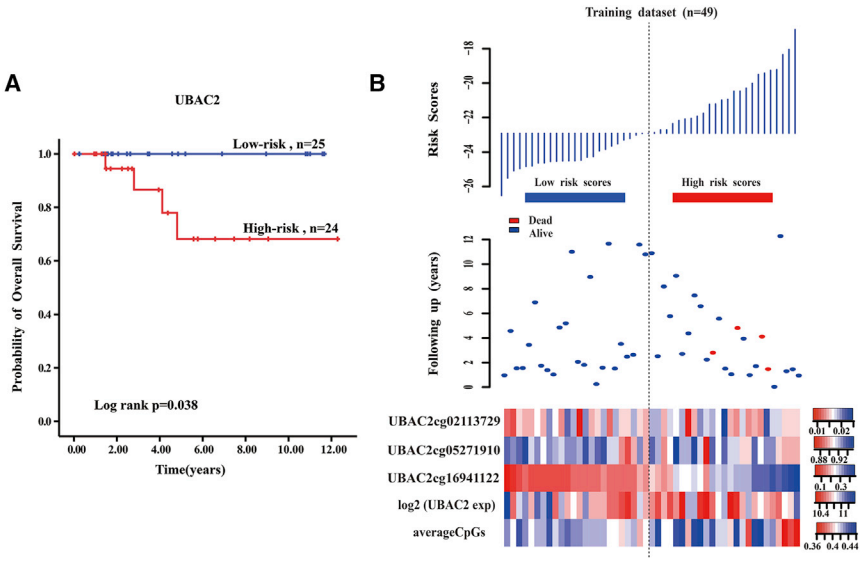

Figure 2. The *UBAC2* Risk Score Model Predicts Overall Survival of Papillary Thyroid Carcinoma Patients in the Training Dataset

(A) Kaplan-Meier analyses for overall survival of patients with high-risk or low-risk scores. The p value was calculated using the two-sided log-rank test. (B) DNA methylation pattern, expression levels, and average methylation levels of the prognostic *UBAC2* that correlated with patients' survival status and increased risk scores.

group were significantly different from that of the low-risk score group in stage III of PTC (p = 0.018) (Figure 5C). The survival times of patients from the high-risk score group were the same as those from the low-risk score group in stages I, II, and IV of PTC (p = 0.39, 0.355, 0.137, respectively). We subgrouped the testing samples based on histological type, and the survival times of the patients in the high-risk score group were significantly different from those of the low-risk score group in the thyroid papillary carcinoma classical/usual group (p = 0.009) (Figure 5D). The thyroid papillary carcinoma tall cell (≥50% tall cell features) histological group contained 33 samples, and there were no samples from patients who had died. Therefore, we could not perform survival analyses. The survival times of the patients in the high-risk score group were the same as those from the low-risk score group in the thyroid papillary carcinoma-follicular (≥99% follicular patterned) histological group (p = 0.366). A higher *UBAC2* risk score predicted a worse PTC patient prognosis.

The DNA methylation pattern of *ELOVL2* that had a high correlation with the CIN significantly predicted the survival of PTC patients in the testing set (Figure 6A). The 5-year survival percentage of the high-risk score patients was 83.4% ± 6.1% and was significantly lower than that of the low-risk score group (97.1% ± 2.4%; p = 0.026). From the low-risk group to the high-risk group, the methylation levels of the *ELOVL2* cg24724428 site showed a significant upward trend and had a strong linear relationship with the risk index, although the expression levels of *ELOVL2* and the average methylation levels had no obvious trend (Figure 6B). The samples from the testing set were subgrouped based on the tumor stage, and the survival times of patients from the high-risk score group were the same as those of the low-risk score group in stages I, II, III, and IV of PTC (p = 0.324, 0.206, 0.19, and 0.05, respectively). We subgrouped the testing samples based on histological type, and the survival times of the patients in the high-risk score group were the same as those from the low-risk score group in the thyroid papillary carcinoma classical/usual group (p = 0.078). The survival times of the patients in the high-risk score group were the same as those of the low-risk score group in the thyroid papillary carcinoma-follicular (≥99% follicular pattern) histological group (p = 0.317). Moreover, the *ELOVL2* risk score was significantly different in the four stages of PTC (Kruskal-Wallis test; $p = 9.915 \times 10^{-13}$) (Figure 6C), suggesting that the DNA methylation pattern of *ELOVL2* that had a high correlation with CIN was related to the progress of PTC. A higher *ELOVL2* risk score was correlated with a worse PTC patient prognosis, indicating that the *ELOVL2* score value that had a high correlation with CIN significantly influenced the patient's clinical condition, progression of disease, and survival time. The AUC values of *ELOVL2* and *UBAC2* were 0.849 and 0.556, respectively (Figure 7).

DISCUSSION

To predict PTC patient clinical prognoses, tumor node metastasis (TNM) staging; patient age, histologic grade of the tumor, tumor extent (extrathyroidal invasion or distant metastases), and size of the primary tumor (AGES) scoring; patient age, presence of distant metastases, extent and size of the primary tumor (AMES) scoring; and metastasis, patient age, completeness of resection, local invasion, and tumor size (MACIS) scoring have been used.[32–35] However, patients with similar clinical phenotypes do not have identical prognoses, suggesting that the present PTC prognostic evaluation system does not provide an accurate clinical prognosis for every patient.[36–39] The PTC prognostic evaluation system therefore needs improvement. The accuracy of PTC prognoses could be significantly improved by the use of molecular markers. Cancer patients with a higher CIN have a worse clinical prognosis, so CIN could be used to evaluate the clinical prognoses of tumor patients.[40–43] To improve the existing PTC prognostic evaluation system, it is important for PTC patient treatment to identify reliable CIN-related prognostic markers. Although CIN-related prognostic markers have been previously reported,[9] the results differed. Therefore, more valid CIN-related prognostic markers are needed to improve the accuracy and credibility of the prognoses.

DNA methylation plays an important role in the regulation of gene expression and CIN. Based on PTC samples from TCGA database,

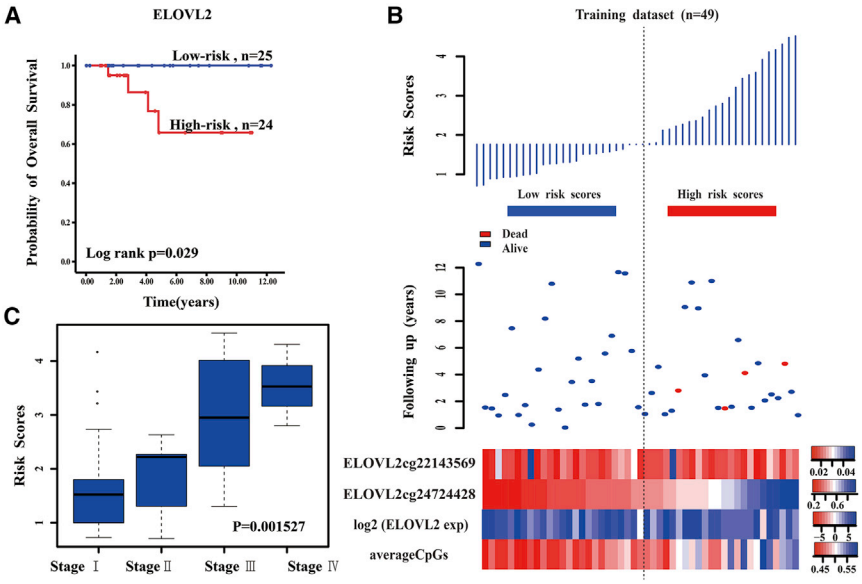

Figure 3. The *ELOVL2* Risk Score Model Predicts Overall Survival of Papillary Thyroid Carcinoma Patients in the Training Dataset

(A) Kaplan-Meier analyses for overall survival of patients with high-risk or low-risk scores. The p value was calculated using the two-sided log-rank test. (B) DNA methylation pattern, expression levels, and average methylation levels of the prognostic *ELOVL2* that correlated with patients' survival status and increased risk scores. (C) Box plot of the *ELOVL2* risk scores in the four stages of papillary thyroid carcinoma.

we used the R Significance Analysis of Microarrays (SAM) package[44] to identify 203,015 differentially methylated CpG sites between PTC and normal samples. Then, we used multiple regression analyses to obtain 7,541 methylation patterns of CpG sites with the strongest correlation with gene expressions. A total of 4,997 genes were obtained by combining the CpG sites, which were represented as featured DNA methylation patterns. The results showed that the distribution of DNA methylation patterns was consistent with the normal distribution, which was the same as gene expression. We subsequently developed a method to characterize CIN based on DNA methylation patterns of genes using the Student's *t* statistics. PCC was used to evaluate the correlation between the DNA methylation patterns and the CIN of each gene. We found that 1,239 genes were highly associated with CIN. With the use of the log-rank test, univariate Cox regression analyses, and the Kaplan-Meier method, DNA methylation patterns of four genes, including *TMEM18*, *UBAC2*, *ELOVL2*, and *ALMS1P*, which had high correlations with CIN, provided good prognostic values for PTC. An independent test set was used to test the validity of the methylation risk score of the four genes. Finally, *ELOVL2* and *UBAC2* remained. In addition, the DNA methylation pattern of *ELOVL2* was involved in different stages of PTC, indicating that the DNA methylation pattern of *ELOVL2* with a high correlation with CIN significantly influenced the patient's clinical condition, progression of disease, and survival time. The DNA methylation pattern associated with CIN may therefore be a good predictor of PTC survival.

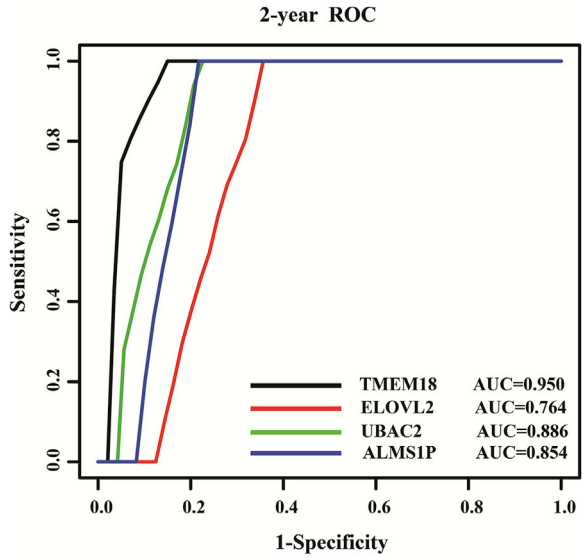

Figure 4. Receiver-Operated Characteristic (ROC) Analyses of the Sensitivity and Specificity for Survival Prediction among the DNA Methylation Patterns of the Four Genes

The time-dependent ROC curve was used to evaluate the prognostic performance for survival predictions. The performance comparison was assessed among the four genes by calculating the area under the ROC curves (AUC) in the training dataset.

Previous studies of *ELOVL2* and *UBAC2* emphasized their relationships with lipid metabolism and obesity.[45–48] González-Bengtsson et al.[45] suggested that ELOVL2 played an important role in docosahexaenoic acid (DHA) synthesis. Kobayashi et al.[46] reported that cells overexpressing ELOVL2 showed enhanced triacylglycerol synthesis and subsequent accumulation of lipid droplets. Pauter et al.[47] suggested that hepatic DHA synthesis of ELOVL2, in addition to controlling *de novo* lipogenesis, also regulated lipid storage and fat mass expansion in an SREBP-1c-independent fashion. Tikhonenko et al.[48] reported that a decrease in long-chain polyunsaturated fatty acids was associated with a decrease in the fatty acid elongases, ELOVL2 and ELOVL4, in diabetes, and additional studies showed that obesity increased the risk of thyroid cancer.[49–52] Han et al.[49] reported that the morbidity of thyroid cancer in female patients was related to a high BMI. Hwang et al.[50] suggested that weight gain and annual increases in obesity indicators in middle-aged adults increased the risk of developing PTC. Kim et al.[51] reported that a higher BMI was associated with more

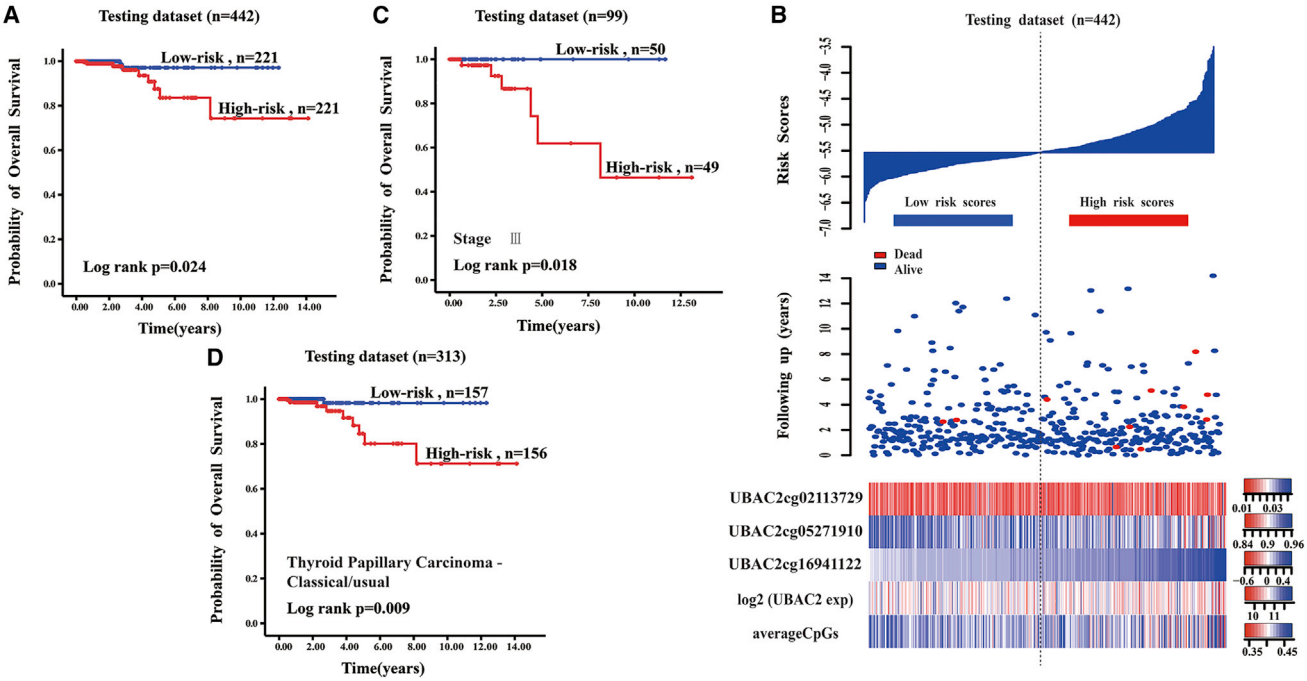

Figure 5. The *UBAC2* Risk Score Model Predicts Overall Survival of Papillary Thyroid Carcinoma Patients in the Testing Dataset

(A) Kaplan-Meier analyses for overall survival of patients with high-risk or low-risk scores. The p value was calculated using the two-sided log-rank test. (B) DNA methylation pattern, expression level, and average methylation level of the prognostic *UBAC2* that correlated with patients' survival status and increased risk scores. (C) Kaplan-Meier analyses for overall survival of patients in the third stage of papillary thyroid carcinoma. The p value was calculated using the two-sided log-rank test. (D) Kaplan-Meier analyses for overall survival of patients with thyroid papillary carcinoma-classical/usual histological types. The p value was calculated using the two-sided log-rank test.

aggressive tumor features, such as lymph node metastasis, lymphatic invasion, and tumor multiplicity in PTC patients. Oberman et al.[52] reported that obesity was significantly associated with thyroid cancer, with BMI, in particular, a strong predictor of thyroid cancer. *ELOVL2* was also significantly related to biosynthesis of unsaturated fatty acids, and *UBAC2* was significantly related to protein localization to the endoplasmic reticulum. Therefore, DNA methylation patterns of *ELOVL2* and *UBAC2* were not only associated with CIN, but might also participate in the initiation and development of PTC.

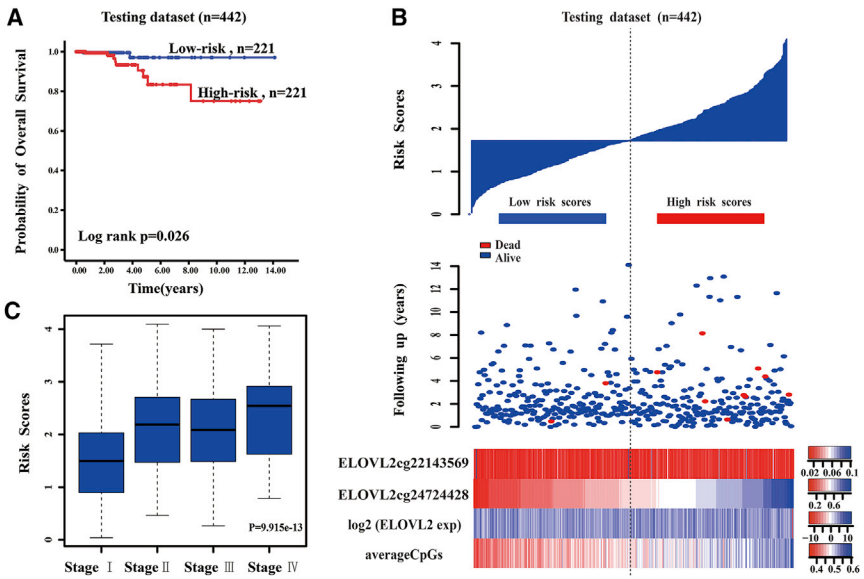

Figure 6. The *ELOVL2* Risk Score Model Predicts Overall Survival of Papillary Thyroid Carcinoma Patients in the Testing Dataset

(A) Kaplan-Meier analyses for overall survival of patients with high-risk or low-risk scores. The P value was calculated using the two-sided log-rank test. (B) DNA methylation pattern, expression level, and average methylation level of the prognostic *ELOVL2* that correlated with the patients' survival status and increased risk scores. (C) Box plot of the *ELOVL2* risk score in the four stages of papillary thyroid carcinoma.

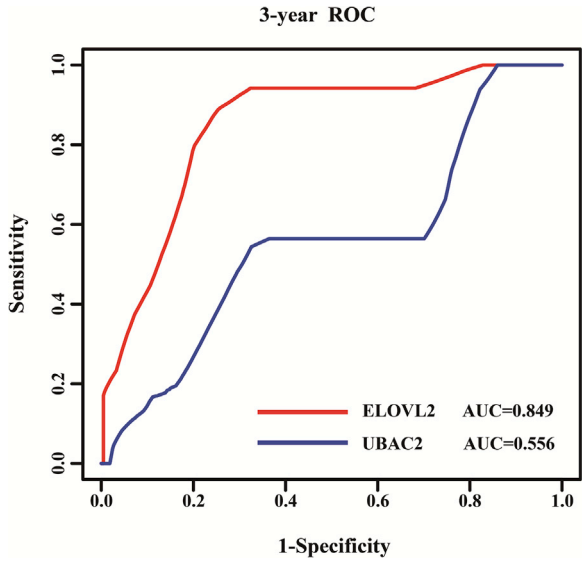

Figure 7. Receiver-Operated Characteristic (ROC) Analyses of the Sensitivity and Specificity for Survival Prediction between the DNA Methylation Patterns of the Two Genes

The time-dependent ROC curve was used to evaluate the prognostic performance for survival predictions. The performance comparison was assessed between the two genes by calculating the area under the ROC curves (AUC) in the testing dataset.

Because of differences among individual patients, presently used prognostic indicators cannot accurately predict the prognosis of each patient. It is therefore difficult to evaluate the clinical prognoses of patients with similar clinical features. Our results showed that the DNA methylation patterns of *ELOVL2* and *UBAC2*, which had high correlations with CIN, could be used to predict the prognosis of PTC. Aberrant DNA methylation was related to the risk of PTC, and the epigenetic markers associated with CIN may be used as predictors of PTC survival. The inclusion of these prognosis markers into the present PTC prognostic evaluation system could therefore assist the clinician in determining the prognoses of patients with similar clinical features and could provide a more appropriate therapeutic schedule for high-risk patients to enhance the efficacies of PTC treatments.

Conclusions

UBAC2 and *ELOVL2*, which had high correlations with CIN, provided good prognostic values for PTC. The DNA methylation pattern associated with CIN may therefore be a good predictor of PTC survival.

MATERIALS AND METHODS

Acquisition of Gene Expression and DNA Methylation Data

With the use of the Illumina Infinium HumanMethylation450 BeadChip assay, raw data (level 3 data), raw UNC RNAseqV2 level 3 expression data, and clinical prognostic information for PTC were collected from TCGA (https://www.cancer.gov/tcga/).

Eliminating Batch Effects

In order to ensure the accuracy of experiments, we used the R Surrogate Variable Analysis (Bioconductor) package to eliminate batch effects of all of the DNA methylation data and the gene-expression data from all samples.

Identifying Differentially Methylated CpGs

To compare the differences of DNA methylation between cancer and normal samples in the training set, we used the R SAM[44] package to identify differentially methylated CpG sites. To control for a FDR of the results, we used the Benjamini-Hochberg method to correct the p value obtained from the statistical test. The threshold for defining the differentially methylated CpG site involved a value of $p < 0.05$, and the differential level (delta beta value) between the cancer and normal samples was >0.1.

Identifying Differentially Methylated CpGs Associated with Gene Expression

For the differentially methylated CpG sites of the PTC samples from the training set, we used multiple regression coefficients to evaluate the correlation between DNA methylation and gene expression. The dependent variable was gene expression of a single gene. The independent variables were all CpG sites mapped to the gene, containing all CpG sites under the gene sequence and all CpG sites in the promoter of the analyzed gene. For a particular CpG site, we calculated correlations with the expression of the nearest gene. A value of $p < 0.05$ was identified as a significant methylation level highly related to gene expression.

Quantitation of the Regulatory Effect of DNA Methylation (Score Value)

With the consideration of the multiple CpGs mapped to the gene and the variability of the DNA methylation levels of multiple CpGs located in the same gene, the average methylation level may not reflect the real ability of DNA methylation to regulate gene expression. Multiple regression analyses were therefore used to quantify the regulatory competence of differential CpG sites and then to quantify the maximal regulatory effect of DNA methylation on gene expression. The methylation pattern score (score value) was defined[31] as follows:

$$Score_k = \widehat{\alpha}_1 \cdot cg_{k1} + \widehat{\alpha}_2 \cdot cg_{k2} + \ldots \widehat{\alpha}_j \cdot cg_{kj}. \quad \text{(Equation 1)}$$

The j CpGs sites represented the significant methylation sites in multiple regression analyses ($p < 0.05$), $\alpha\hat{}j$ ($_{j = 1,2,\ldots,j}$) represented the multiple regression coefficient of the jth CpG sites of the gene, and *cgkj* represented the methylation level of the jth CpG sites of the gene in the kth sample.

Calculating the CIN Score (MFA)

This study downloaded the cytoband coordinate file (GRCh37/hg19) from the UCSC Genome Bioinformatics and then mapped the genes to chromosomal sub-bands. If fewer than five genes were present in a given cytoband, we considered the statistical

measure unreliable, and that cytoband was eliminated from further analysis. The f_{Bi} described the net deviation in score value contained in each chromosomal region relative to the remainder of the sampled score value:[9]

$$f_{Bi} = \frac{\mu_{Bi} - \mu_{Gi}}{\sqrt{(\sigma_{Bi}^2/N_B) + (\sigma_{Gi}^2/N_G)}}. \quad \text{(Equation 2)}$$

The μBi represented the average score value of all of the genes under sample i in the band, σ_{Bi}^2 represented the variance of the score value of all the genes under sample i in the band, μGi represented the average score value of the rest of the genes under sample i, σ_{Gi}^2 represented the variance of the score value of the rest of the genes under sample i, N_B represented the number of genes in the band, and N_G represented the number of the remaining genes.

As a measure of overall CIN, we defined the MFA of a sample as the sum of the magnitudes of its f_{Bi} features:[9]

$$MFA_i = \sum_{Band} \left| f_{Bi} \right|. \quad \text{(Equation 3)}$$

Identification of Genes Related to MFA

PCC was used to evaluate the correlation between the score value and the MFA of each gene. In order to control for a FDR of the PCC, we adopted a permutation test to correct the p value of the statistical test. For each gene, the MFA was permutated 1,000 times to calculate its PCC value, and if the value of p was <0.05, this gene was identified as highly related to MFA as follows:

$$p = \frac{c+1}{1001}, \quad \text{(Equation 4)}$$

where c was the number of PCC square values that were no less than the actual PCC square value of the gene from the 1,000 permutations.

Enrichment Analyses for the GO and KEGG Pathways

To analyze further the biological significance of the genes related to CIN, we used Database for Annotation, Visualization and Integrated Discovery (DAVID) software to perform GO function analyses for the genes related to CIN.[53] The Fisher exact test with multiple test corrections (FDR < 0.05) was used to obtain significant GO terms associated with PTC. We acquired the KEGG pathway terms using the same method.

Prognosis Analyses

The log-rank test was used to obtain p values and to identify a subset of genes for which a score value that had high correlations with CIN showed significant differences between the high and low groups. The high and low groups were groups with high and low score values, grouped by their median values. The survival times were compared between these two groups. Genes with $p < 0.05$ were used in the study. The p values were uncorrected p values. Univariate Cox regression analyses were performed to assess the survival prognosis capabilities of the selected gene set using the overall survival time as a dependent variable. The HRs and 95% CIs of the clinical parameters for mortality were calculated using the univariate Cox proportional hazard model. The risk score formula for each patient was calculated as follows:

$$Risk\ \ Score_k = c_i \bullet gene_{ik}, \quad \text{(Equation 5)}$$

where k was the kth sample, i denoted the feature genes filtered by the univariate Cox proportional hazards models, and c_i was the regression coefficient estimated by the univariate Cox proportional hazards model. The 5-year overall survival for each score value scoring group (high versus low) was calculated using the Kaplan-Meier method, and the statistical significance was assessed using the log-rank test. The significance level of all statistical tests was $p < 0.05$. We performed time-dependent ROC curve analyses to compare the sensitivities and specificities for survival predictions between the predicted genes. The ROC AUC values were obtained from ROC analyses and were compared between the selected genes.

In order to verify the reproducibility and accuracy of the gene prognostic model, as predicted in the training set, we used the testing set. The regression coefficients and the thresholds of risk scores derived from the training set were directly applied to the testing set, and then the patients in the testing set were divided into high-risk and low-risk groups. The evaluation of survival times and the comparison of differences between the two groups were the same as that of the training set.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.omtn.2019.09.027.

AUTHOR CONTRIBUTIONS

H.Q. conceived and designed the experiments. J.H., M.C., and Q.F. acquired the experiment data. J.H., M.C., and Y.Z. performed the study. J.H., M.C., and Y.W. carried out the data analysis. J.H. and J.E. wrote the manuscript. All authors read and approved the final manuscript.

CONFLICTS OF INTEREST

The authors declare no competing interests.

ACKNOWLEDGMENTS

This work was funded by the National Natural Science Foundation of China (grant number 81673108), the Science and Technology Innovation Talent Research Foundation of Harbin (grant number 2016RAXYJ088), and the Nurture Research Fund from The Fourth Affiliated Hospital of Harbin Medical University (grant number HYDSYPY201910).

REFERENCES

1. Bădulescu, I.C., Bărbuş, E., and Piciu, D. (2017). Circulating tumor cells in thyroid carcinoma - the prognostic role of this biomarker. Review of the literature. Clujul Med. *90*, 256–261.
2. Zhang, H., Teng, X., Liu, Z., Zhang, L., and Liu, Z. (2015). Gene expression profile analyze the molecular mechanism of CXCR7 regulating papillary thyroid carcinoma growth and metastasis. J. Exp. Clin. Cancer Res. *34*, 16.
3. Wang, P., Pei, R., Lu, Z., Rao, X., and Liu, B. (2013). Methylation of p16 CpG islands correlated with metastasis and aggressiveness in papillary thyroid carcinoma. J. Chin. Med. Assoc. *76*, 135–139.
4. Lewiński, A., and Adamczewski, Z. (2017). Papillary thyroid carcinoma: a cancer with an extremely diverse genetic background and prognosis. Pol. Arch. Intern. Med. *127*, 388–389.
5. Hay, I.D., Thompson, G.B., Grant, C.S., Bergstralh, E.J., Dvorak, C.E., Gorman, C.A., Maurer, M.S., McIver, B., Mullan, B.P., Oberg, A.L., et al. (2002). Papillary thyroid carcinoma managed at the Mayo Clinic during six decades (1940-1999): temporal trends in initial therapy and long-term outcome in 2444 consecutively treated patients. World J. Surg. *26*, 879–885.
6. Cancer Genome Atlas Research Network (2014). Integrated genomic characterization of papillary thyroid carcinoma. Cell *159*, 676–690.
7. Ma, M., and Yu, N. (2017). Over-Expression of TBL1XR1 Indicates Poor Prognosis of Serous Epithelial Ovarian Cancer. Tohoku J. Exp. Med. *241*, 239–247.
8. Wang, M., Zhang, G., Wang, Y., Ma, R., Zhang, L., Lv, H., Fang, F., and Kang, X. (2017). DHX32 expression is an indicator of poor breast cancer prognosis. Oncol. Lett. *13*, 942–948.
9. Carter, S.L., Eklund, A.C., Kohane, I.S., Harris, L.N., and Szallasi, Z. (2006). A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. Nat. Genet. *38*, 1043–1048.
10. Bastians, H. (2015). Causes of Chromosomal Instability. Recent Results Cancer Res. *200*, 95–113.
11. Carbone, L., and Chavez, S.L. (2015). Mammalian pre-implantation chromosomal instability: species comparison, evolutionary considerations, and pathological correlations. Syst Biol Reprod Med *61*, 321–335.
12. Meena, J., Rudolph, K.L., and Günes, C. (2015). Telomere Dysfunction, Chromosomal Instability and Cancer. Recent Results Cancer Res. *200*, 61–79.
13. Wu, Z.H. (2016). Phenotypes and genotypes of the chromosomal instability syndromes. Transl. Pediatr. *5*, 79–83.
14. Kawano, H., Saeki, H., Kitao, H., Tsuda, Y., Otsu, H., Ando, K., Ito, S., Egashira, A., Oki, E., Morita, M., et al. (2014). Chromosomal instability associated with global DNA hypomethylation is associated with the initiation and progression of esophageal squamous cell carcinoma. Ann. Surg. Oncol. *21* (*Suppl 4*), S696–S702.
15. Nishida, N., Kudo, M., Nishimura, T., Arizumi, T., Takita, M., Kitai, S., Yada, N., Hagiwara, S., Inoue, T., Minami, Y., et al. (2013). Unique association between global DNA hypomethylation and chromosomal alterations in human hepatocellular carcinoma. PLoS ONE *8*, e72312.
16. Rodriguez, J., Frigola, J., Vendrell, E., Risques, R.A., Fraga, M.F., Morales, C., Moreno, V., Esteller, M., Capellà, G., Ribas, M., and Peinado, M.A. (2006). Chromosomal instability correlates with genome-wide DNA demethylation in human primary colorectal cancers. Cancer Res. *66*, 8462–9468.
17. Suzuki, K., Suzuki, I., Leodolter, A., Alonso, S., Horiuchi, S., Yamashita, K., and Perucho, M. (2006). Global DNA demethylation in gastrointestinal cancer is age dependent and precedes genomic damage. Cancer Cell *9*, 199–207.
18. Bird, A. (2007). Perceptions of epigenetics. Nature *447*, 396–398.
19. Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat. Genet. *33* (*Suppl*), 245–254.
20. Laird, P.W. (2010). Principles and challenges of genomewide DNA methylation analysis. Nat. Rev. Genet. *11*, 191–203.
21. Taby, R., and Issa, J.P. (2010). Cancer epigenetics. CA Cancer J. Clin. *60*, 376–392.
22. Lu, X.X., Ge, M.H., Ling, Z.Q., Hu, S.S., Xu, J., Zheng, C.M., Tan, Z., and Chen, C. (2013). [Aberrant methylation of hMLH1 gene promoter in papillary thyroid cancer and its clinical significance]. Zhonghua Zhong Liu Za Zhi *35*, 833–836.
23. Shou, F., Xu, F., Li, G., Zhao, Z., Mao, Y., Yang, F., Wang, H., and Guo, H. (2017). RASSF1A promoter methylation is associated with increased risk of thyroid cancer: a meta-analysis. OncoTargets Ther. *10*, 247–257.
24. Wang, D., Cui, W., Wu, X., Qu, Y., Wang, N., Shi, B., and Hou, P. (2014). RUNX3 site-specific hypermethylation predicts papillary thyroid cancer recurrence. Am. J. Cancer Res. *4*, 725–737.
25. El-Tahan, R.R., Ghoneim, A.M., and El-Mashad, N. (2016). TNF-α gene polymorphisms and expression. Springerplus *5*, 1508.
26. Fujiuchi, N., Matoba, N., and Matsuda, R. (2016). Environment Control to Improve Recombinant Protein Yields in Plants Based on Agrobacterium-Mediated Transient Gene Expression. Front. Bioeng. Biotechnol. *4*, 23.
27. Rivas, H.G., Schmaling, S.K., and Gaglia, M.M. (2016). Shutoff of Host Gene Expression in Influenza A Virus and Herpesviruses: Similar Mechanisms and Common Themes. Viruses *8*, 102.
28. Salminen, A., Kauppinen, A., and Kaarniranta, K. (2016). AMPK/Snf1 signaling regulates histone acetylation: Impact on gene expression and epigenetic functions. Cell. Signal. *28*, 887–895.
29. Judson, H., Stewart, A., Leslie, A., Pratt, N.R., Baty, D.U., Steele, R.J., and Carey, F.A. (2006). Relationship between point gene mutation, chromosomal abnormality, and tumour suppressor gene methylation status in colorectal adenomas. J. Pathol. *210*, 344–350.
30. Schultz, M.D., He, Y., Whitaker, J.W., Hariharan, M., Mukamel, E.A., Leung, D., Rajagopal, N., Nery, J.R., Urich, M.A., Chen, H., et al. (2015). Human body epigenome maps reveal noncanonical DNA methylation variation. Nature *523*, 212–216.
31. Zhang, M., Zhang, S., Wen, Y., Wang, Y., Wei, Y., Liu, H., Zhang, D., Su, J., Wang, F., and Zhang, Y. (2015). DNA Methylation Patterns Can Estimate Nonequivalent Outcomes of Breast Cancer with the Same Receptor Subtypes. PLoS ONE *10*, e0142279.
32. Chruścik, A., and Lam, A.K. (2015). Clinical pathological impacts of microRNAs in papillary thyroid carcinoma: A crucial review. Exp. Mol. Pathol. *99*, 393–398.
33. Gong, W., Yang, S., Yang, X., and Guo, F. (2016). Blood preoperative neutrophil-to-lymphocyte ratio is correlated with TNM stage in patients with papillary thyroid cancer. Clinics (São Paulo) *71*, 311–314.
34. Lee, Y.S., Kim, Y., Jeon, S., Bae, J.S., Jung, S.L., and Jung, C.K. (2015). Cytologic, clinicopathologic, and molecular features of papillary thyroid carcinoma with prominent hobnail features: 10 case reports and systematic literature review. Int. J. Clin. Exp. Pathol. *8*, 7988–7997.
35. Pillai, S., Gopalan, V., Smith, R.A., and Lam, A.K. (2015). Diffuse sclerosing variant of papillary thyroid carcinoma–an update of its clinicopathological features and molecular biology. Crit. Rev. Oncol. Hematol. *94*, 64–73.
36. Ito, Y., Miyauchi, A., Kihara, M., Kobayashi, K., and Miya, A. (2014). Prognostic values of clinical lymph node metastasis and macroscopic extrathyroid extension in papillary thyroid carcinoma. Endocr. J. *61*, 745–750.
37. Ito, Y., Miyauchi, A., Kobayashi, K., Kihara, M., and Miya, A. (2014). Static and dynamic prognostic factors of papillary thyroid carcinoma. Endocr. J. *61*, 1145–1151.
38. Ito, Y., Miyauchi, A., Kobayashi, K., and Miya, A. (2014). Prognosis and growth activity depend on patient age in clinical and subclinical papillary thyroid carcinoma. Endocr. J. *61*, 205–213.
39. Pelizzo, M.R., Merante Boschin, I., Toniato, A., Pagetta, C., Casal Ide, E., Mian, C., and Rubello, D. (2008). Diagnosis, treatment, prognostic factors and long-term outcome in papillary thyroid carcinoma. Minerva Endocrinol. *33*, 359–379.
40. Genga, K.R., Filho, F.D., Ferreira, F.V., de Sousa, J.C., Studart, F.S., Magalhães, S.M., Heredia, F.F., and Pinheiro, R.F. (2015). Proteins of the mitotic checkpoint and spindle are related to chromosomal instability and unfavourable prognosis in patients with myelodysplastic syndrome. J. Clin. Pathol. *68*, 381–387.
41. Giaretti, W., Monteghirfo, S., Pentenero, M., Gandolfo, S., Malacarne, D., and Castagnola, P. (2013). Chromosomal instability, DNA index, dysplasia, and subsite in oral premalignancy as intermediate endpoints of risk of cancer. Cancer Epidemiol. Biomarkers Prev. *22*, 1133–1141.

42. Krem, M.M., Press, O.W., Horwitz, M.S., and Tidwell, T. (2015). Mechanisms and clinical applications of chromosomal instability in lymphoid malignancy. Br. J. Haematol. *171*, 13–28.

43. Taiakina, D., Dal Pra, A., and Bristow, R.G. (2014). Intratumoral hypoxia as the genesis of genetic instability and clinical prognosis in prostate cancer. Adv. Exp. Med. Biol. *772*, 189–204.

44. Li, J., and Tibshirani, R. (2013). Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. Stat. Methods Med. Res. *22*, 519–536.

45. González-Bengtsson, A., Asadi, A., Gao, H., Dahlman-Wright, K., and Jacobsson, A. (2016). Estrogen Enhances the Expression of the Polyunsaturated Fatty Acid Elongase Elovl2 via ERα in Breast Cancer Cells. PLoS ONE *11*, e0164241.

46. Kobayashi, T., Zadravec, D., and Jacobsson, A. (2007). ELOVL2 overexpression enhances triacylglycerol synthesis in 3T3-L1 and F442A cells. FEBS Lett. *581*, 3157–3163.

47. Pauter, A.M., Olsson, P., Asadi, A., Herslöf, B., Csikasz, R.I., Zadravec, D., and Jacobsson, A. (2014). Elovl2 ablation demonstrates that systemic DHA is endogenously produced and is essential for lipid homeostasis in mice. J. Lipid Res. *55*, 718–728.

48. Tikhonenko, M., Lydic, T.A., Wang, Y., Chen, W., Opreanu, M., Sochacki, A., McSorley, K.M., Renis, R.L., Kern, T., Jump, D.B., et al. (2010). Remodeling of retinal Fatty acids in an animal model of diabetes: a decrease in long-chain polyunsaturated fatty acids is associated with a decrease in fatty acid elongases Elovl2 and Elovl4. Diabetes *59*, 219–227.

49. Han, J.M., Kim, T.Y., Jeon, M.J., Yim, J.H., Kim, W.G., Song, D.E., Hong, S.J., Bae, S.J., Kim, H.K., Shin, M.H., et al. (2013). Obesity is a risk factor for thyroid cancer in a large, ultrasonographically screened population. Eur. J. Endocrinol. *168*, 879–886.

50. Hwang, Y., Lee, K.E., Park, Y.J., Kim, S.J., Kwon, H., Park, D.J., Cho, B., Choi, H.C., Kang, D., and Park, S.K. (2016). Annual Average Changes in Adult Obesity as a Risk Factor for Papillary Thyroid Cancer: A Large-Scale Case-Control Study. Medicine (Baltimore) *95*, e2893.

51. Kim, S.H., Park, H.S., Kim, K.H., Yoo, H., Chae, B.J., Bae, J.S., Jung, S.S., and Song, B.J. (2015). Correlation between obesity and clinicopathological factors in patients with papillary thyroid cancer. Surg. Today *45*, 723–729.

52. Oberman, B., Khaku, A., Camacho, F., and Goldenberg, D. (2015). Relationship between obesity, diabetes and the risk of thyroid cancer. Am. J. Otolaryngol. *36*, 535–541.

53. Huang, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

OMTN, Volume *18*

Supplemental Information

Prediction of the Prognosis Based on Chromosomal Instability-Related DNA Methylation Patterns of *ELOVL2* and *UBAC2* in PTCs

Jun Han, Meijun Chen, Qingxiao Fang, Yanqing Zhang, Yihan Wang, Jamaspishvili Esma, and Hong Qiao

Supplementary Figures

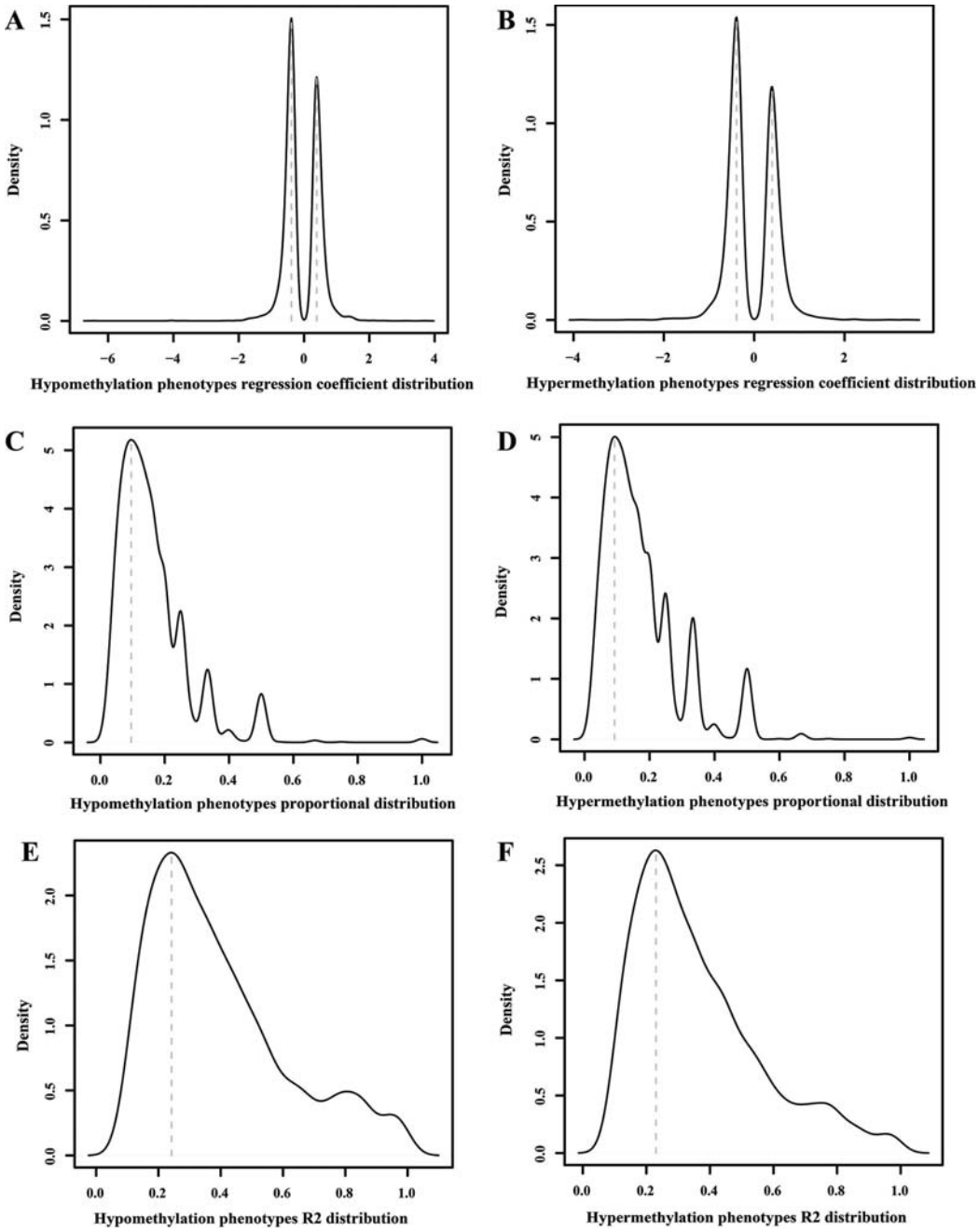

Figure S1. Correlation distribution. (A) More than 50% of regression coefficient distribution of the hypomethylated sites were negatively related to gene expression. (B) More than 50% of regression coefficient distribution of the hypermethylated sites were negatively related to gene expression. (C) Proportion of the hypomethylated sites significantly related to gene expression, peak value was 9.607311%, 2,035 genes≥9.607311%, which was 70% of total, 7 of them with 100% proportion. (D) Proportion of the hypermethylated sites significantly related to gene expression, peak value was 9.219488%, 2,082 genes≥9.219488%, which was 72% of total, 3 of

them with 100% proportion. (E) Distribution of the hypomethylated sites multiple regression correlation coefficient (R^2), 1,758 genes R^2>0.25, which was more than 50% of total. (F) Distribution of the hypermethylated sites multiple regression correlation coefficient (R^2), 1,694 genes R^2>0.25, which was more than 50% of total.

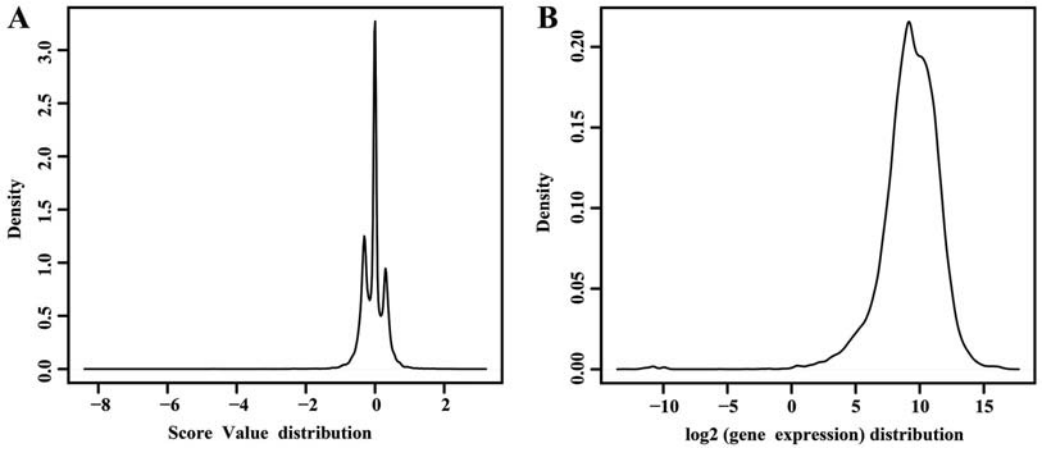

Figure S2. Data distribution. (A) The distribution of Score Value is consistent with the normal distribution. (B) The distribution of gene expression is consistent with the normal distribution.

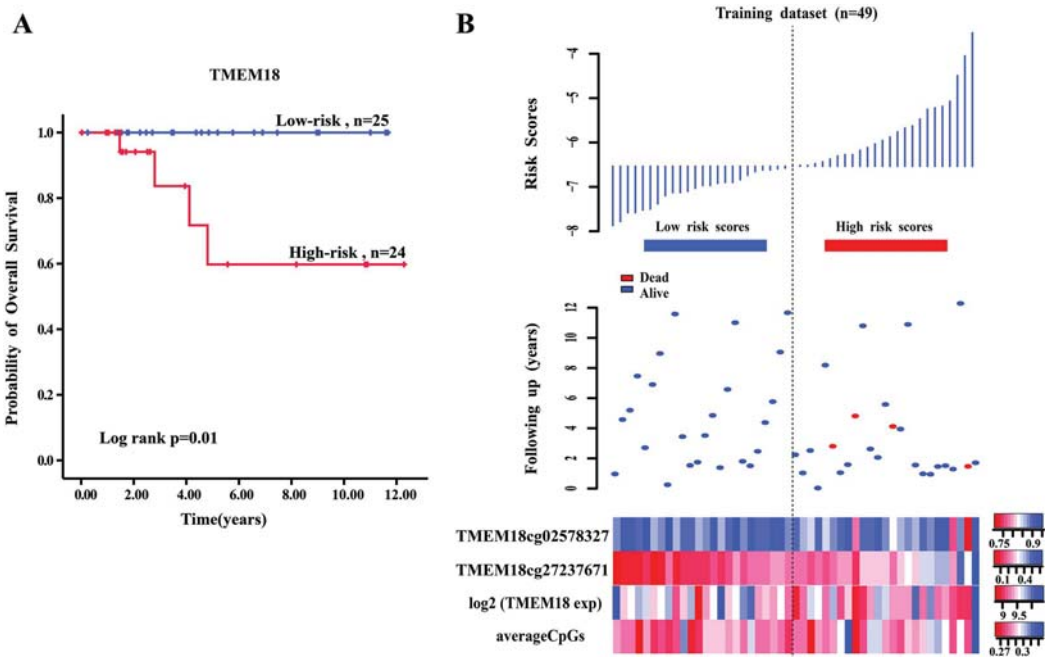

Figure S3. The *TMEM18* risk score model predicts overall survival of papillary thyroid carcinoma patients in the training dataset. (A) Kaplan-Meier analysis for overall survival of patients with high-risk or low-risk scores. P value was calculated using the two-sided log-rank test. (B) DNA methylation pattern, expression level and average methylation level of the prognostic *TMEM18* that correlates with patients' survival status and increased risk scores.

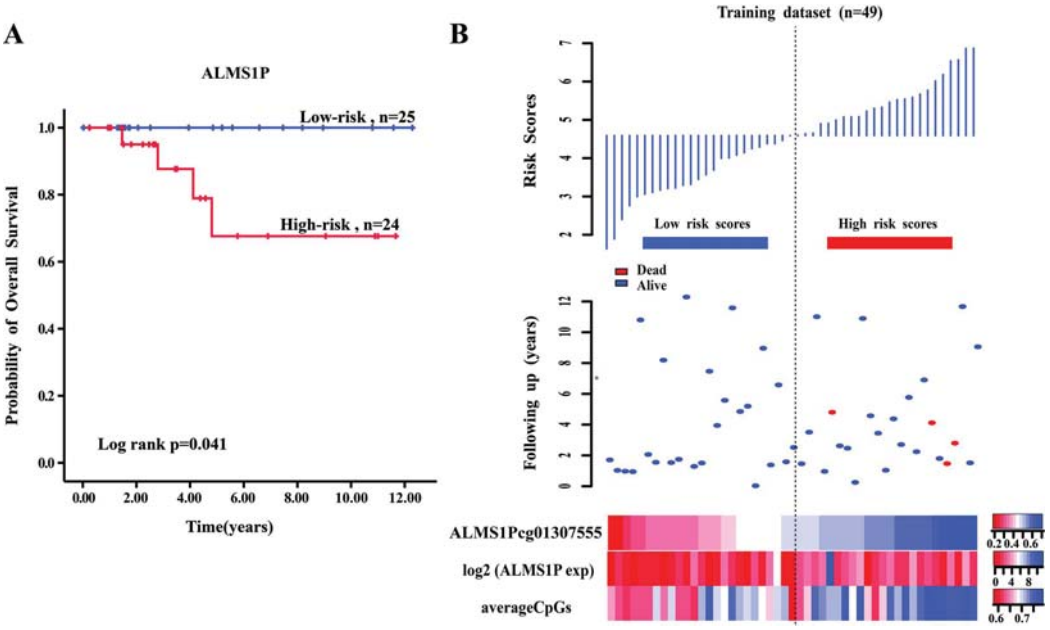

Figure S4. The *ALMS1P* risk score model predicts overall survival of papillary thyroid carcinoma patients in the training dataset. (A) Kaplan-Meier analysis for overall survival of patients with high-risk or low-risk scores. P value was calculated using the two-sided log-rank test. (B) DNA methylation pattern, expression level and average methylation level of the prognostic *ALMS1P* that correlates with patients' survival status and increased risk scores.

Table S1. 1,239 genes were significantly related to chromosomal instability.

Gene Name					
PARM1	SLC7A10	CTNNA2	PDE4C	GOLT1A	KLHDC2
CACNB3	F10	RNASE1	GALR2	FER1L6	MKX
KRTDAP	SLC16A11	RRP15	NES	SYDE2	AGAP7
CCDC48	NRN1L	ALG12	PML	SNRNP48	CACNA1A
PLOD2	STRN3	CLDN14	PON3	C9orf170	CALM1
IRF5	PHF1	ANKH	ZNF138	PGS1	GLI3
GGT6	PPARD	CMAH	SNX11	TPD52	AIF1
ZNF517	PIK3R1	KCNJ15	FJX1	GLCCI1	PITPNA
GJB3	ZC3HAV1	HEXIM2	KCTD15	CBX3	SLC17A7
HCRTR2	MPL	TGM2	COL28A1	MCM3	SYT10
CRBN	SUSD2	PLEKHG1	TPSAB1	TEX2	ALPL
KY	PDIA6	ZNF300	CTF1	ABCA1	MGLL
IGSF11	MYLK4	FAM109A	DNAJC3	TF	SLC4A4
SNX22	DEPDC1B	BMPR1A	TBC1D22B	FFAR2	EDARADD
MAN2B1	BSCL2	C8orf31	CISH	DNAJC5B	ENAH
GLT25D2	IRX1	SPINK5	SHROOM1	OLR1	SYNE2
MRPL44	DPF3	HLA-DMB	CLUAP1	SMAD7	CUL4A
ALS2CL	IL17RE	AKR7L	DARS	HSPC157	ATG9B
PTGDS	C21orf70	LOC728613	DGKB	PEX19	TMEM175
PMM1	ACPL2	SPSB2	ZBTB20	FAM55D	BEND7
PTPRA	SNX29	TRPV2	BEND4	HCFC2	SLC6A7
UBE2Z	ARHGAP10	LAMA4	GPR56	FAM63A	ADAMTSL3
AIF1L	SLC25A37	LEPREL2	LRRC16A	TNS3	MAP3K8
MYLK2	GPHN	KLF5	MMP9	FAM47E	MDGA2
C20orf117	ADD3	SPIN3	RHOBTB1	CYR61	NT5C2
TMEM215	NPHP1	SATL1	CHST11	FSTL1	C14orf4
PVRL4	ERN2	ACIN1	SLC26A11	ANXA4	VARS2
UPK2	PRKCD	RASSF8	TIGIT	PSRC1	FFAR1
DNASE2	GATA3	LASS4	SNTG2	C13orf38	VANGL2
DGKI	ASPRV1	ITFG3	SAFB	MAFG	IL12RB2
MTMR11	FERMT3	C9orf163	LINGO3	SLC45A2	TTC28
PLB1	MYB	LFNG	ELOVL5	CLIC1	C5orf20
SERPINA1	EP400	ATP6V0A2	HIPK2	SPRED1	STL
C1orf70	SEL1L3	DDRGK1	C1orf161	GABRP	LHFPL5
PLXNA2	SOX1	TTC24	CCDC105	GPR31	HSPA1A
TMEM108	OTUD7A	TIFAB	DCTD	F11	NAV3
APEH	MFNG	NAALADL2	HLA-E	ANXA11	PROCR
PTPN22	VAC14	CMTM4	TBX1	SGK269	PTK2B
ZNF204P	HGFAC	MORN4	SEMA3B	FLII	MAPRE3
DIAPH1	TMC8	VDR	FGF1	IFT172	GNLY
HHIPL1	SDHA	C6orf155	HOXA2	RNH1	CCR9
ADORA1	ZFYVE21	SPINT2	ACCN1	KIAA1644	PTBP1

PRRT3	C5orf32	ARF6	FZR1	CFH	MAP7D1
TRPC4AP	DISP2	ACCN3	NSUN7	SHC4	SLC16A4
P4HA2	FAM13AOS	LY9	C2orf54	TMEM179	SNX10
SEC14L4	TRAF2	SERPINA3	CDC42BPB	WWP1	ZSWIM4
PDK4	DQX1	C17orf58	NSMCE1	ARIH2	SPARC
ALMS1P	WBSCR16	SGCB	CCDC42	MACROD2	RIN3
IKZF2	PTCRA	RASSF6	SLC22A17	DISP1	THBS4
SHD	FAM184B	ENTPD7	KCTD1	PCYT1A	TAP1
MLLT4	AEBP1	NET1	SATB1	IL4	ST6GALNAC3
NPEPL1	ASCL3	CXorf56	THAP4	HGSNAT	TARS
CAND2	RNF222	WDR19	HS3ST4	GPR120	KRT4
MLPH	C16orf52	POU6F2	RAVER2	IMPACT	TMEM121
CHI3L2	EPHA7	KIRREL3	PUM1	STK33	FCER1G
SLC24A4	TMEM156	TBC1D9	MYOF	CSGALNACT2	ASB18
OR6F1	PDE6B	GMPR	BAI3	CTTN	FIBP
UTRN	GKN2	DCP2	ADRA2A	POLR3K	RBMS2
RAG1AP1	EGFR	SLC4A2	BAK1	NTSR1	SPINT1
NFATC4	LOC400657	PROKR2	C10orf11	POLA2	IQCJ
WISP2	PKNOX2	ACSM5	C2orf55	CLN8	OPN3
ADRM1	FGF23	STK17A	SMYD2	KRTCAP3	AKR1B1
MEGF9	PEBP4	ROPN1L	KIAA0495	ANKS1B	CNNM1
CTNNA1	TRMT1	FCRL5	LMOD2	TREML2	IL10RB
ABTB2	NODAL	C22orf43	SPTBN4	VAV3	FAM46B
TMEM233	IL21R	EXT2	TRAK1	JAK1	LOC285033
ESYT3	MICALL1	MATR3	TBR1	ZC3H6	TUFT1
UACA	TMEM18	APLP2	DYNC1I2	GDF6	SNX33
DNAJC6	OSR1	XPR1	FRAS1	EPHA10	PPP3CC
ERC1	PTPRCAP	COBL	CLDN11	TNFAIP3	PTGIR
PI4KB	LAG3	THRA	CHL1	RCN1	CYP27C1
C6orf168	INPP5F	CEACAM1	AHR	C6orf174	IMMP2L
C1orf111	EIF4G1	ALDH4A1	NEDD4L	NPBWR2	ACLY
TRAP1	RAB5C	BRE	DPYD	BTN3A3	ALOX12
CST6	SMTN	RNF212	JRK	CCM2	MGC29506
AP3S2	RGL3	TMIGD2	VPS13D	EPHA8	ZAK
MGEA5	ZBTB24	C17orf82	ADCY6	LOC285954	FOSL2
CARD9	CPAMD8	ZNF217	TBC1D22A	ABI2	C1orf173
LOX	E2F6	YWHAB	PSMB8	TOR1AIP1	ARHGAP23
KLRAQ1	TMEM132E	ZNF704	MBOAT1	SLC39A4	DHRS11
NRIP1	TCFL5	DAGLA	GZF1	GNA14	KCTD3
FAM92B	C1orf230	TAS2R16	MMP14	MTUS1	GFI1
PTGER1	PPAPDC1A	NCOA4	DEDD2	IL15	ASCL4
ZNF133	LOC283050	C7orf4	CYTH1	CYTH4	RFC1

NECAB2	GNB1	GPR83	C11orf75	ST6GAL2	SMAD3
PHACTR2	BTN1A1	ZNF606	ESPNP	USP10	ATP6V1H
MN1	BCAM	ZNF259	SHMT2	TESC	KCNH4
TCF23	NDST1	C1orf187	ZFP37	EOMES	UBE2QL1
NKX2-8	PDGFD	SLC6A12	TJP3	PRSS42	INHBB
KIAA1191	MICB	SPAG17	RAD9A	ASB9	CDRT4
UNC5CL	IL16	RPP14	KLHDC7A	GRM2	NIPA1
NHSL1	FNDC3B	TMEM87A	ZNF544	ACY1	C6orf186
GATSL3	NECAB1	ZNF876P	RABGAP1	SLC6A3	PDE4DIP
CREB3L1	ABCC1	EBAG9	DBC1	SGSM1	ZFP64
ABCB9	B3GALT4	KLF6	ZWINT	CLIP4	EN2
MPHOSPH6	SGMS1	PTPLA	GMIP	HCG9	GBX2
RUFY3	KCNN4	IGF1	G0S2	RAPGEF3	XPNPEP2
KCNA6	DENND1C	LGALS12	NAT8B	SLC5A9	RFX4
AXL	NR1H3	SFTA3	EPHB1	MFSD6	DDX47
TRIM54	SLFN12L	FBXO32	ABHD15	NUAK2	ATP8B4
MED29	PHACTR1	RAB17	TINAGL1	HMG20B	NKD2
FBXL18	PDCD1	CD80	FZD8	MTF1	COMMD3
ADAM15	TMC3	RAB31	ZNF774	RAPGEF1	SPATA7
TXNDC11	FLG2	SYNJ2	ATP6V1B2	WRNIP1	BAIAP2L1
NTHL1	STEAP3	GNAT1	SETD3	SHB	CTNNBIP1
SUSD1	RIBC2	RHOBTB2	C2orf3	ARHGEF4	EGFLAM
CAPN5	PGCP	LDB2	ZNF502	FZD5	SETMAR
ADCY7	SOAT2	PRPF4B	FAT3	SGEF	MASP1
NXF3	TAF4	SAMD11	LRRC1	WDR66	RAN
CDK6	VSNL1	FAM129B	C12orf56	WNK2	PTPN7
TLR2	RORA	MAPKAP1	POU2F2	TNC	C20orf108
SLC12A6	TEDDM1	CDH2	LPAR4	NFIL3	TMEM140
TP53I3	RAD51L1	MFAP3L	APBA2	TMEM72	ELOVL2
FAM101A	CTBP1	FAM54B	TMEM30B	SCTR	TRHR
ZFAT	CTHRC1	HOMER2	CYP2W1	RFC2	KCTD10
NDUFA10	TSPAN9	SORT1	B4GALNT4	KIAA1045	PVT1
CPO	RCSD1	AK3	MARCH10	C10orf26	CRTAM
GTSF1	XYLT1	HOXA10	NDUFA11	SLC38A3	PRMT2
SPIB	CRTAP	HCN4	MXRA5	EXT1	DPP9
FMOD	KRBA1	CCND1	DEPDC5	SERPINC1	C10orf58
PGAP3	NUPR1	EDIL3	DCAKD	TMEM149	ADARB1
MXD4	SLC16A5	MCTS1	GNAI1	GTPBP8	FAM13C
UBE2C	ITGA4	PYCARD	PDPN	ZNF594	CRELD2
PYROXD2	CCDC81	ZNF560	SIGLEC6	ABCA2	NT5C3
RSPH4A	IRF8	INADL	B4GALT6	SHC3	RAB11FIP1
ETS1	GDF11	YAP1	ESRP1	CPT1C	AKAP12
ABCA3	MFGE8	TRIM10	PROX1	SLC20A2	POU2F1
RGS9BP	ZFPM2	UBAC2	IL24	DOK2	ORC4L

CDK5R1	BAIAP3	TM9SF1	ASIP	MPDZ	MKRN3
RNF135	C15orf23	CCDC85C	ZNF322A	DCAF12	EGR3
KIAA1274	ZNF839	JARID2	ITGAL	GUCY2D	ZNF793
C10orf82	RCOR1	TMEM110	SLC38A5	SSH2	PPP1R14C
MED16	LOC153328	GDI2	MGST3	KIAA0020	THSD1
ZSWIM5	RAP2B	C1R	LOC285692	FOXB2	SCML4
LRP12	TRAF3	EDNRA	SDSL	ZNF703	FBXL21
SDCBP2	ALX4	GLB1L3	FLT3	KEL	SLC25A29
SUV420H1	PABPC4	FBXO21	ZNF397OS	HUNK	COL1A1
LOC148696	UBE2I	DGKD	CASP4	TPRKB	EPHA5
GPATCH1	KRT6A	WNT5A	NFKBIL2	PADI4	TMEM150C
CRYBB1	BMP5	EFNA3	NOV	KRT7	HOXA1
ULK2	HDLBP	ELOVL4	SIAH2	GRIN3A	NRBP2
SLC38A8	ASPH	PSORS1C1	TUB	PKLR	SLC25A42
PNPLA7	EPB49	CD79A	MLF1	FAM105B	CPSF4
WDFY4	C1orf113	STK39	SMAD6	C2orf73	TTC7A
RSPO1	ASCL2	CORO1A	GCOM1	C6orf132	TCOF1
CAPG	LAMA5	PSMB9	SNX4	CDH22	UBTF
CYP27B1	PLEKHA1	C11orf24	SOX6	TNFRSF13C	SPEN
WNT7A	THEM4	CCDC12	EPOR	WNK4	DNAL4
LOC399959	SPRY4	GOLGA3	PIP4K2B	RASGEF1A	ZNF814
SLC38A7	APPL1	SLC13A3	HAVCR2	LOC339674	C2orf85
FHL3	SLC5A11	CNRIP1	MTHFR	GLUL	RNF185
FAM84B	PPP1R1B	ZNF438	ABHD6	UBR4	TPCN1
RDH13	PABPN1L	GATA4	FARS2	RBP7	PLAC2
C6orf142	PUSL1	SPATA2L	L3MBTL4	HPN	C21orf88
PPFIBP2	CABLES1	TPM4	BMP2	RAB19	KRT12
CHAC1	SPATS2L	FLJ42709	NUDT16	ROBO1	GPR124
ZNF311	SNTG1	CYB5R4	CHRM3	ICOS	SPATC1
PDE4A	KIFC3	GNAQ	CEACAM3	LAPTM5	C1orf168
IL6	MAP1LC3B2	ABCB6	SHISA2	CASC3	RHOG
TMPRSS13	LMNB1	AQP2	TPBG	CRHR2	BCL2L15
VPS37B	MAPKAPK2	BCAR1	KIAA0040	BTN2A3	RASGRF1
SIAH1	EVC2	C1orf93	TMEM63A	BAI2	PARP1
MRGPRE	PLCG1	C14orf23	GPR158	TBC1D10C	GOLIM4
FAM163A	KHDC1L	RNF126P1	LATS2	APOA4	PCCA
RNF186	ELMO3	KRT78	C1orf103	PPP1R1C	VEGFA
N4BP2L2	ZDHHC23	HSD11B1	CRYBG3	RDX	GPR20
ICK	VWA3B	TMC6	ZSCAN22	SLC40A1	UCP2
ZDHHC1	FAM3D	GPR61	C6orf176	ENTPD4	CPA5
GRINA	HOXB9	ENPP7	AP1M2	PPP2R5C	C10orf110
ZFP1	RUNX1	CYB561	LOC284009	TPCN2	CCDC116
DNAH6	MIPOL1	PAX2	CCRL2	RNPEP	SMYD5

ABCC3	C6orf118	LAX1	NAA25	NRM	TGFA
CDH7	TRPM5	ANKK1	NFKBIE	C6orf25	NRP1
SLC30A2	NRD1	IMPDH1	ST8SIA1	ONECUT1	CELSR2
HES5	BCL10	LAT	SERPINI2	C2orf29	IFI30
SEC11A	ARL9	GRB2	GAL3ST3	SPATA20	CX3CR1
FREM2	SLC22A5	C1orf86	SORCS1	KIAA0415	COL17A1
LRFN1	IL6R	C22orf15	SHMT1	LIPC	NID2
LOC100132354	KALRN	FCGRT	NMI	C10orf81	APOD
FASLG	DYNLL2	ARPC1B	MYO16	ZNF528	CDH1
BTBD19	SACS	PRDX1	FBXL2	KIAA0427	SYT2
MBNL1	CCDC68	KCNJ1	RPS6KA1	ADCYAP1R1	SNCA
MORN1	PDCD1LG2	PHLPP1	IL5RA	MARCH4	KISS1
CRLF3	KRT8	RFWD2	FAAH	FAM20C	RNF40
PSMA8	MC1R	TNFAIP8L3	MAP7	ZNF529	CENPN
SGIP1	IFIT3	TMEM132A	METRNL	EML1	FLJ33360
REG4	IFIH1	ZKSCAN2	EFNA1	MAPK10	RGL1
ABCB10	TTPA	GLDN	SETD6	RBM45	CMPK1
HMGCS1	IL31RA	TBC1D12	ITPK1	FBXO6	NINJ1
LOXL1	SLC25A38	CCR7	CDKL1	TRUB1	PLCG2
DECR2	IGF2BP2	SH3BP1	MET	HLA-DPB1	TACC2
TRAF3IP2	PARD3	FAM172A	GRIN2C	SDCBP	NAA40
RAC1	CASP7	SLC4A9	GABARAPL1	ASB16	SLC39A11
GRAP2	CNTN6	CLPTM1L	ANKFN1	DDR2	LOC100190939
ICAM4	STRADA	FAM115C	SLC9A2	EEPD1	TFPI
TAF12	SLCO2A1	SIL1	CYB5R1	FAM57A	GTF3C1
CXCL1	ZNF853	ORAOV1	SLC41A3	NR2E1	RFX1
CHN2	MED12L	RAP1A	MLN	NSD1	MTIF2
VAMP1	LOC146880	RNF166	SLC6A1	FOXG1	RARRES2
TGFB1I1	FOXL1	GEMIN8	DIRC2	FAM181B	MOXD1
UTF1	WRB	SOCS2	MRAS	CIB4	SIDT1
NCK2	CBLB	CD3D	IGFBP5	FAM171A2	CLEC18B
TAP2	ARHGAP6	DDX51	CX3CL1	PRDM10	UNKL
ST3GAL1	ADK	CAMK2D	LRRC42	PLXNA4	C12orf42
MOBKL1B	ACVR1C	CDH16	DNAH14	LCP1	CAMTA2
PHF21B	E4F1	TLE2	C17orf97	MAPKBP1	TRIM58
PAX9	KIAA1199	SLC2A9			

Table S2. 572 of 1,239 genes were hypermethylation phenotypes.

Gene Name					
PARM1	RUFY3	PLCG1	CXorf56	NDUFA11	TPD52
CACNB3	KCNA6	KHDC1L	WDR19	MXRA5	GLCCI1
KRTDAP	AXL	ELMO3	POU6F2	DEPDC5	CBX3
CCDC48	TRIM54	ZDHHC23	KIRREL3	DCAKD	MCM3
PLOD2	MED29	VWA3B	TBC1D9	GNAI1	TEX2
IRF5	FBXL18	FAM3D	GMPR	PDPN	ABCA1
GGT6	ADAM15	HOXB9	DCP2	SIGLEC6	TF
ZNF517	TXNDC11	RUNX1	SLC4A2	B4GALT6	FFAR2
GJB3	NTHL1	MIPOL1	PROKR2	ESRP1	DNAJC5B
HCRTR2	SUSD1	C6orf118	ACSM5	PROX1	OLR1
CRBN	CAPN5	TRPM5	STK17A	IL24	SMAD7
KY	ADCY7	NRD1	ROPN1L	ASIP	PEX19
IGSF11	NXF3	BCL10	FCRL5	ZNF322A	FAM55D
SNX22	CDK6	ARL9	C22orf43	ITGAL	HCFC2
MAN2B1	TLR2	SLC22A5	EXT2	SLC38A5	FAM63A
GLT25D2	SLC12A6	IL6R	MATR3	MGST3	TNS3
MRPL44	TP53I3	KALRN	APLP2	LOC285692	FAM47E
ALS2CL	FAM101A	DYNLL2	XPR1	SDSL	CYR61
PTGDS	ZFAT	SACS	COBL	FLT3	FSTL1
PMM1	NDUFA10	CCDC68	THRA	ZNF397OS	ANXA4
PTPRA	CPO	PDCD1LG2	CEACAM1	CASP4	PSRC1
UBE2Z	GTSF1	KRT8	ALDH4A1	NFKBIL2	C13orf38
AIF1L	SPIB	MC1R	BRE	NOV	MAFG
MYLK2	FMOD	IFIT3	RNF212	SIAH2	SLC45A2
C20orf117	PGAP3	IFIH1	TMIGD2	TUB	CLIC1
TMEM215	MXD4	TTPA	C17orf82	MLF1	SPRED1
PVRL4	UBE2C	IL31RA	ZNF217	SMAD6	GABRP
UPK2	PYROXD2	SLC25A38	YWHAB	GCOM1	GPR31
DNASE2	RSPH4A	IGF2BP2	ZNF704	SNX4	F11
DGKI	ETS1	PARD3	DAGLA	SOX6	ANXA11
MTMR11	ABCA3	CASP7	TAS2R16	EPOR	SGK269
PLB1	RGS9BP	CNTN6	NCOA4	PIP4K2B	FLII
SERPINA1	CDK5R1	STRADA	C7orf4	HAVCR2	IFT172
C1orf70	RNF135	SLCO2A1	GPR83	MTHFR	RNH1
PLXNA2	KIAA1274	ZNF853	ZNF606	ABHD6	KIAA1644
TMEM108	C10orf82	MED12L	ZNF259	FARS2	CFH
APEH	MED16	LOC146880	C1orf187	L3MBTL4	SHC4
PTPN22	ZSWIM5	FOXL1	SLC6A12	BMP2	TMEM179
ZNF204P	LRP12	WRB	SPAG17	NUDT16	WWP1
DIAPH1	SDCBP2	CBLB	RPP14	CHRM3	ARIH2
HHIPL1	SUV420H1	ARHGAP6	TMEM87A	CEACAM3	MACROD2
ADORA1	LOC148696	ADK	ZNF876P	SHISA2	DISP1

PRRT3	GPATCH1	ACVR1C	EBAG9	TPBG	PCYT1A
TRPC4AP	CRYBB1	E4F1	KLF6	KIAA0040	IL4
P4HA2	ULK2	KIAA1199	PTPLA	TMEM63A	HGSNAT
SEC14L4	SLC38A8	CTNNA2	IGF1	GPR158	GPR120
PDK4	PNPLA7	RNASE1	LGALS12	LATS2	IMPACT
ALMS1P	WDFY4	RRP15	SFTA3	C1orf103	STK33
IKZF2	RSPO1	ALG12	FBXO32	CRYBG3	CSGALNACT2
SHD	CAPG	CLDN14	RAB17	ZSCAN22	CTTN
MLLT4	CYP27B1	ANKH	CD80	C6orf176	FFAR1
NPEPL1	WNT7A	CMAH	RAB31	AP1M2	VANGL2
CAND2	LOC399959	KCNJ15	SYNJ2	LOC284009	IL12RB2
MLPH	SLC38A7	HEXIM2	GNAT1	CCRL2	TTC28
CHI3L2	FHL3	TGM2	RHOBTB2	NAA25	C5orf20
SLC24A4	FAM84B	PLEKHG1	LDB2	NFKBIE	STL
OR6F1	RDH13	ZNF300	PRPF4B	ST8SIA1	LHFPL5
UTRN	C6orf142	FAM109A	SAMD11	SERPINI2	HSPA1A
RAG1AP1	PPFIBP2	BMPR1A	FAM129B	GAL3ST3	NAV3
NFATC4	CHAC1	C8orf31	MAPKAP1	SORCS1	PROCR
WISP2	ZNF311	SPINK5	CDH2	SHMT1	PTK2B
ADRM1	PDE4A	HLA-DMB	MFAP3L	NMI	MAPRE3
MEGF9	IL6	AKR7L	FAM54B	MYO16	GNLY
CTNNA1	TMPRSS13	LOC728613	HOMER2	FBXL2	CCR9
ABTB2	VPS37B	SPSB2	SORT1	RPS6KA1	PTBP1
TMEM233	SIAH1	TRPV2	AK3	IL5RA	MAP7D1
ESYT3	MRGPRE	LAMA4	ZFP37	FAAH	SLC16A4
UACA	FAM163A	LEPREL2	TJP3	MAP7	SNX10
DNAJC6	RNF186	KLF5	RAD9A	METRNL	ZSWIM4
ERC1	N4BP2L2	SPIN3	KLHDC7A	EFNA1	SPARC
PI4KB	ICK	SATL1	ZNF544	SETD6	RIN3
C6orf168	ZDHHC1	ACIN1	RABGAP1	ITPK1	THBS4
C1orf111	GRINA	RASSF8	DBC1	CDKL1	TAP1
TRAP1	ZFP1	LASS4	ZWINT	MET	ST6GALNAC3
CST6	DNAH6	ITFG3	GMIP	GRIN2C	TARS
AP3S2	ABCC3	C9orf163	G0S2	GABARAPL1	KRT4
MGEA5	CDH7	LFNG	NAT8B	ANKFN1	TMEM121
CARD9	SLC30A2	ATP6V0A2	EPHB1	SLC9A2	FCER1G
LOX	HES5	DDRGK1	ABHD15	CYB5R1	ASB18
KLRAQ1	SEC11A	TTC24	TINAGL1	SLC41A3	FIBP
NRIP1	FREM2	TIFAB	FZD8	MLN	RBMS2
FAM92B	THEM4	NAALADL2	ZNF774	SLC6A1	SPINT1
PTGER1	SPRY4	CMTM4	ATP6V1B2	DIRC2	IQCJ

ZNF133	APPL1	MORN4	SETD3	MRAS	OPN3
NECAB2	SLC5A11	VDR	C2orf3	IGFBP5	AKR1B1
PHACTR2	PPP1R1B	C6orf155	ZNF502	CX3CL1	CNNM1
MN1	PABPN1L	SPINT2	FAT3	LRRC42	IL10RB
TCF23	PUSL1	ARF6	LRRC1	DNAH14	FAM46B
NKX2-8	CABLES1	ACCN3	C12orf56	C17orf97	LOC285033
KIAA1191	SPATS2L	LY9	POU2F2	TRIM58	TUFT1
UNC5CL	SNTG1	SERPINA3	LPAR4	GOLT1A	SNX33
NHSL1	KIFC3	C17orf58	APBA2	FER1L6	HSPC157
GATSL3	MAP1LC3B2	SGCB	TMEM30B	SYDE2	PGS1
CREB3L1	LMNB1	RASSF6	CYP2W1	SNRNP48	MARCH10
ABCB9	MAPKAPK2	ENTPD7	B4GALNT4	C9orf170	NET1
MPHOSPH6	EVC2				

Table S3. 667 of 1,239 genes were hypomethylation phenotypes.

Gene Name					
LRFN1	PPAPDC1A	RNF126P1	FRAS1	LOC339674	GBX2
LOC100132354	LOC283050	KRT78	CLDN11	GLUL	XPNPEP2
FASLG	GNB1	HSD11B1	CHL1	UBR4	RFX4
BTBD19	BTN1A1	TMC6	AHR	RBP7	DDX47
MBNL1	BCAM	GPR61	NEDD4L	HPN	ATP8B4
MORN1	NDST1	ENPP7	DPYD	RAB19	NKD2
CRLF3	PDGFD	CYB561	JRK	ROBO1	COMMD3
PSMA8	MICB	PAX2	VPS13D	ICOS	SPATA7
SGIP1	IL16	LAX1	ADCY6	LAPTM5	BAIAP2L1
REG4	FNDC3B	ANKK1	TBC1D22A	CASC3	CTNNBIP1
ABCB10	NECAB1	IMPDH1	PSMB8	CRHR2	EGFLAM
HMGCS1	ABCC1	LAT	MBOAT1	BTN2A3	SETMAR
LOXL1	B3GALT4	GRB2	GZF1	BAI2	MASP1
DECR2	SGMS1	C1orf86	MMP14	TBC1D10C	RAN
TRAF3IP2	KCNN4	C22orf15	DEDD2	APOA4	PTPN7
RAC1	DENND1C	FCGRT	CYTH1	PPP1R1C	C20orf108
GRAP2	NR1H3	ARPC1B	C11orf75	RDX	TMEM140
ICAM4	SLFN12L	PRDX1	ESPNP	SLC40A1	TRHR
TAF12	PHACTR1	KCNJ1	SHMT2	ENTPD4	KCTD10
CXCL1	PDCD1	PHLPP1	POLR3K	PPP2R5C	PVT1
CHN2	TMC3	RFWD2	NTSR1	TPCN2	CRTAM
VAMP1	FLG2	TNFAIP8L3	POLA2	RNPEP	PRMT2

TGFB1I1	STEAP3	TMEM132A	CLN8	NRM	DPP9
UTF1	RIBC2	ZKSCAN2	KRTCAP3	C6orf25	C10orf58
NCK2	PGCP	GLDN	ANKS1B	ONECUT1	ADARB1
TAP2	SOAT2	TBC1D12	TREML2	C2orf29	FAM13C
ST3GAL1	TAF4	CCR7	VAV3	SPATA20	CRELD2
MOBKL1B	VSNL1	SH3BP1	JAK1	KIAA0415	NT5C3
PHF21B	RORA	FAM172A	ZC3H6	LIPC	RAB11FIP1
PAX9	TEDDM1	SLC4A9	GDF6	C10orf81	AKAP12
SLC7A10	RAD51L1	CLPTM1L	EPHA10	ZNF528	POU2F1
F10	CTBP1	FAM115C	TNFAIP3	KIAA0427	ORC4L
SLC16A11	CTHRC1	SIL1	RCN1	ADCYAP1R1	MKRN3
NRN1L	TSPAN9	ORAOV1	C6orf174	FAM20C	EGR3
STRN3	RCSD1	RAP1A	NPBWR2	ZNF529	ZNF793
PHF1	XYLT1	RNF166	BTN3A3	EML1	PPP1R14C
PPARD	CRTAP	GEMIN8	CCM2	MAPK10	THSD1
PIK3R1	KRBA1	SOCS2	EPHA8	RBM45	SCML4
ZC3HAV1	NUPR1	CD3D	LOC285954	FBXO6	FBXL21
MPL	SLC16A5	DDX51	ABI2	TRUB1	SLC25A29
SUSD2	ITGA4	CAMK2D	TOR1AIP1	HLA-DPB1	COL1A1
PDIA6	CCDC81	CDH16	SLC39A4	SDCBP	EPHA5
MYLK4	IRF8	TLE2	GNA14	ASB16	TMEM150C
DEPDC1B	GDF11	SLC2A9	MTUS1	DDR2	HOXA1
BSCL2	MFGE8	PDE4C	IL15	EEPD1	NRBP2
IRX1	ZFPM2	GALR2	CYTH4	FAM57A	SLC25A42
DPF3	BAIAP3	NES	ST6GAL2	NR2E1	CPSF4
IL17RE	C15orf23	PML	USP10	NSD1	TTC7A
C21orf70	ZNF839	PON3	EOMES	FOXG1	TCOF1
ACPL2	RCOR1	ZNF138	PRSS42	FAM181B	UBTF
SNX29	LOC153328	SNX11	ASB9	CIB4	SPEN
ARHGAP10	RAP2B	FJX1	GRM2	FAM171A2	DNAL4
SLC25A37	TRAF3	KCTD15	ACY1	PRDM10	ZNF814
GPHN	ALX4	COL28A1	SLC6A3	PLXNA4	C2orf85
ADD3	PABPC4	TPSAB1	SGSM1	LCP1	RNF185
NPHP1	UBE2I	CTF1	CLIP4	MAPKBP1	TPCN1
ERN2	KRT6A	DNAJC3	HCG9	CAMTA2	PLAC2
PRKCD	BMP5	TBC1D22B	RAPGEF3	KLHDC2	C21orf88
GATA3	HDLBP	CISH	SLC5A9	MKX	KRT12
ASPRV1	ASPH	SHROOM1	MFSD6	AGAP7	GPR124
FERMT3	EPB49	CLUAP1	NUAK2	CACNA1A	SPATC1
MYB	C1orf113	DARS	HMG20B	CALM1	C1orf168
EP400	ASCL2	DGKB	MTF1	GLI3	RHOG
SEL1L3	LAMA5	ZBTB20	RAPGEF1	AIF1	BCL2L15
SOX1	PLEKHA1	GPR56	WRNIP1	PITPNA	RASGRF1

OTUD7A	HOXA10	LRRC16A	SHB	SLC17A7	PARP1
MFNG	HCN4	MMP9	ARHGEF4	SYT10	GOLIM4
VAC14	CCND1	RHOBTB1	FZD5	ALPL	PCCA
HGFAC	EDIL3	CHST11	SGEF	MGLL	VEGFA
TMC8	MCTS1	SLC26A11	WDR66	SLC4A4	GPR20
SDHA	PYCARD	TIGIT	WNK2	EDARADD	UCP2
ZFYVE21	ZNF560	SNTG2	TNC	ENAH	CPA5
C5orf32	INADL	SAFB	NFIL3	SYNE2	C10orf110
DISP2	YAP1	LINGO3	TMEM72	CUL4A	CCDC116
FAM13AOS	TRIM10	ELOVL5	SCTR	ATG9B	SMYD5
TRAF2	UBAC2	HIPK2	RFC2	TMEM175	TGFA
DQX1	TM9SF1	C1orf161	KIAA1045	BEND7	NRP1
WBSCR16	CCDC85C	CCDC105	C10orf26	SLC6A7	CELSR2
PTCRA	JARID2	DCTD	SLC38A3	ADAMTSL3	IFI30
FAM184B	TMEM110	HLA-E	EXT1	MAP3K8	CX3CR1
AEBP1	GDI2	TBX1	SERPINC1	MDGA2	COL17A1
ASCL3	C1R	SEMA3B	TMEM149	NT5C2	NID2
RNF222	EDNRA	FGF1	GTPBP8	C14orf4	APOD
C16orf52	GLB1L3	HOXA2	ZNF594	VARS2	CDH1
EPHA7	FBXO21	ACCN1	ABCA2	PPP3CC	SYT2
TMEM156	DGKD	FZR1	SHC3	PTGIR	SNCA
PDE6B	WNT5A	NSUN7	CPT1C	CYP27C1	KISS1
GKN2	EFNA3	C2orf54	MARCH4	IMMP2L	RNF40
EGFR	ELOVL4	CDC42BPB	SLC20A2	ACLY	CENPN
LOC400657	PSORS1C1	NSMCE1	DOK2	ALOX12	FLJ33360
PKNOX2	CD79A	CCDC42	MPDZ	MGC29506	RGL1
FGF23	STK39	SLC22A17	DCAF12	ZAK	CMPK1
PEBP4	CORO1A	KCTD1	GUCY2D	FOSL2	NINJ1
TRMT1	PSMB9	SATB1	SSH2	C1orf173	PLCG2
NODAL	C11orf24	THAP4	KIAA0020	ARHGAP23	TACC2
IL21R	CCDC12	HS3ST4	FOXB2	DHRS11	NAA40
MICALL1	GOLGA3	RAVER2	ZNF703	KCTD3	SLC39A11
TMEM18	SLC13A3	PUM1	KEL	GFI1	LOC100190939
OSR1	CNRIP1	MYOF	HUNK	ASCL4	TFPI
PTPRCAP	ZNF438	BAI3	TPRKB	RFC1	GTF3C1
LAG3	GATA4	ADRA2A	PADI4	SMAD3	RFX1
INPP5F	SPATA2L	BAK1	KRT7	ATP6V1H	MTIF2
EIF4G1	TPM4	C10orf11	GRIN3A	KCNH4	RARRES2
RAB5C	FLJ42709	C2orf55	PKLR	UBE2QL1	MOXD1
SMTN	CYB5R4	SMYD2	FAM105B	INHBB	SIDT1
RGL3	GNAQ	KIAA0495	C2orf73	CDRT4	CLEC18B
ZBTB24	ABCB6	LMOD2	C6orf132	NIPA1	UNKL
CPAMD8	AQP2	SPTBN4	CDH22	C6orf186	C12orf42

E2F6	BCAR1	TRAK1	TNFRSF13C	PDE4DIP	ELOVL2
TMEM132E	C1orf93	TBR1	WNK4	ZFP64	TESC
TCFL5	C14orf23	DYNC1I2	RASGEF1A	EN2	BEND4