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 Qiao2

1Department of Endocrinology and Metabolism, The Second Affiliated Hospital, Harbin Medical University, Harbin 150001, China; 2Department of Endocrinology and Metabolism, The Second Affiliated Hospital, Harbin Medical University, Harbin 150001, China; 3Department of Endocrinology and Metabolism, The Fourth Affiliated Hospital, Harbin Medical University, Harbin 150001, China; 4Surgical Oncology, The Second Affiliated Hospital, Harbin Medical University, Harbin 150001, China; 5Hematological Department, The Second Affiliated Hospital, Harbin Medical University, Harbin 150001, China; 6College 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 by 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 did differentiate 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’ suggested that TBL1XR1 overexpression was an unfavorable prognostic factor for epithelial ovarian cancer, and Wang et al.7 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.8 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.9 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
Genomic DNA hypomethylation has been associated with increased CIN, which plays a central role in tumorigenesis.\(^{14}\) Genomic DNA hypomethylation has been associated with livers. Rodriguez et al.\(^{16}\) reported that CIN was correlated with cellular carcinoma, especially in the background of noncirrhotic thyellation of cancers, and Suzuki et al.\(^{17}\) reported that global DNA demethyla-
tion is associated with CIN.

\begin{table}
\small
\centering
\caption{Clinical Information on PTC Patients}
\begin{tabular}{lrr}
\hline
Characteristics & Training Set (n = 49) & Testing Set (n = 442) \\
\hline
\textbf{State} & & \\
Living & 45 & 431 \\
Dead & 4 & 11 \\
\hline
\textbf{Survival (years)} & & \\
Mean ± SD & 4.22 ± 3.52 & 2.46 ± 2.37 \\
\hline
\textbf{Gender} & & \\
Male & 14 & 117 \\
Female & 35 & 325 \\
\hline
\textbf{Age (years)} & & \\
Mean ± SD & 45.29 ± 17.22 & 47.57 ± 15.72 \\
\hline
\textbf{Stage} & & \\
I & 30 & 246 \\
II & 5 & 47 \\
III & 11 & 99 \\
IV & 3 & 50 \\
\hline
\textbf{Histological Type} & & \\
Thyroid papillary carcinoma & & \\
Classical/usual & 42 & 313 \\
Tall cell (≥ 50% tall cell features) & 3 & 33 \\
Follicular (≥ 99% follicular patterned) & 4 & 96 \\
\hline
\end{tabular}
\end{table}

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 \textit{UBAC2} and \textit{ELOVL2} that had high correlations with CIN provided good prognostic values for PTC. Moreover, \textit{UBAC2} and \textit{ELOVL2} were hypomethylation phenotypes. The DNA methylation patterns associated with CIN may therefore be a good predictor of PTC survival.

**RESULTS**

\textbf{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
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 three genes was influenced by all of their hypomethylated 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).
subdivided the PTC patients into high-risk and low-risk groups. The DNA methylation of the risk gene was bad for the survival time. We and (21.25 \times 10^{-9}) were 0.95, 0.886, 0.764, and 0.854, 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 = 0.01) (Figure S4). Therefore, a higher gene risk score predicted a worse PTC patient prognosis.

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 models. 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 \times \text{TMEM18}) + (−41.68 \times \text{UBAC2}) + (−6.20807 \times \text{ELOVL2}) + (21.25 \times \text{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 \times 10^{-9} (2.351e−17-0.1619), 7.909 \times 10^{-19} (3.658e−33-0.000171), 0.002013 (5.646e−06-0.7177), and 1.7 \times 10^{9} (1.075-2.687e+18), respectively (Table 2).

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 also 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

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) | −6.20807 | 0.0384 |
| ALMS1P   | 1.7 × 10^{9} (1.075-2.687e+18) | 21.25 | 0.0492 |

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
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 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 ELOVL2 were significantly different. Moreover, the ELOVL2 risk score was significantly different in the four stages of PTC (Kruskal-Wallis test; p = 9.915 × 10^{-15}) (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. 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.

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. 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, 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,
we used the R Significance Analysis of Microarrays (SAM) package 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.

Previous studies of ELOVL2 and UBAC2 emphasized their relationships with lipid metabolism and obesity. González-Bengtsson et al. suggested that ELOVL2 played an important role in docosahexaenoic acid (DHA) synthesis. Kobayashi et al. reported that cells overexpressing ELOVL2 showed enhanced triacylglycerol synthesis and subsequent accumulation of lipid droplets. Pauter et al. 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. 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. Han et al. reported that the morbidity of thyroid cancer in female patients was related to a high BMI. Hwang et al. suggested that weight gain and annual increases in obesity indicators in middle-aged adults increased the risk of developing PTC. Kim et al. reported that a higher BMI was associated with more

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.

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.
aggressive tumor features, such as lymph node metastasis, lymphatic invasion, and tumor multiplicity in PTC patients. Oberman et al. 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 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.

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.
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 prognosis of patients with similar clinical features. Our results showed that the DNA methylation patterns of \textit{ELOVL2} and \textit{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 prognosis of patients with similar clinical features and could provide a more appropriate therapeutic schedule for high-risk patients to enhance the efficacy of PTC treatments.

\textbf{Conclusions}

\textit{UBAC2} and \textit{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.

\textbf{MATERIALS AND METHODS}

\textbf{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/).

\textbf{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.

\textbf{Identifying Differentially Methylated CpGs}
To compare the differences of DNA methylation between cancer and normal samples in the training set, we used the R SAM44 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.

\textbf{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.

\textbf{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 as follows:

\[
\text{Score}_k = \alpha_1 \cdot c_{g1} + \alpha_2 \cdot c_{g2} + \ldots + \alpha_j \cdot c_{gj},
\]  

(Equation 1)

The \( j \) CpGs sites represented the significant methylation sites in multiple regression analyses (\( p < 0.05 \)), \( \alpha_j \) \( (j = 1, 2, \ldots) \) represented the multiple regression coefficient of the \( j \)th CpG sites of the gene, and \( c_{gj} \) represented the methylation level of the \( j \)th CpG sites of the gene in the \( k \)th sample.

\textbf{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:

\[
f_{Bi} = \frac{\mu_{Bi} - \mu_{Gi}}{\sqrt{(\sigma_{Bi}^2/N_B) + (\sigma_{Gi}^2/N_C)}}.
\]  

(Equation 2)

The \( \mu_{Bi} \) represented the average score value of all of the genes under sample \( i \) in the band, \( \sigma_{Bi}^2 \) represented the variance of the score value of all the genes under sample \( i \) in the band, \( \mu_{Gi} \) represented the average score value of the rest of the genes under sample \( i \), \( \sigma_{Gi}^2 \) 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 \( N_C \) 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:

\[
MFA_i = \sum_{Band} \left| f_{Bi} \right| .
\]  

(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 permuted 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:

\[
p = \frac{c + 1}{1001}
\]  

(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.\(^5\) 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:

\[
Risk\ Score_k = c \cdot genec_k,
\]  

(Equation 5)

where \( k \) was the \( k \)th 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 HYDSTPY201910).
REFERENCES

1. Budílesku, L.C., Bărbuș, E., and Piciu, D. (2017). Circulating tumor cells in thyroid carcinoma - the prognostic role of this biomarker. Review of the literature. Chujil 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. Lewinski, 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., McVier, 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. Tshohu 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., Saki, H., Kitao, H., Tsuda, Y., Otsubo, 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 - the prognostic role of this biomarker. Cancer Res. 21, 676–687.

16. 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.

17. 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.

18. 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.

19. El-Tahan, R.R., Ghoneim, A.M., and El-Mashad, N. (2016). TNF-α gene polymorphisms and expression. Springerplus 5, 1508.

20. Fujii, N., Matoba, N., and Matsuda, R. (2016). Environment Control to Improve Recombinant Protein Yields in Plants Based on Agrobacterium-Mediated Transient Gene Expression. Front. Biotech. Biotechnol. 4, 23.

21. 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.

22. Lu, X.X., Ge, M.H., Ling, Z.Q., Hu, S.S., Xu, J., Zheng, C.M., Tan, Z., and Chen, C. (2013). [Ablentration 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.
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. Taiaquina, 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.I., 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.I., 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.