DNA Methylation Haplotype Block Markers Efficiently Discriminate Follicular Thyroid Carcinoma from Follicular Adenoma

Hui Zhang,1,* Zhenzhen Zhang,2,* Xiaoding Liu,1 Huanli Duan,3 Tianmin Xiang,2 Qiye He,2 Zhixi Su,2 Huanwen Wu,1 and Zhiyong Liang1

1Department of Pathology, Molecular Pathology Research Center, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Science, Beijing, 100730, China; 2Singlera Genomics Inc. Shanghai 201318, China; and 3Department of Pathology, Xuanwu Hospital, Capital Medical University, Beijing, 100053, China

ORCID number: 0000-0003-4787-2925 (Z. Liang).

*These are joint first authors.

Abbreviations: DMC, differentially methylated CpG; EA, encapsulated angioinvasive; FA, follicular adenoma; FDR, false discovery rate; FFPE, formalin-fixed paraffin-embedded; FTC, follicular thyroid carcinoma; GO, Gene Ontology; MHB, methylation haplotype block; MHL, methylated haplotype load; MI, minimally invasive; NT, normal tissue; PDR, proportion of discordant read; PTC, papillary thyroid carcinoma; ROC, receiver operating characteristic; RRBS, reduced representation bisulfite sequencing; TC, thyroid carcinoma; TT-UMP, thyroid tumors with uncertain malignant potential; WI, widely invasive.

Received: 2 May 2020; Editorial Decision: 16 December 2020; First Published Online: 4 January 2021; Corrected and Typeset: 16 January 2021.

Abstract

Context: Follicular thyroid carcinoma (FTC) is the second most common type of thyroid carcinoma and must be pathologically distinguished from benign follicular adenoma (FA). Additionally, the clinical assessment of thyroid tumors with uncertain malignant potential (TT-UMP) demands effective indicators.

Objective: We aimed to identify discriminating DNA methylation markers between FA and FTC.

Methods: DNA methylation patterns were investigated in 33 FTC and 33 FA samples using reduced representation bisulfite sequencing and methylation haplotype block-based analysis. A prediction model was constructed and validated in an independent cohort of 13 FTC and 13 FA samples. Moreover, 36 TT-UMP samples were assessed using this model.

Results: A total of 70 DNA methylation markers, approximately half of which were located within promoters, were identified to be significantly different between the FTC and FA samples. All the Gene Ontology terms enriched among the marker-associated genes were related to “DNA binding,” implying that the inactivation of DNA binding played a role in FTC development. A random forest model with an area under the curve of 0.994 was constructed using those markers for discriminating FTC from FA in the
Thyroid carcinoma (TC) is the most common endocrine malignancy. The majority of TCs originate from follicular cells. There are 4 main pathological types of TC: papillary, follicular, medullary, and anaplastic (1). Among these, approximately 10% of all diagnosed TC cases are follicular thyroid carcinoma (FTC) (2). The diagnosis of FTC depends upon pathologic confirmation of follicular cells that lack the nuclear atypia seen in papillary thyroid carcinoma (PTC), in addition to capsular and/or vascular invasion (1). Follicular adenoma (FA) is the most common form of benign thyroid neoplasm and the benign counterpart of FTC (3). In general, it is difficult to distinguish FA from FTC based on cytologic features alone. However, FA differs significantly from FTC, in that it has no capsular invasion, vascular invasion, extrathyroidal tumor extension, or metastases (4).

In 2017, the World Health Organization proposed the concept of borderline thyroid tumors including hyalinizing trabecular tumors and other encapsulated recurrent pattern tumors (1). The latter tumors are divided into 2 types of thyroid tumors: thyroid tumors of uncertain malignant potential (TT-UMP) and noninvasive follicular thyroid neoplasm with papillary-like nuclear features (1). TT-UMP are further classified as encapsulated and follicular-patterned nodules representing suspicious capsular/vascular invasion or with incompletely developed PTC nuclear changes, and they fail to fulfill the criteria for malignancy of carcinoma (1). The accurate assessment of malignancy in individual TT-UMP is challenging. Although numerous efforts have been made to distinguish FTC, TT-UMP, and FA, one of the major challenges remaining is the interobserver variability in the histologic interpretation of capsular or vascular invasion. Immense progress has been made in developing molecular markers for clinical use, which are capable of identifying follicular tumors and assessing the malignancy of TT-UMP samples, including somatic gene mutations/rearrangements and gene expression classifiers (5, 6). However, ~8% of suspected high-risk lesions display no gene alterations (7). Furthermore, some common genetic alterations like RAS mutations and PAX8-PPARG rearrangements have been identified in FA and FTC (8), while gene expression profiling has revealed that FA and FTC are highly similar in enriched gene ontologies and pathways (9). These facts highlight the need for developing highly specific and feasible markers that can evaluate TT-UMP malignancy, in order to prevent either undertreatment or overtreatment of patients.

DNA methylation is one of the major epigenetic mechanisms that has been shown to play critical roles in carcinogenesis and tumor progression in multiple cancer types (10). Abnormal DNA methylation, frequently preceding mutations (2), may have the potential to serve as a biomarker for the early diagnosis of malignant nodules. Recently, multiple studies have demonstrated that different methylation patterns may be useful in clinically differentiating TC from benign nodules (11-13), implying that DNA methylation alterations may help us to distinguish FTC from FA, while concurrently determining the malignant potential of TT-UMP nodules.

Despite these efforts, a highly predictive model to distinguish FTC from FA has still not been developed, indicating that further in-depth investigations to identify robust FTC and/or FA-specific DNA methylation markers are urgently needed. It has been observed that adjacent CpG sites on the same DNA molecules tend to possess similar methylation statuses, attributable to the locally coordinated activities of the DNA methylation and/or demethylation enzymes (10). Those comethylated CpG sites may further form methylation haplotype blocks (MHBs). In this study, an MHB-based methylation analysis was performed on cohorts of specimens including malignant FTC tissues, benign FA lesions, and TT-UMP lesions. We focused on the difference of methylation patterns between FA and FTC to determine the necessary DNA methylation markers, and subsequently used them to build a prediction model for FTC and FA tissues, as well as to assess the malignant potentials of TT-UMP lesions.

**Patients and Methods**

**Patients and sampling**

A total of 153 tissue samples of surgically resected thyroid tumor initially diagnosed as FA, TT-UMP, and FTC were collected between June 2010 and May 2015 at the
Department of Pathology, Peking Union Medical College Hospital. All corresponding archived hematoxylin and eosin–stained tumor sections were collected and independently reevaluated by 2 experienced pathologists, based on the 2017 World Health Organization classification of endocrine tumors. A third pathologist was available if necessary. FTC samples were further classified into minimally invasive (MI) group, encapsulated angioinvasive (EA) group, and widely invasive (WI) group. All the samples were examined to establish whether Hürthle cells were present in their thyroid nodules. Twenty-five patients were excluded owing to the paucity in DNA for genetic testing. Finally, 128 patients, ie, 46 FA, 36 TT-UMP, and 46 FTC cases were included in the study. The clinical characteristics of the thyroid tumor patients are provided in Table 1 and (14). All enrolled patients provided consent, and this study was approved by the institutional review board of Peking Union Medical College Hospital (S-K1109).

Patients pathologically diagnosed with FTC or FA were randomly divided into a training cohort (33 FTC and 33 FA samples) and a validating cohort (13 FTC and 13 FA). An additional 36 patients with TT-UMP were recruited and their methylation profiles were investigated.

Formalin-fixed paraffin-embedded (FFPE) tumor specimens were collected from the enrolled patients. Fifteen randomly selected cases of adjacent normal tissues (NTs) were collected and examined as negative control. DNA was extracted from the FFPE samples using the QIAamp DNA FFPE Tissue Kit (Qiagen, Hilden, Germany) in accordance with the manufacturer’s instructions. For quality control, it was ensured that the majority of the DNA fragments were larger than 500 bp in size.

Preparation of DNA methylation libraries and sequencing

All DNA samples were prepared using the reduced representation bisulfite sequencing (RRBS) approach. Briefly, 50 to 100 ng of input DNA was converted using the Methylcode Bisulfite Conversion Kit (ThermoFisher, MECOV50), and subsequently dephosphorylated and ligated to a universal adapter containing sequences required for Illumina-based sequencing platforms. Ligated DNA was then amplified by PCR to add indices. The libraries were sequenced on an Illumina X10 platform.

DNA methylation data preprocessing

Both paired-end reads were adapter trimmed and merged to form single-end reads. Subsequently, the trimmed reads were mapped to the bisulfite-converted human reference genome (version hg19) with Bismark and Bowtie v.1. BAM files, and further processed for the following analysis.

To identify the differentially methylated CpGs (DMCs) of FTC and FA, with respect to the NT samples, CpG methyRate of the training cohort was calculated using Bismark. DMCs were called using the metilene program with the methyRate and identified on criteria of an adjusted false discovery rate (FDR) $P < .05$ as well as a mean delta beta ($\Delta \beta$) value that was either more than .2 or less than –.2. Heatmaps based on the FTC-specific DMCs and combined DMCs (including both FTC-specific DMCs and FA-specific DMCs) were generated using Morpheus (https://software.broadinstitute.org/morpheus) by applying hierarchical clustering and using the Euclidean distance as a similarity measure.

Identifying methylation haplotype blocks for samples

MHBs are defined as regions in which a minimum of 2 adjacent CpG sites exhibit a linkage disequilibrium measurement $r^2$ greater than 0.5 (10). In a previous study (10), 147 888 MHBs were defined based on the published DNA methylation datasets including human primary tissues, the

| Variable                           | FTC (n = 46) | FA (n = 46) | TT-UMP (n = 36) |
|------------------------------------|-------------|------------|----------------|
| Age, years (range)                 | 51.5 (12-77) | 46 (13-69) | 43.5 (24-67)   |
| Tumor size, cm (range)             | 2.5 (0.5-11) | 2.5 (0.4-6) | 2.4 (0.5-10.5) |
| Sex                                |             |            |                |
| Male                               | 14          | 15         | 11             |
| Female                             | 32          | 31         | 25             |
| Hürthle cell change                |             |            |                |
| With Hürthle cell change (<75%)    | 11          | 10         | 4              |
| Without Hürthle cell change        | 35          | 36         | 32             |
| Histologic type                    |             |            |                |
| Minimally invasive                 | 28          | \          | \              |
| Encapsulated angioinvasive         | 8           | \          | \              |
| Widely invasive                    | 10          | \          | \              |
| Disease recurrence                 |             |            |                |
| Yes                                | 5           | 0          | 0              |
| No                                 | 38          | 44         | 29             |
| Unknown                            | 3           | 2          | 7              |
| Oncogenic gene alteration          |             |            |                |
| NRAS mutation                      | 9           | 3          | 8              |
| TERTp mutation                     | 13          | 0          | 0              |
| HRAS mutation                      | 4           | 4          | 2              |
| PTEN mutation                      | 2           | 3          | 0              |
| TSHR mutation                      | 0           | 3          | 1              |
| KRAS mutation                      | 1           | 0          | 3              |
| TP53 mutation                      | 0           | 0          | 3              |
| PPARG rearrangement                | 1           | 0          | 1              |
| BRAF mutation                      | 1           | 0          | 0              |

Table 1. Characteristics of 128 patients with follicular thyroid tumors
H1 human embryonic stem cells, in vitro–derived progenitors and human cancer cell lines. We screened candidate MHBs in the FA and FTC samples by searching these previously defined MHBs in the sequencing data of each sample. The candidate MHBs for the subsequent analysis were required to be detected in no fewer than 90% of the samples.

Quantifying the methylation status of MHB
MHBs were quantified by metrics of the proportion of discordant reads (PDRs) and the methylated haplotype load (MHL), for which the calculation formulae are presented in Fig. 1. PDR is a measurement of a process that stochastically increases variation in methylation, which is considered as a characteristic of the cancer epigenome (15). MHL is the weighted mean of the fraction of fully methylated/unmethylated haplotypes and their substrings at different lengths, that is capable of distinguishing MHBs with the same average levels of methylation but various degrees of coordinated methylation (10). According to available literature (10), parameters of MHL like $P(MH_i)/P(uMH_i)$ (the fraction of fully methylated/unmethylated CpGs with $i$ loci) and $w_i$ (the weight for $i$ locus haplotype) were considered optional in different investigations. For an MHB of length $l$, all substrings with a length from $1$ to $l$ were considered for those calculations. In this study, we have tried various combinations of $P(MH_i)/P(uMH_i)$ and $w_i$ (equal to $i$, $i^2$, or $i^3$) for the MHL calculation of each MHB. PDR and MHL of each MHB were compared between the FA and FTC samples of the training cohort. MHBs with a PDR or MHL standard deviation of no less than 0.02 and an FDR less than 0.05 (Wilcox rank sum test) were selected as MHB markers. Finally, 4 types of MHB markers were identified (Fig. 1): MHBs quantified by PDR were indicated as pdr-based markers; MHBs quantified by MHL, with $w_i$ equal to $i$, were indicated as mhl-based markers; $w_i$ was considered to be equal to $i^2$ when calculating mhl2-based markers; and $P(MH_i)$ was replaced by $P(uMH_i)$ when calculating unmhl-based markers. A heatmap based on the selected MHB markers was generated using Morpheus, by applying hierarchical clustering and using the Euclidean distance as a similarity measure.

A random forest model was built and tested by cross-validation of the data (R package “caret,” 3-fold, 10 repeats). The model possessing the highest accuracy was selected for further independent validation. Furthermore, Gene Ontology (GO) enrichment analyses were conducted in WebGestalt (http://www.webgestalt.org) using overrepresentation enrichment analysis method.

DNA variation
For the DNA variation screening, targeted next generation sequencing was performed using the Next-seq500 sequencer (Illumina, Inc., CA, USA), as previously reported (16). Briefly, all the samples were sequenced using a custom-designed panel (Agilent) targeting DNA alteration of 14 thyroid tumor–related genes. The panel covered selected exons, introns or promoters of BRAF, CTNNB1, GNAS, HRAS, KRAS, NRAS, PTEN, PIK3CA, TERT, TP53, TSHR, and RET to detect any mutations. Additionally, this panel also identifies gene rearrangements of RET, PPAR, and ALK. Mutations classified as tolerated, benign, and neutral via in silico algorithms, or predicted as neutral in the COSMIC database, were excluded from further analysis.

Statistical analysis
Fisher’s exact test was used to analyze the differences in gene alteration across the TT-UMP samples with high and low risk, as evaluated by methylation score. $P < .05$ was considered statistically significant.

Results
DNA methylation status based on CpG dinucleotides
A total of 8294 DMCs and 7723 DMCs unique to the FTC and FA samples, respectively, against NT controls were identified during screening by RRBS. Within these DMCs, 3872 CpGs were hypermethylated in both the FTC and FA samples, while 221 CpGs were hypomethylated in both the samples. These were excluded from downstream analyses, which left 4201 FTC-specific DMCs (3710 hypermethylated and 491 hypomethylated), and 3630 FA-specific DMCs (3536 hypermethylated and 94 hypomethylated) as shown in (14). Unsupervised hierarchical clustering of samples using FTC-specific DMCs, as well as FTC-specific combined FA-specific DMCs identified 2 main groups. However, both groups consisted of FTC as well as FA samples (Fig. 1 (17)).

Candidate methylation biomarkers of FTC
In this study, 46 960 to 80 681 (mean = 69 626) MHBs were identified for each FA or FTC sample. The candidate MHBs were required to be detected in no less than 90% of the samples. Finally, a total of 47 900 MHBs were screened to identify differential MHBs between FTC and FA samples.

Each MHB-based marker consisted of an MHB region and a score, which was assigned by one of the several methods that quantify MHB (see “Patients and Methods”). These resulted in the identification of 2 or more markers from a single MHB. A total of 70 MHB-based markers from 64 MHBs were identified, with significant different methylation
values (FDR <0.05) between the 33 FTC and 33 FA samples. Details of all the 70 markers have been listed in (14). Twenty-two of them were pdr-based markers; of the rest, 48 MHBs consisted of 5 mhl-based markers, 33 mhl3-based markers, and 10 umhl-based markers. No association was observed between the MHB markers and other clinical feathers (like MI/EA/WI, or with/without Hürthle cell change) of our cohort (data not shown). These MHB markers were able to cluster the 33 FTC samples, 33 FA samples, and 15 normal samples in an unsupervised manner (Fig. 1).

Additionally, 9 of them are among the 18 genes abnormally methylated in FTC; for the rest, 7 are located in 6 exons, 6 in 6 introns, 2 in 2 3′-untranslated regions, and the remaining 21 lie within intergenic regions (Fig. 2A).

GO analyses showed that among the 60 marker-associated genes, 11 were enriched in 6 categories of molecular functions. Among them, 4 were related to “RNA Polymerase II Transcription Factor Activity, sequence specific DNA Binding,” and the remaining 2 were related to “Transcription Regulatory Region Sequencing specific DNA Binding” (Fig. 2B, and (14)). Nine out of the 11 gene regions related to GO terms were hypermethylated and 2 were hypomethylated in the FTC compared with the FA samples.

A random forest model of 500 trees was constructed to distinguish the FTC from the FA samples based on the 70 DNA methylation markers identified in the training cohort.
A score of 0.5 for the methylation markers was most accurate in distinguishing benign from malignant samples. The importance value (Gini index) of each marker is listed in (14).

Validation of DNA methylation biomarkers in an independent cohort

A total of 26 samples including, 13 FTC and 13 FA tissues, were used to test the diagnostic accuracy of the DNA methylation-based FTC predicting model. When compared with the matched pathological features, only 1 sample’s prediction result by this model at a cutoff of 0.5 was inconsistent with the result from its pathological diagnosis (Fig. 3A), The methylation score of the exceptional specimen was 0.462 (14), which was slightly smaller than the cutoff value. Overall, this model achieved an area under the curve value of 0.994, with an estimated specificity of 100% and sensitivity of 92.3% (Fig. 3B) in the validation cohort. These results indicate that the FTC-specific DNA methylation markers were highly consistent with clinicopathological characteristics, which strongly confirmed their applicability in diagnosing FTC specimens.

Methylation pattern in patients with TT-UMP

Additionally, in order to further confirm the diagnosis potential and subject range of the FTC-predicting model, 32

Table 2. DNA methylation markers located in promoter regions

| Marker | Gene | Tumor-related function | Reference |
|--------|------|------------------------|-----------|
| chr10:102822370:102822647_pdr | KAZALD1 | Epigenetic silencing confers a better prognosis | (32) |
| chr10:124896784:124897154_mhl3 | HMX3<sup>b</sup> | / | |
| chr13:92051635:92051674_mhl3 | GPC5 | A tumor suppressor | (33) |
| chr13:92051635:92051674_mhl | SIX6<sup>b</sup> | Involved in the tumor initiation and progression | (34) |
| chr14:60977845:60977865_mhl3 | WNK4<sup>b</sup> | Associated with tumorigenesis | (35) |
| chr14:60977845:60977865_mhl | HOXB3<sup>b</sup> | Promotes cancer cell progression | (36) |
| chr17:19747115:19747127_mhl3 | GATA6-AS1 | Overexpression inhibits cancer progression | (37) |
| chr18:19747115:19747127_mhl | CBLN2 | Mutated in cancer cells | (38) |
| chr2:105474371:105474381_mhl3 | POU3F3 | Promotes cancer cell proliferation | (39) |
| chr2:80350770:80350826_mhl3 | LRRT1M | Increased methylation in carcinoma | (40) |
| chr4:40632502:40632519_mhl3 | RBM47 | Suppress tumor growth | (41) |
| chr7:158798629:158798674_mhl3 | LINCO0689 | Promotes cell growth and metastasis | (42) |
| chr7:27182264:27183525_mhl3 | HOXA5<sup>b</sup> | A tumor suppressor | (43) |
| chr9:139740669:139740676_mhl3 | AJM1 | / | |
| chr15:48704025:48705556_umhl | MYEFT2 | / | |
| chr15:60883371:60883395_umhl | RORA<sup>b</sup> | A tumor suppressor | (44) |
| chr22:38071168:38071189_umhl | LGALS1 | Contributes to the tumor development | (45) |
| chr10:102822370:102822647_umhl | KAZALD1 | Epigenetic silencing confers a better prognosis | (32) |
| chr10:105233375:105233477_pdr | NEURL1 | Promoter methylated in carcinoma | (46) |
| chr12:50350585:50350570_pdr | AQP5 | Upregulates cancer proliferation | (47) |
| chr13:32605660:32605843_pdr | FRY | Plays a role in cancer progression | (48) |
| chr13:77459521:77459792_pdr | KCTD12<sup>a</sup> | Induces cancer cell proliferation | (49) |
| chr17:80009015:80009025_pdr | RFNG | Mutated in carcinoma | (50) |
| chr19:12831808:12832195_pdr | TNPO2 | Promotes cancer cell proliferation | (51) |
| chr19:13213485:13213513_pdr | LYL1<sup>b</sup> | Correlated with upregulation of cancer-related pathways | (52) |
| chr19:13213644:13213814_pdr | EPHX3 | Increased methylation in carcinoma | (40) |
| chr19:86746748:86747497_pdr | ADAMTS10 | Binds to tumor suppressor protein | (53) |
| chr3:194208192:194208617_pdr | LINCO0884 | / | |
| chr6:31696240:31696334_pdr | DDAH2<sup>b</sup> | Promotes tumor angiogenesis | (54) |

<sup>a</sup>These coordinates were based on UCSC hg19 reference.
<sup>b</sup>MHB marker–related genes that were reported previously have been indicated.
TT-UMP samples were independently tested to evaluate their potential of malignancy. Based on the scores of the model, all the tested TT-UMP samples were grouped into 2 categories: (1) 12 in the low-risk group, whose scores ranged from 0 to 0.5; and (2) 24 in the high-risk group (0.5-1). When coanalyzed with their gene alteration status (Table 3), only 1 case carrying deleterious variation was identified in the low risk TT-UMP samples (1/12, 8.3%).

Figure 2. Genomic locations of identified MHBs. (A) The distribution of MHB markers in the genome. Markers located on 3′ untranslated region were merged into the exon group. (B) Gene Ontology (GO) terms enrichment by genes associated with MHBs. Number of genes in each group is indicated.

Figure 3. Prediction of the validation cohort based on the DNA methylation biomarkers and model. (A) Cancer risk scores for the thyroid nodule specimens in the validation cohort. The threshold of cancer risk score is 0.5. (B) A ROC curve of the methylation-based malignancy predicting model on the validation cohort. The sensitivity, specificity, and area under the curve with the 95% confidence interval values have been presented.
Table 3. Oncogenic gene alterations of the TT-UMP samples

| Sample ID | Methylation score | Gene alterations |
|-----------|-------------------|------------------|
| UMP1      | 0.426             | NRAS Q61R        |
| UMP2      | 0.504             | NRAS Q61R        |
| UMP3      | 0.534             | NRAS Q61R        |
| UMP4      | 0.548             | HRAS Q61R        |
| UMP5      | 0.584             | TP53 M246I       |
| UMP6      | 0.614             | TSHR D633H       |
| UMP7      | 0.638             | NRAS Q61R        |
| UMP8      | 0.640             | PPARG-PAX8 fusion|
| UMP9      | 0.680             | KRAS Q61R        |
| UMP10     | 0.686             | NRAS Q61R        |
| UMP11     | 0.692             | HRAS Q61K        |
| UMP12     | 0.692             | KRAS G12D        |
| UMP13     | 0.702             | NRAS Q61K        |
| UMP14     | 0.774             | KRAS G13D; TP53 H214R |
| UMP15     | 0.812             | NRAS Q61K        |
| UMP16     | 0.812             | NRAS Q61R; TP53 C141W; TP53 Y220C |

Figure 4. Correlation between the methylation-based diagnostic scores and gene alteration status in 36 TT-UMP samples. Hollow circles indicate the TT-UMP samples without pathologic gene alteration; solid circles indicate the samples with deleterious gene alterations identified. Asterisk (*) indicates a significant difference with P < .05.

and 15 cases with gene alterations in high risk cases (15/24, 62.5%) (Fig. 4). The ratio of gene alteration in the high-risk group was significantly higher than that in the low-risk group (P = .004).

Discussion

DNA methylation plays important roles in human cancer development by regulating gene expression, or inducing genomic instability which promotes tumor progression (2). Methylation analysis based on CpG dinucleotides primarily predicts the gene expression status. However, in a previous research, 7.9% (93/1172) of the DMCs-associated genes were differentially expressed in the FTC, compared to control samples (11). Moreover, only 2.9% (12416) of the hypermethylated genes were downregulated and 1.3% (1/77) of the hypomethylated genes were upregulated in the FTC samples in another study (18). It was also reported that only 1.5% of the genes hypermethylated in FTC were downregulated, and no association was found between the hypomethylated genes and changes in gene expression (19).

In our study, we did not observe differential expression of any of the identified MHB-associated genes in the FTC samples, from either the public datasets or previous publications. This could be attributed to the relatively small number of genes. Contrarily, it may also suggest that DNA methylation variations have little direct and/or robust effect on the gene expression. Our study did not show satisfactory CpG-based clustering of the FTC and FA samples. This phenomenon was also observed in other single CpG site–based analyses (13, 20). Methylation of the genes TPO and UCHL1 discriminated the FTC from the FA samples with a low accuracy (receiver operating characteristic [ROC] = 0.607, sensitivity 0.78%) (13). Three CpGs (cg10705422, cg17707274, and cg26849382) differentiated nonmalignant (FA, Hürthle cell adenoma, and noninvasive follicular thyroid neoplasm with papillary-like nuclear features) tumors from thyroid cancers, with ROC curve of 0.83, 0.83, and 0.80, respectively (20). Moreover, it has been reported that MHB can capture discrete entities of epigenetic regulation in the human genome, thus improving the effectiveness of DNA methylation analysis based on single-CpG sites (10, 21, 22). However, MHB-based methylation profiles were never analyzed in previous investigations of thyroid cancer.

In this study, we have profiled changes in the DNA methylation patterns in FTC, using RRBS method and MHB-based analysis. Based on the training cohort, 70 MHB markers were identified for discriminating between the FA and FTC samples at a high accuracy (ROC = 0.994). Among the 60 genes that were associated with the identified MHB markers, 18 have been previously reported to show an abnormal methylation status in FTC samples (12, 18). However, the trends of methylation change for these markers were not always in the same direction. For example, cg18055007 located in the promoter of gene DDAH2 was reported as a FA-specific marker that is hypermethylated compared to control samples (12), whereas it was reported to be also hypermethylated in FTC (18). In our study, methylation status of this gene is recorded as a pdr-based marker which has a lower score in the FTC, as compared to the FA samples. PDR was calculated as a proportion of the discordant reads in the total reads of a given region.
Based on this concept, we can suppose that a portion of CpGs (excluding cg18055007) in this region were methylated in the normal cells, exhibiting a high score for this pdr-based marker. In the benign lesions, the methylation level was accumulated, and the proportion of methylated CpG (partly including cg18055007) was increased, which resulted in hypermethylation of this site, in conjunction with a high score for this pdr-based marker. Lastly, almost all the CpGs in this region, including cg18055007, were methylated in the malignant tumor, which resulted in a decrease in the pdr score of this MHB marker.

It is well established that the inactivation of certain tumor suppressor genes by hypermethylation in their promoters occurs in multiple cancer types (2). In our study, about half of the identified DNA methylation markers were located in promoters. Most of the genes associated with these markers were reported to possess roles in tumorogenesis. In the GO analysis, all the enriched terms of these markers were related to “sequence specific DNA binding.” Collectively, these evidences suggest that the inactivation of DNA binding by transcriptional regulators through promoter hypermethylation may play a role in FTC development, which is a mechanism that has been reported to cause abnormal gene expression in many tumors (23).

The methylation prediction model based on the MHB markers had displayed robust classification for distinguishing FTC from FA with a much higher sensitivity (92.3%) and specificity (100%) than other molecular markers, like gene mutation (24) and mRNA expression (9), both of which have shown less than 90% of either sensitivity or specificity. The presence of specific mutations like TERTp can be a useful tool for predicting malignancy (25). TERTp testing is quick and cost effective, even for a large number of patients. However, it remains difficult to distinguish many thyroid tumors without specific mutations from benign nodules. For those patients, the identified DNA methylation markers may be highly useful for refining the mutation-driven classifications.

The difficulty in determining the malignancy of TT-UMP, which is a heterogeneous group of tumors with unspecified, borderline and uncertain behavior, arises from the fact that these tumors pathologically present suspicious capsular/vascular invasion or with an ambiguous PTC nuclear change (26). Additionally, the pathological assessment criteria of thyroid lesions are prone to subjectively and display interobserver variability, even among acknowledged experts, which adds to the overall uncertainty of diagnosing those lesions (27). In recent years, screening for TERTp mutations has been routinely used to clinically predict malignant potential in follicular thyroid tumors (25). However, TERTp mutations only occurred in approximately 10% to 35% FTC patients and is quite rare in TT-UMP (25, 28). On the other hand, many genetic alterations in genes like HRAS, NRAS, and RET are frequently observed in both FTC and FA (5, 29). Alterations in these drive genes appear to lack sensitivity and specificity for FTC, which may be limited in suggesting the risk of TT-UMP lesions. In our investigation of TT-UMP, we found that high methylation scores showed more concordance with a high frequency of gene variations, which is indicative of increased risk of malignancy, compared with those with a low frequency of mutations. This result implied that DNA MHB markers can be used to assess the malignant potential of TT-UMP. However, the clinical significance of this result should be further investigated in more patients with TT-UMP lesions and over a longer follow-up period. In the present study, 3 TT-UMP with deleterious TP53 mutations were all in the high-risk group. Moreover, a strong correlation between TP53 mutations and DNA methylation has been reported in previous studies (30, 31), further indicating the significance of these markers. Moreover, 2 TT-UMP with TP53 “pathogenic mutation” (p.His214Arg, p.Tyr220Cys) presented much higher methylation scores than those with p.Met246Ile, suggesting that different genetic variations of TP53 display distinct capacities for modulating DNA methylation (30).

In summary, 70 DNA methylation markers were identified based on MHB analysis that showed significant differential methylation scores between FTC and FA samples. A random forest model based on these DNA methylation markers was constructed to discriminate malignant FTC from benign FA. This model can also be used to predict the malignant potential of TT-UMP thyroid nodules. For the patients lacking any aggressive gene alterations and possessing thyroid tumors that are difficult to diagnose pathologically, DNA methylation markers of those samples may be highly instrumental for the clinical judgement.

Acknowledgments

Financial Support: This work was supported by Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences (2016-I2M-1-002, 2017-I2M-1-001).

Author Contributions: H.Z. designed the study and wrote the manuscript. Z.Z. and T.X. analyzed the data and wrote the manuscript. X.L. and H.D. contributed to the experiments. Q.H. and Z.S. contributed to the data collection. H.W. and Z.L. are the guarantors of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Additional Information

Correspondence: Zhiyong Liang, PhD, Chinese Academy of Medical Sciences and Peking Union Medical College Hospital, Beijing, China. Email: liangzhiyong1220@yahoo.com; or Huanwen Wu,
References

1. Lloyd RV, Osamura RY, Kloppel G, et al. eds. *WHO Classification of Tumours of Endocrine Organs*. 4th ed. International Agency for Research on Cancer; 2017.

2. Baloch ZW, LiVolsi VA. Our approach to follicular-patterned lesions of the thyroid. *J Clin Pathol*. 2007;60(3):244-250.

3. McHenry CR, Phitayakorn R. Follicular adenoma and carcinoma of the thyroid gland. *Oncologist*. 2011;16(5):585-593.

4. Duan H, Liu X, Ren X, Zhang H, Wu H, Liang Z. Mutation profiles of follicular thyroid tumors by targeted sequencing. *Diagn Pathol*. 2019;14(1):39.

5. Labourier E, Shifrin A, Busseniers AE, et al. Molecular testing for miRNA, mRNA, and DNA on fine-needle aspiration improves the preoperative diagnosis of thyroid nodules with indeterminate cytology. *J Clin Endocrinol Metab*. 2015;100(7):2743-2750.

6. Le Mercier M, D’Haene N, De Nève N, et al. Next-generation sequencing improves the diagnosis of thyroid FNA specimens with indeterminate cytology. *Histopathology*. 2015;66(2):215-224.

7. Nikiforova MN, Lynch RA, Biddinger PW, et al. RAS point mutations and PAX8-PPAR gamma rearrangement in thyroid tumors: evidence for distinct molecular pathways in thyroid follicular carcinoma. *J Clin Endocrinol Metab*. 2003;88(5):2318-2326.

8. Dom G, Frank S, Floor S, et al. Thyroid follicular adenomas and carcinomas: molecular profiling provides evidence for a continuous evolution. *Oncotarget*. 2018;9(12):10343-10359.

9. Guo S, Diep D, Plongthongkum N, Fung HL, Zhang K, Zhang K. Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. *Nat Genet*. 2017;49(4):635-642.

10. Affinoto O, Salerno P, D’Alessio A, et al. Association between DNA methylation profile and malignancy in follicular-patterned thyroid neoplasms. *Endocr Relat Cancer*. 2019;26(4):451-462.

11. Mancikova V, Buj R, Castelblanco E, et al. DNA methylation profiling of well-differentiated thyroid cancer uncovers markers of recurrence free survival. *Int J Cancer*. 2014;135(3):598-610.

12. Stephen JK, Chen KM, Herritt J, Chitale D, Divine G, Worsham MJ. Methylation markers differentiate thyroid cancer from benign nodules. *J Endocrinol Invest*. 2018;41(2):163-170.

13. Zhang H, Zhang Z, Liu X, et al. Data from: supplementary data for “DNA methylation haplotype block markers efficiently discriminate follicular thyroid carcinoma from follicular adenoma”. Deposited August 31, 2020. ProMED-mail website. [http://doi.org/10.6084/m9.figshare.12895469](http://doi.org/10.6084/m9.figshare.12895469).

14. Landau DA, Clement K, Ziller MJ, et al. Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. *Cancer Cell*. 2014;26(6):813-825.

15. Duan H, Li Y, Hu P, et al. Mutational profiling of poorly differentiated and anaplastic thyroid carcinoma by the use of targeted next-generation sequencing. *Histopathology*. 2019;75(6):890-899.

16. Zhang H, Zhang Z, Liu X, et al. Data from: supplementary figure 1 for “DNA methylation haplotype block markers efficiently discriminate follicular thyroid carcinoma from follicular adenoma”. Deposited August 31, 2020. ProMED-mail website. [http://doi.org/10.6084/m9.figshare.12895439](http://doi.org/10.6084/m9.figshare.12895439).

17. Bisarro Dos Reis M, Barros-Filho MC, Marchi FA, et al. Prognostic classifier based on genome-wide DNA methylation profiling in well-differentiated thyroid tumors. *J Clin Endocrinol Metab*. 2017;102(11):4089-4099.

18. Dom G, Frank S, Floor S, et al. Thyroid follicular adenomas and carcinomas: molecular pathology of sperm DNA methylation and their potential implications in cattle. *BMC Genomics*. 2019;20(1):888.

19. Acton RJ, Bell CG. Cancer detection and tissue of origin determination with novel annotation and scoring of cell-free methylated DNA. *AME Med J*. 2017;2:110.

20. Poma AM, Giannini R, Piaggi P, et al. A six-gene panel to label follicular adenoma, low- and high-risk follicular thyroid carcinoma. *Endocr Connect*. 2018;7(1):124-132.

21. Baser H, Topaloglu O, Tam AA, et al. Comparing clinicopathologic and radiographic findings between TT-UMP, classical, and non-encapsulated follicular variants of papillary thyroid carcinomas. *Endocr Pathol*. 2016;27(3):233-242.

22. Duggal R, Rajwanshi A, Gupta N, Vasishta RK. Interobserver variability amongst cytopathologists and histopathologists in the diagnosis of neoplastic follicular patterned lesions of thyroid. *Diagn Cytopathol*. 2011;39(4):235-241.

23. Baser H, Topaloglu O, Tam AA, et al. Comparing clinicopathologic and radiographic findings between TT-UMP, classical, and non-encapsulated follicular variants of papillary thyroid carcinomas. *Endocr Pathol*. 2016;27(3):233-242.

24. Duggal R, Rajwanshi A, Gupta N, Vasishta RK. Interobserver variability amongst cytopathologists and histopathologists in the diagnosis of neoplastic follicular patterned lesions of thyroid. *Diagn Cytopathol*. 2011;39(4):235-241.

25. Acquaviva G, Visani M, Repaci A, et al. Molecular pathology of thyroid tumours of follicular cells: a review of genetic alterations and their clinicopathological relevance. *Histopathology*. 2018;72(1):6-31.

26. Borowczyk M, Szczepanek-Parulska E, Debicki S, et al. Differences in mutational profile between follicular thyroid carcinoma and follicular thyroid adenoma identified using next generation sequencing. *Int J Mol Sci*. 2019;20(13):3126.

27. Rebbani K, Marchio A, Ezzikouri S, et al. TP53 R72P polymorphism modulates DNA methylation in hepatocellular carcinoma. *Mol Cancer*. 2015;14:74.

28. Chen YC, Gotea V, Margolin G, Elnitski L. Significant associations between driver gene mutations and DNA methylation alterations across many cancer types. *PLoS Comput Biol*. 2017;13(11):e1005840.
32. Wang H, Feng Y, Bao Z, et al. Epigenetic silencing of KAZALD1 confers a better prognosis and is associated with malignant transformation/progression in glioma. Oncol Rep. 2013;30(5):2089-2096.
33. Yuan S, Yu Z, Liu Q, et al. GPC5, a novel epigenetically silenced tumor suppressor, inhibits tumor growth by suppressing Wnt/β-catenin signaling in lung adenocarcinoma. Oncogene. 2016;35(47):6120-6131.
34. Liu Q, Li A, Tian Y, et al. The expression profile and clinic significance of the SIX family in non-small cell lung cancer. J Hematol Oncol. 2016;9(1):119.
35. Moniz S, Jordan P. Emerging roles for WNK kinases in cancer. Cell Mol Life Sci. 2010;67(8):1265-1276.
36. Chen J, Zhu S, Jiang N, Shang Z, Quan C, Niu Y. HoxB3 promotes prostate cancer cell progression by transactivating CDCA3. Cancer Lett. 2013;330(2):217-224.
37. Li ZT, Zhang X, Wang DW, et al. Overexpressed lncRNA GATA6-AS1 inhibits LNM and EMT via FZD4 through the Wnt/β-catenin signaling pathway in GC. Mol Ther Nucleic Acids. 2020;19:827-840.
38. Wang F, Liu DB, Zhao Q, et al. The genomic landscape of small cell carcinoma of the esophagus. Cell Mol Life Sci. 2018;28(7):771-774.
39. Wan X, Xiang J, Zhang Q, Bian C. Long noncoding RNA POU3F3 promotes cancer cell proliferation by upregulating rho-associated protein kinase 1. J Cell Biochem. 2019;120(5):8195-8200.
40. Morandi L, Gissi D, Tarsitano A, et al. CpG location and methylation level are crucial factors for the early detection of oral squamous cell carcinoma in brushing samples using bisulfite sequencing of a 13-gene panel. Clin Epigenetics. 2017;9:85.
41. Sakurai T, Isogaya K, Sakai S, et al. RNA-binding motif protein 47 inhibits Nrf2 activity to suppress tumor growth in prostate carcinoma. Oncogene. 2016;35(38):5000-5009.
42. Liu X, Zhu Q, Guo Y, Xiao Z, Hu L, Xu Q. LncRNA LINCO0689 promotes the growth, metastasis and glycolysis of glioma cells by targeting miR-338-3p/PKM2 axis. Biomed Pharmacother. 2019;117:109069.
43. Peng X, Zha L, Chen A, Wang Z. HOXA5 is a tumor suppressor gene that is decreased in gastric cancer. Oncol Rep. 2018;40(3):1317-1329.
44. Sun X, Dongol S, Qiu C, et al. miR-652 promotes tumor proliferation and metastasis by targeting RORA in endometrial cancer. Mol Cancer Res. 2018;16(12):1927-1939.
45. Chou FC, Chen HY, Kuo CC, Sytwu HK. Role of galectins in tumors and in clinical immunotherapy. Int J Mol Sci. 2018;19(2):430.
46. van Vlodrop IJH, Joosten SC, De Meyer T, et al. A four-gene promoter methylation marker panel consisting of GREM1, NEURL, LAD1, and NEFH predicts survival of clear cell renal cell cancer patients. Clin Cancer Res. 2017;23(8):2066-2018.
47. Lee SJ, Kang BW, Kim JG, et al. AQP5 variants affect tumoral expression of AQP5 and survival in patients with early breast cancer. Oncology. 2017;92(3):153-160.
48. Ren X, Graham JC, Jing L, et al. Mapping of Mcs30, a new mammary carcinoma susceptibility quantitative trait locus (QTL30) on rat chromosome 12: identification of Ify as a candidate Mcs gene. PLoS One. 2013;8(9):e70930.
49. Zhong Y, Yang J, Xu WW, et al. KCTD12 promotes tumorigenesis by facilitating CDC25B/CDK1/Aurora A-dependent G2/M transition. Oncogene. 2017;36(44):6177-6189.
50. Liu K, Ren Y, Pang L, et al. Papillary renal cell carcinoma: a clinicopathological and whole-genome exon sequencing study. Int J Clin Exp Pathol. 2015;8(7):8311-8335.
51. Gong L, Wen T, Li Z, et al. TNPO2 operates downstream of DYNC1H1 and promotes gastric cancer cell proliferation and inhibits apoptosis. Cancer Med. 2019;8(17):7299-7312.
52. Kim SI, Lee JW, Lee N, et al. LYL1 gene amplification predicts poor survival of patients with uterine corpus endometrial carcinoma: analysis of the Cancer genome atlas data. BMC Cancer. 2018;18(1):494.
53. Matsuzech A, Kitayama H, Omura A, Nishimoto E, Alexander DB, Noda M. The RECK tumor-suppressor protein binds and stabilizes ADAMTS10. Biol Open. 2018;7(10):1-10.
54. Shiozawa T, Iyama S, Toshima S, et al. Dimethylarginine dimethylaminohydrolase 2 promotes tumor angiogenesis in lung adenocarcinoma. Virchows Arch. 2016;468(2):179-190.