

<https://doi.org/10.1038/s41698-025-00932-7>

Epigenomic profiling of papillary thyroid carcinoma reveals distinct subtypes with clinical implications



Yeongun Lee^{1,4}, Hyo Jin Park^{1,4}, Jin Seok Lee², Ho Jung Jeong², Su Min Kim¹, Minjeong Cho¹, Wonjin Woo¹, So Hee Dho¹, Seok-Mo Kim²✉ & Lark Kyun Kim^{1,3}✉

Papillary thyroid carcinoma (PTC) is the most prevalent form of thyroid cancer with generally favorable outcomes. However, surgeons often face challenges regarding optimal surgical timing, extent of surgery, and identifying patients at risk for metastasis or progression to more aggressive subtypes. The ongoing debate over immediate surgery versus active surveillance emphasizes the need for reliable, minimally invasive diagnostic tools to inform surgical decision-making. This study aims to develop an epigenetic biomarker-based prediction system using fine-needle aspiration biopsy (FNAB) samples to assess PTC aggressiveness preoperatively. We conducted a comprehensive analysis of methylome data to identify approximately 7200 CpG islands with altered methylation levels in thyroid cancer tissues. These candidate regions were further examined in our cohort of 55 PTC patients to develop methylation-specific primers suitable for FNAB samples. Methylation patterns allowed us to stratify patients into two distinct prognostic groups, one of which exhibited a poorer survival rate. Our methylation-specific primers effectively classified FNAB samples into these groups, demonstrating their potential as a preoperative tool for assessing tumor aggressiveness. This stratification aids in informing surgical planning and personalizing treatment strategies. DNA methylation profiling of PTC identifies key epigenetic biomarkers associated with tumor aggressiveness. Utilizing these biomarkers in FNAB samples provides a minimally invasive method for preoperative risk assessment, assisting surgeons in tailoring surgical interventions and potentially improving patient outcomes.

Papillary thyroid carcinoma (PTC) is the most prevalent subtype of thyroid cancer, and its early detection through ultrasound screening results in improved survival rates^{1–4}. However, despite its typically indolent nature, concerns persist regarding patient anxiety, optimal timing for surgery, metastasis, and potential progression to aggressive cancers, such as anaplastic thyroid cancer. The ongoing debate surrounding the necessity of surgery highlights the need to develop molecular-based early diagnostic markers to address these concerns. Although conventional clinicopathological indicators offer some insights, they remain insufficient for predicting PTC prognosis, emphasizing the urgency for novel prognostic biomarkers in PTC management⁵.

DNA methylation, a pivotal epigenetic mechanism that regulates gene expression, has emerged as a promising biomarker for cancer prognosis^{6,7}. A recent study by Grail Inc. on the development of a prognostic prediction system based on machine learning and DNA methylation using non-invasive liquid biopsy represents a significant advancement in this field⁸. It demonstrates high specificity and extremely low false-positive rates across various types of cancer. However, current research has predominantly focused on highly lethal cancers, with less emphasis on achieving high predictability for thyroid cancer.

Efforts to predict and diagnose diseases non-invasively using liquid biopsy have yielded significant advancements in recent years⁹. However, its effectiveness may not match that of direct tissue examination. In the case of

¹Department of Biomedical Sciences, Graduate School of Medical Science, Brain Korea 21 Project, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea. ²Department of Surgery, Thyroid Cancer Center, Gangnam Severance Hospital, Institute of Refractory Thyroid Cancer, Yonsei University College of Medicine, Seoul, Republic of Korea. ³Division of Biology, Pohang University of Science and Technology, Pohang, Republic of Korea. ⁴These authors contributed equally: Yeongun Lee, Hyo Jin Park.

✉ e-mail: SEOKMOKIM@yuhs.ac; LKKIM@yuhs.ac

thyroid cancer, tissues can be obtained through fine-needle aspiration biopsy (FNAB), which is highly regarded for its minimal invasiveness and provides a preoperative malignancy diagnosis with reduced patient discomfort^{10,11}. The discovery of highly specific biomarkers that enable robust prognostic prediction using FNAB samples alone holds great promise for clinical applications.

This study aimed to delineate PTC subtypes by identifying differentially methylated CpG islands using targeted bisulfite sequencing in a cohort of 55 patients with PTC and matched normal tissues. Through the analysis of methylation level changes derived from The Cancer Genome Atlas (TCGA) and public reduced-representative bisulfite sequencing (RRBS) data (GSE107738), distinct differentially methylated regions (DMRs) were identified^{12,13}. Extensive molecular characterization of PTC subgroups was conducted based on methylation levels, incorporating RNA sequencing (RNA-seq) and Assay for Transposase-Accessible Chromatin (ATAC) sequencing from public datasets (TCGA and GSE162515)^{12,14}; this facilitated the identification of seven candidate genes that exhibited significant differences in DNA methylation, RNA expression, and chromatin accessibility between PTC subgroups. A sophisticated quantitative methylation-specific PCR (qMSP) system was developed for the precise assessment of DNA methylation levels in these genes. The qMSP system effectively distinguished between PTC subtypes based on methylation variation. Finally, the feasibility of differentiating PTC samples using FNAB complemented by real-time PCR was explored.

Results

Identification of DMRs in PTC

In our previous study, we found that accurately measuring methylation changes at all CpG sites within selected CpG islands enabled the development of highly specific MSP primers¹⁵. To identify candidate CpG island regions, we analyzed DNA methylation microarray data from TCGA and public data (RRBS, GSE107738)^{12,13}, ultimately identifying 7217 CpG island regions of interest (Supplementary Fig. 1). To determine the TBS in these regions, we isolated thyroid tumor tissues and adjacent normal tissues from 55 Korean patients with PTC. Following sequencing, we assessed the data quality in terms of base sequence quality, alignment on the reference genome, and read depth, revealing a significantly higher read depth than other sequencing data (Supplementary Fig. 2). Using our selection workflow, we identified 333 DMRs (Supplementary Fig. 3), with 329 DMRs available for downstream analysis after excluding four that were unavailable in TCGA data.

Interestingly, K-means clustering revealed two distinct patient subgroups within our cohort and TCGA cohort (Fig. 1A and B). PCA using the DNA methylation levels of the 329 DMRs confirmed a clear division into the two subgroups in our cohort (Fig. 1C). To explore the correlation between the subgroups in our cohort and TCGA, we assessed the correlation coefficient of each subgroup using the DNA methylation levels of the 329 DMRs. This analysis revealed a high correlation coefficient of 0.863 between the PTC1 subgroups of each cohort and a coefficient of 0.442 between the PTC2 subgroups, suggesting the presence of corresponding subgroups within each cohort (Fig. 1D). Notably, the difference in the DNA methylation levels of DMRs between PTC1 and PTC2 was more pronounced in our cohort than in TCGA cohort. This discrepancy in the detection of DNA methylation differences could be attributed to the fact that our targeted bisulfite sequencing reads the methylation levels of every CpG site in the target regions of interest. In summary, heterogeneous subgroups of PTCs can be detected using targeted bisulfite sequencing.

Furthermore, we examined the clinical significance of the subgroups defined by DNA methylation levels of the DMRs. As our cohort was prospective, survival analysis between the two groups was not feasible. Therefore, we used clinical data from TCGA for survival analysis between the two groups. The Kaplan–Meier plot indicated that PTC1 had worse overall survival than PTC2 (Fig. 1E). To further assess the prognostic significance, we performed a multivariate Cox regression analysis incorporating key clinicopathological factors (Sex, TNM stage, TCGA methylation subgroups,

and TCGA-defined PTC1/PTC2 clusters). The results showed a trend toward worse survival outcomes in the PTC1 group (HR = 3.6, 95% CI: 0.91–14.1, $p = 0.068$), though statistical significance was not reached (Supplementary Fig. 4). Notably, despite this survival trend, there were no significant differences in clinicopathological characteristics between PTC1 and PTC2 in our cohort, except for age at diagnosis (Supplementary Table 1). In TCGA cohort, the RAS-like PTC class and RAS mutation were frequently found in PTC2 patients (Supplementary Table 2). The thyroid differentiation score, determined from the expression levels of 16 genes related to thyroid metabolism and function in a previous study¹², was significantly lower in PTC1 than in PTC2. Additionally, the ERK score, calculated using a 52-gene signature from a previous study¹⁶ to assess ERK (MAPK) activation, was significantly higher in PTC1 than in PTC2. ERK activity is tightly regulated in normal thyroid tissues; otherwise, it results in aberrant cell proliferation, differentiation, and apoptosis^{17,18} (Supplementary Table 2). This suggests that PTC1 may exhibit a more aggressive phenotype than PTC2. In summary, DNA methylation levels of the 329 identified DMRs allowed us to categorize the PTC cohort into two groups that exhibited prognostic and clinicopathological differences.

Characterization of DNA methylation patterns and functional significance

Before exploring the functional relevance of the selected 329 DMRs, we categorized them based on the average level of DNA methylation in the subgroups of our cohort. We calculated the average DNA methylation levels in the PTC1 and PTC2 subgroups, leading to the redefinition of DMRs into 245 hypermethylated and 84 hypomethylated regions in PTC1 compared to PTC2 (Fig. 2A). Annotation of DMR locations in the genome revealed a higher occurrence of DNA methylation in the promoters and distal intergenic regions of PTC1 (Fig. 2B). To investigate the functional roles of the genes associated with these DMRs, we conducted an ontology analysis of the annotated genes, uncovering a significant distribution of genes related to developmental processes in the hypermethylated DMRs of PTC1 (Fig. 2C). Furthermore, exploration of transcription factors associated with DMRs was performed using motif analysis^{19,20}, which showed the enrichment of binding motifs for components of the PRC1 and PRC2 complexes in DMRs (Fig. 2D). Visualization of the genome-wide location of the DMRs through coverage plot analysis revealed hypermethylation concentrated in the HOXA and HOXB clusters (Fig. 2E), consistent with observations in various cancers, including thyroid cancer, as demonstrated in previous studies^{21–23}. Taken together, our results demonstrated a distinct methylation pattern in PTC subgroups, characterized by significant hypermethylation of developmentally crucial genes. This finding suggests that epigenetic changes in PTC1 are associated with developmental processes and contribute to the understanding of the heterogeneity of PTC.

Transcriptomic profiling and differential gene expression analysis of PTC subgroups

To compare the transcriptomic profiles of the two subgroups and normal tissues, we analyzed RNA-seq data from TCGA. As expected, the PCA revealed a clear separation of subgroups based on principal components 1 and 2 (Fig. 3A). To investigate the DEGs between each subgroup, we employed DESeq2²⁴ and generated volcano plots to visualize the DEGs (Fig. 3B, Supplementary Fig. 5A and B). K-means clustering of all the DEGs categorized them into three clusters based on their expression patterns (Fig. 3C). DEG cluster 1 consisted of genes commonly downregulated in PTC1 and PTC2 compared with those in normal tissues, DEG cluster 2 comprised genes specifically upregulated in PTC1, and DEG cluster 3 included genes commonly upregulated in PTC1 and PTC2 compared with those in normal tissues. Consistent with findings from another study²⁵, ontology analysis revealed that DEG cluster 1 contained genes known to be downregulated in thyroid cancer, whereas DEG cluster 3 contained genes known to be upregulated in thyroid cancer. Interestingly, DEG cluster 2 showed enrichment of immune-related terms (Fig. 3D). Furthermore, GSEA

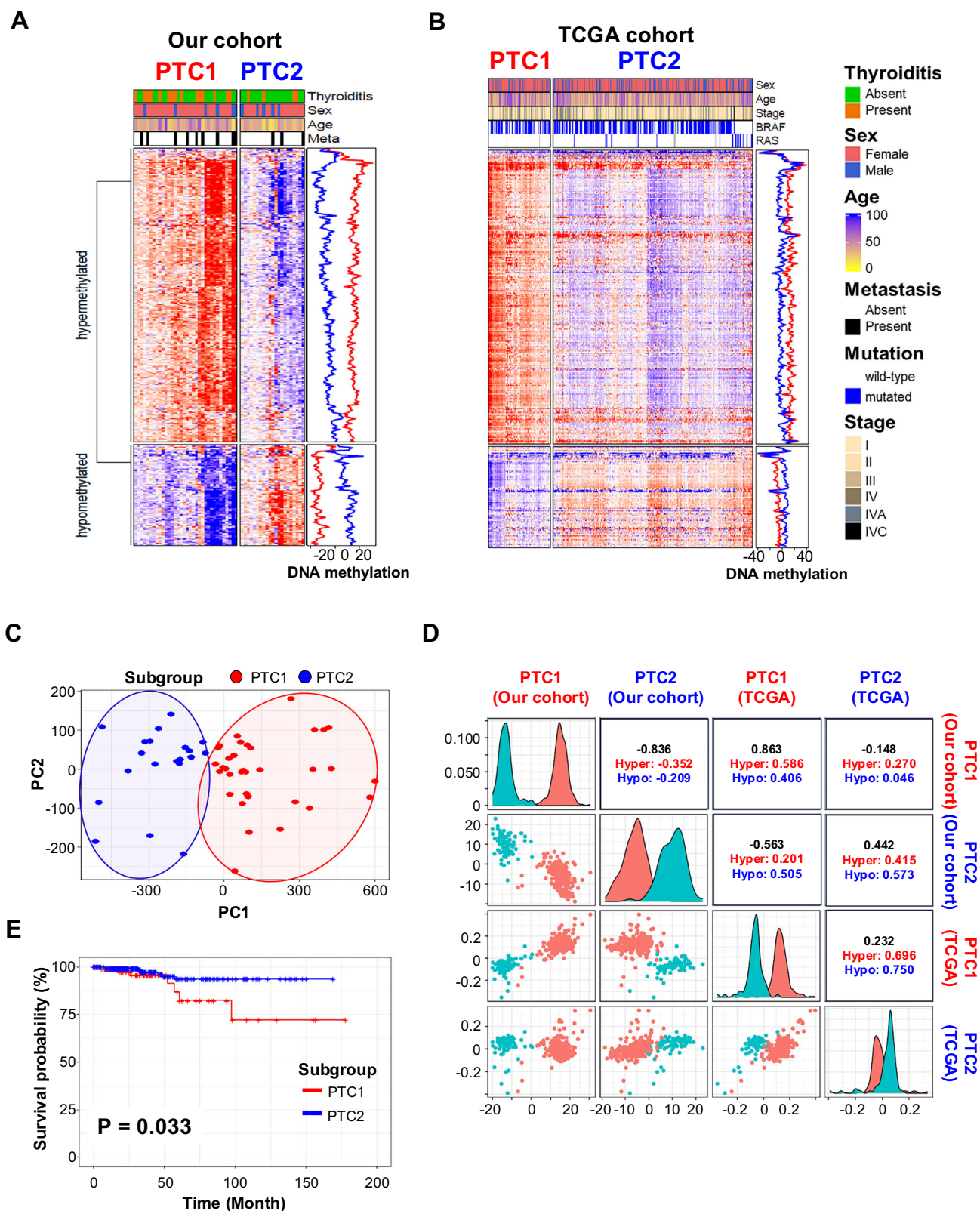


Fig. 1 | Stratification of PTC based on 329 DMRs. Heatmap of differential DNA methylation levels across 329 DMRs in our cohort (A) and TCGA cohort (B). The DNA methylation levels of each tumor sample were normalized by subtracting the DNA methylation levels of paired normal tissues. Hypermethylation and hypomethylation were defined based on the DNA methylation levels in each subgroup.

C PCA of our PTC cohort conducted with 329 DMRs. D Correlation plot of DNA methylation levels of 329 DMRs between DNA methylation subgroups of each cohort. The correlation coefficients of hypermethylated and hypomethylated CpG islands were separately calculated. E Overall survival analysis of DNA methylation of TCGA PTC cohort. Statistical significance was calculated using the log-rank test.

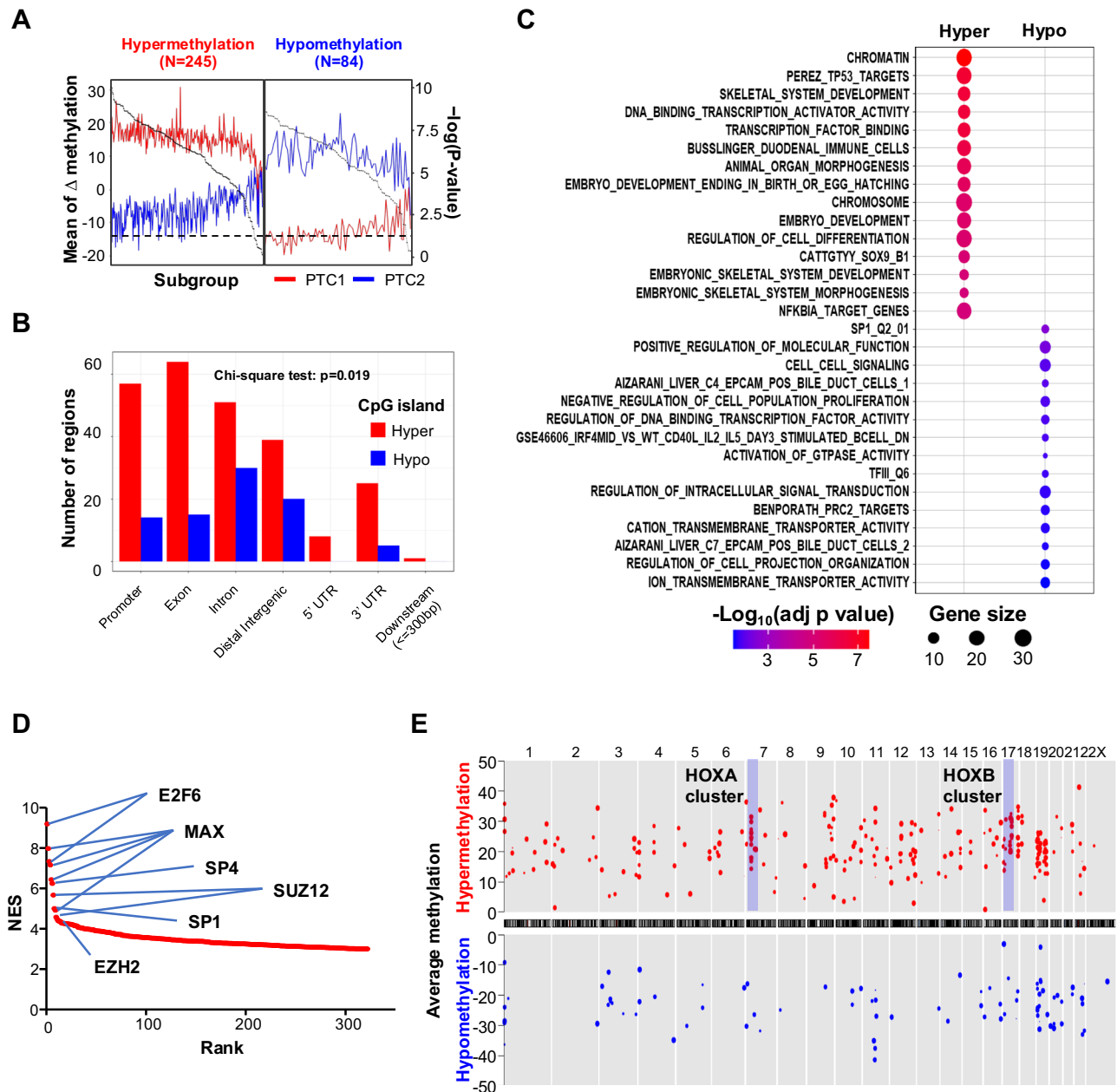


Fig. 2 | Categorization and functional analysis of DMRs in PTC subgroups.

A Classification of DMRs as hypermethylated or hypomethylated in PTC1 compared with PTC2. The classification was based on the relative average methylation levels, and DMRs are ranked by p -values. **B** Genomic annotation of the DMRs. The DMRs were annotated with the reference of TxDB and org.HS.eg.db. The statistical

evaluation of the DMR annotation was conducted using the chi-square test.

C Ontology analysis of the annotated genes closely located to DMRs. **D** Motif analysis of the DMRs. This plot was ranked by the normalized enrichment score of transcription factor binding motifs in the 329 DMRs. **E** Coverage plot visualizing the location of the DMRs in a genome-wide view.

indicated that genes related to tumorigenesis and tumor evasion were upregulated in PTC1 (Fig. 3E, Supplementary Fig. 5C and D).

Integrative analysis of DNA methylation, gene expression, and chromatin accessibility of PTC subgroups

To identify genes directly affected by changes in methylation, we integrated our methylation data with publicly available RNA-seq and ATAC-seq data for comprehensive analysis. First, we selected 77 genes exhibiting differential DNA methylation and expression levels in PTC1 compared to those in PTC2 (Fig. 4A). To examine chromatin accessibility in each subgroup, we used publicly available ATAC-seq data (GSE162515)¹⁴ and inferred DNA methylation subgroups by comparing the expression patterns of these 77 genes with TCGA RNA-seq dataset. Hierarchical clustering enabled us to classify 5 patients as PTC1 and 23 patients as PTC2 (Fig. 4B). PCA of the

ATAC-seq data showed segregation by subgroup (Supplementary Fig. 6A). Using Diffbind²⁶, we observed differential chromatin accessibility and defined differentially accessible regions (DARs) for each subgroup (Supplementary Fig. 6B–D). Strikingly, by annotating the location of DARs in the genome, we found a significantly enriched proportion of promoters in regions with increased chromatin accessibility in PTC1 compared to that in PTC2, whereas regions with decreased accessibility in PTC1 showed few promoters and a higher distribution in distal intergenic regions (Supplementary Fig. 6E). Gene ontology analysis of the genes nearest to the differentially accessible regions in PTC1 revealed regions enriched in immune cell activation, which is consistent with the gene ontology results from TCGA RNA-seq data (Fig. 4C). Conversely, regions with decreased accessibility were associated with development, differentiation, and proliferation of various cell types (Fig. 4D). Our integrated analysis revealed a set of

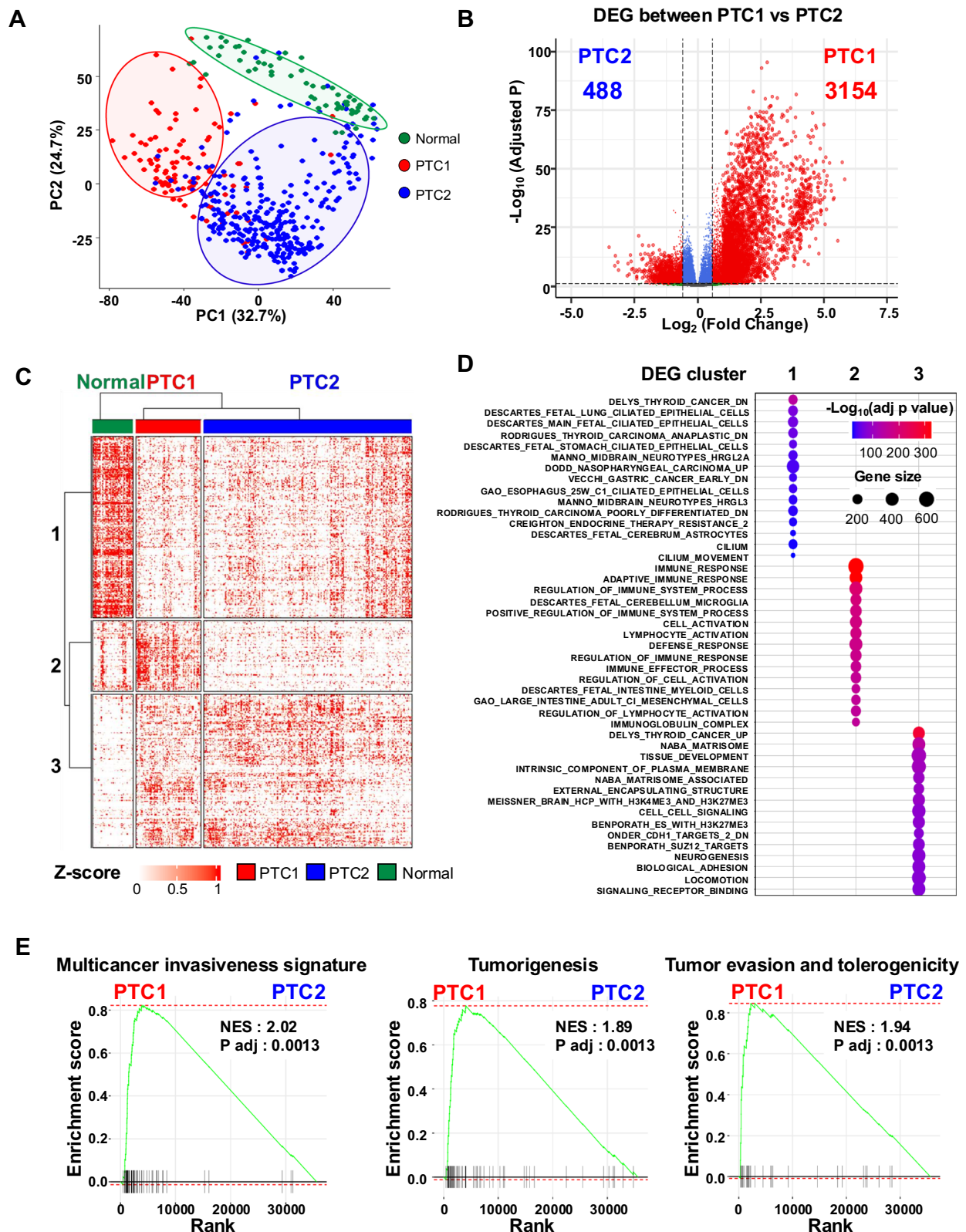


Fig. 3 | Transcriptomic analysis of PTC subgroups in TCGA cohort. **A** PCA of PTC based on RNA-seq data from TCGA cohort. **B** Volcano plot highlighting DEGs between PTC1 and PTC2. Red dots indicate DEGs between PTC1 and PTC2. **C** Heatmap showing expression patterns of DEGs across PTC1, PTC2,

and normal thyroid tissues. The DEGs were clustered via K-means clustering ($K = 3$). **D** Gene ontology analysis of each gene cluster, as defined in Fig. 3C. **E**. GSEA comparing PTC1 to PTC2. The GSEA terminologies are from MSigDB.

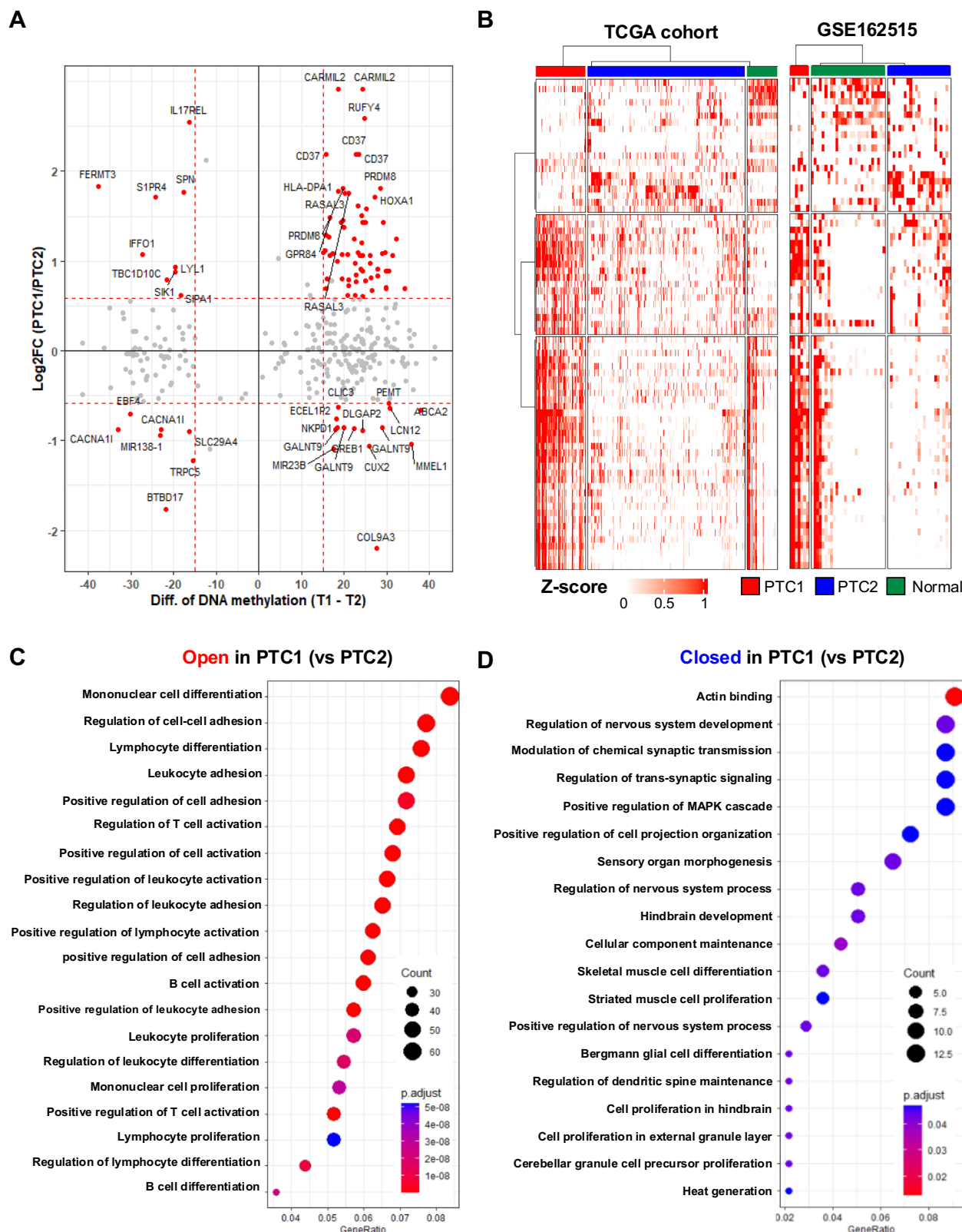


Fig. 4 | Differential chromatin accessibility and its association with gene expression in PTC methylation subgroups. **A** Dot plot illustrating genes with differential methylation and expression levels of PTC1 and PTC2 across our cohort and TCGA cohort. Red dots indicate differentially methylated and expressed genes between

PTC1 and PTC2 in our cohort and TCGA cohort, respectively. **B** Classification of patients based on the expression profile of 77 genes shortlisted in Fig. 4A, comparing TCGA (left) and GSE162515 (right) cohorts. Gene ontology analyses for genes near regions of increased (**C**) and decreased (**D**) chromatin accessibility.

Table 1 | Integrated catalog of candidate genes combined from targeted bisulfite sequencing, RNA-seq, and ATAC-seq analysis

Genomic location	Gene	Targeted bisulfite sequencing	RNA-seq (TCGA)		RNA-seq (GSE162515)		ATAC-seq (GSE162515)
		Met diff (T1 - T2)	LFC	Padj	LFC	Padj	LFC
chr11:65352231-65353134	<i>EHP1L1</i>	34.28	0.70	4.4E-20	1.28	5.5E-03	0.44
chr11:65359292-65360328	<i>EHP1L1</i>	29.75	0.70	4.4E-20	1.28	5.5E-03	0.27
chr12:58132478-58132734	<i>AGAP2</i>	29.19	1.43	1.4E-36	2.74	1.6E-06	1.60
chr4:81109887-81110460	<i>PRDM8</i>	28.44	1.81	1.8E-47	2.29	2.1E-04	0.48
chr5:10649367-10650352	<i>ANKRD33B</i>	27.73	0.91	3.1E-12	2.20	1.2E-04	0.39
chr14:93153278-93154759	<i>RIN3</i>	26.37	0.82	2.6E-28	1.89	3.4E-05	0.39
chr12:58130870-58132047	<i>AGAP2</i>	24.94	1.43	1.4E-36	2.74	1.6E-06	0.58
chr12:58119909-58121551	<i>AGAP2</i>	24.36	1.43	1.4E-36	2.74	1.6E-06	0.36
chr16:67686860-67687674	<i>CARMIL2</i>	24.28	2.92	1.7E-77	4.30	3.4E-06	0.25
chr15:40583093-40583526	<i>PLCB2</i>	23.26	1.61	2.1E-47	2.62	3.4E-05	0.28
chr19:49841187-49841628	<i>CD37</i>	22.72	2.19	5.2E-57	3.85	4.0E-06	0.98
chr19:17877468-17877777	<i>FCHO1</i>	22.70	0.77	2.0E-06	2.47	3.1E-03	0.28
chr5:43037259-43037520	<i>ANXA2R</i>	22.32	1.24	1.1E-27	2.06	7.4E-05	1.99
chr2:11774310-11774521	<i>GREB1</i>	22.24	-0.87	4.7E-05	-2.15	2.0E-02	0.72
chr19:15563869-15564223	<i>RASAL3</i>	21.03	1.76	5.7E-57	3.01	3.2E-06	0.32
chr19:15568027-15569227	<i>RASAL3</i>	20.10	1.76	5.7E-57	3.01	3.2E-06	0.87
chr21:45789090-45789373	<i>TRPM2</i>	20.00	1.38	2.0E-36	1.95	3.1E-04	0.63
chr4:81128229-81128691	<i>PRDM8</i>	19.74	1.81	1.8E-47	2.29	2.1E-04	0.29
chr16:67681975-67683924	<i>CARMIL2</i>	18.68	2.92	1.7E-77	4.30	3.4E-06	0.76
chr2:198029068-198029438	<i>ANKRD44</i>	17.22	1.08	8.1E-35	2.75	5.1E-07	0.34
chr12:54764065-54764510	<i>GPR84</i>	16.72	1.48	8.2E-33	1.90	1.4E-02	0.30
chr17:3847999-3848570	<i>ATP2A3</i>	16.40	1.27	6.8E-32	2.84	3.4E-06	1.47
chr17:14201726-14202052	<i>HS3ST3B1</i>	15.68	1.29	8.7E-28	3.00	3.0E-07	1.88
chr19:49842654-49843628	<i>CD37</i>	15.67	2.19	5.2E-57	3.85	4.0E-06	0.64
chr19:1070985-1071812	<i>ARHGAP45</i>	15.06	1.09	5.4E-33	2.61	3.4E-05	0.52
chr7:5336513-5336894	<i>SLC29A4</i>	-16.26	-0.90	5.2E-06	-2.22	9.6E-03	2.48
chr22:50483350-50483579	<i>IL17REL</i>	-16.26	2.55	1.1E-29	5.48	1.9E-03	0.32
chr16:29675845-29676120	<i>SPN</i>	-17.67	1.77	2.5E-44	2.64	6.5E-04	2.61
chr11:65408344-65408631	<i>SIPA1</i>	-18.32	0.62	3.8E-14	1.72	5.4E-05	2.42
chr19:13207375-13207621	<i>LYL1</i>	-19.67	0.94	1.0E-16	2.40	1.1E-05	1.84
chr11:67176945-67177169	<i>TBC1D10C</i>	-21.63	0.79	9.2E-15	1.91	1.5E-04	2.71
chr17:72347924-72348322	<i>BTBD17</i>	-21.72	-1.76	1.1E-05	-3.49	3.5E-02	0.70
chr19:3178741-3179986	<i>S1PR4</i>	-24.25	1.71	2.0E-45	3.32	4.7E-07	2.61
chr12:6664425-6665336	<i>IFFO1</i>	-27.25	1.07	3.0E-37	1.46	4.1E-04	2.66
chr11:63974829-63975048	<i>FERMT3</i>	-37.59	1.83	2.4E-54	2.68	1.0E-05	2.47

Met diff difference in average DNA methylation, LFC Log2(Fold Change); Padj Adjusted *p* value.

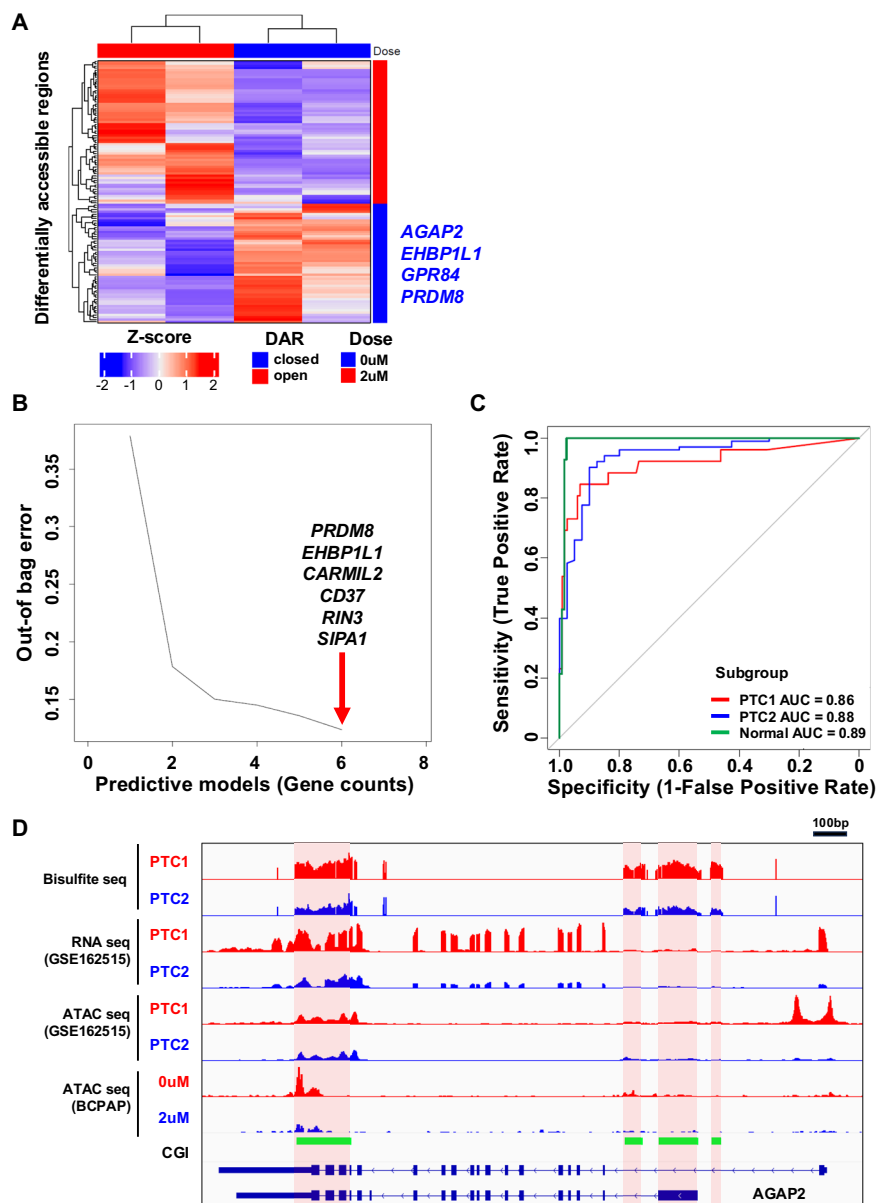
potential biomarkers for PTC1, highlighting the importance of considering DNA methylation, gene expression, and chromatin accessibility. Furthermore, the significant enrichment of immune-related regions in PTC1 reinforces the potential role of the immune system in thyroid cancer progression.

Identification and prioritization of candidate genes for methylation detection system for PTC subgrouping

After conducting an integrated analysis of our bisulfite sequencing, public RNA-seq, and ATAC-seq data, we observed 35 genomic regions in which distinct changes were detected in each dataset, and it was anticipated that these changes were correlated. A total of 28 genes were identified within these regions (Table 1). Before developing a system to sensitively detect methylation changes in these regions, we narrowed down the target genes. To achieve this, we treated the thyroid cancer cell line BCPAP with the

demethylating agent 5-azacytidine (5-Aza), followed by ATAC-seq analysis, to determine which regions exhibited the most significant changes in chromatin accessibility. We identified changes in the expression of the following four genes: *AGAP2*, *EHP1L1*, *GPR84*, and *PRDM8* (Fig. 5A). Additionally, efforts were made to narrow down the target genes using the random forest algorithm. Using VSURF to select fit and non-redundant classifiers of the DNA methylation subgroup, we selected six genes (*PRDM8*, *EHP1L1*, *CARMIL2*, *CD37*, *RIN3*, and *SIPA1*) (Fig. 5B). On applying the algorithm to the test set, we observed a total accuracy of 87.4% (Fig. 5C). Two genes (*EHP1L1* and *PRDM8*) consistently showed changes in both the BCPAP cell line ATAC-seq results and the random forest analysis. *AGAP2* was a representative candidate biomarker observed in all datasets (Fig. 5D). Thus, based on the eight genes that exhibited differences between the BCPAP cell line ATAC-seq results and random forest analysis, we constructed a system for detecting methylation changes.

Fig. 5 | Biomarker identification and validation using ATAC-seq in the BCPAP cells treated with 5-Aza and machine learning. **A** Heatmap displaying the differential accessibility of regions upon 5-Aza treatment. Open DAR means more accessible regions upon BCPAP treatment with 5-Aza. **B** Six key biomarkers were selected through machine learning of TCGA THCA RNA-seq data to distinguish between normal tissue and PTC subtypes. **C** ROC curves evaluating the performance of the random forest model in classifying DNA methylation subgroups and normal tissue. The balanced accuracy scores, computed as the average of sensitivity and specificity, were 0.89 for normal tissue, 0.86 for PTC1, and 0.88 for PTC2. **D** Integrative multi-omic status at the AGAP2 locus.

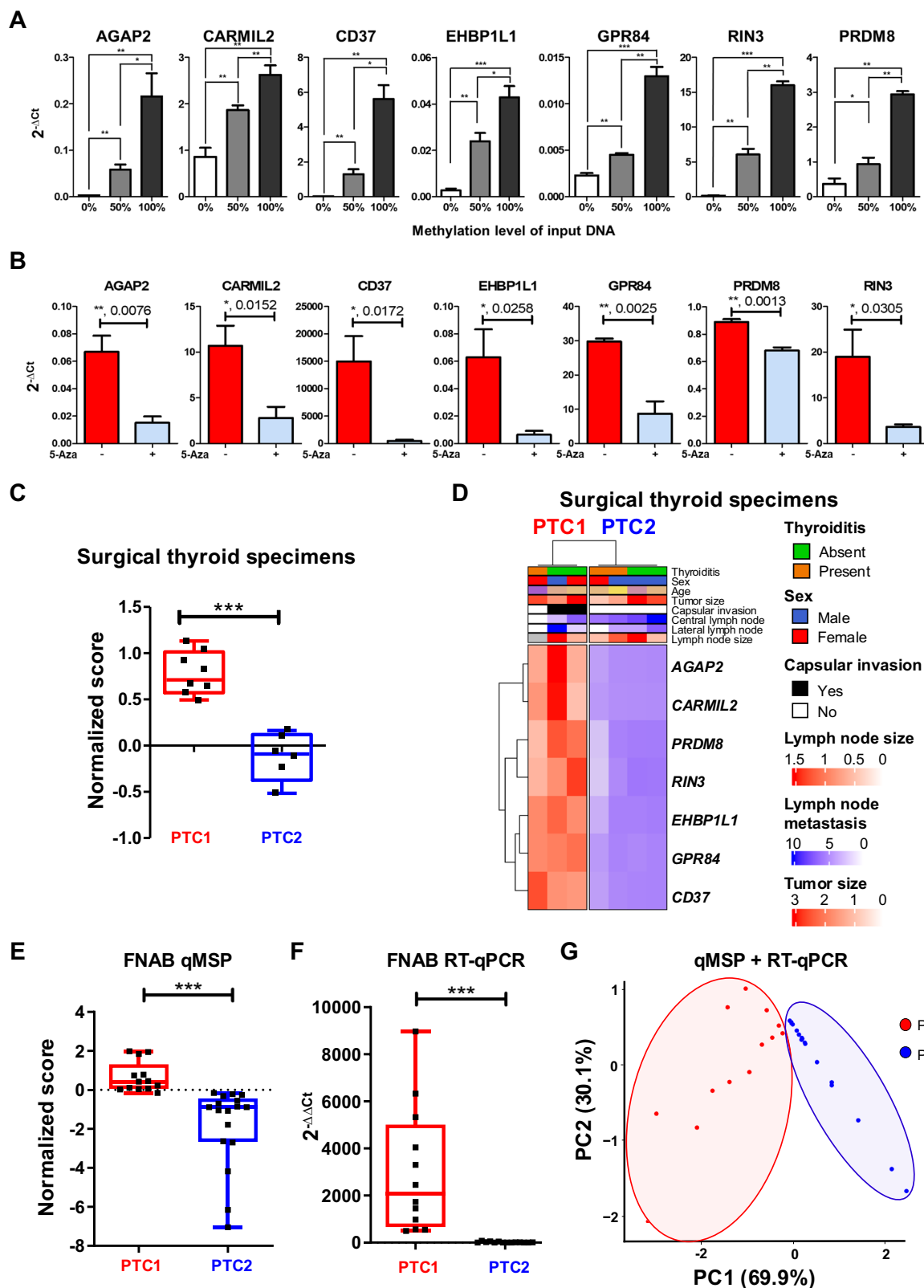


Among identified candidate biomarkers, we selected CD37 as a representative example to validate its role in promoting tumor aggressiveness in vitro. We overexpressed CD37 in 293 T and BCPAP cell lines by transfection and subsequently performed functional assays, including proliferation, invasion, wound healing, and colony-forming assays, to assess the impact of CD37 on tumorigenic potential (Supplementary Fig. 7). The results demonstrated that CD37 significantly enhanced the tumorigenic properties of both cell lines and this suggests that CD37 plays a critical role in promoting tumor progression. These findings suggest that the candidate biomarkers identified through our analysis, including CD37, may be involved in influencing the aggressiveness of PTC.

Validation and clinical application of the MSP-based methylation subgrouping system

MSP is a widely used and simple method for evaluating DNA methylation levels in specific genomic regions²⁷. To establish a system for molecularly categorizing patients with PTC into PTC1 or PTC2, we used the MSP technique. Our validation experiments prioritized DNA hypermethylated regions, which were more detectable than hypomethylated regions because of the clearer presence of methylation signals. Thus, the CpG island in the

SIPA1 region showing hypomethylation in PTC1 was excluded from further validation to ensure methodological consistency and a clear interpretation of the results. We ranked CpG sites within candidate DMRs based on the variance in DNA methylation levels between PTC1 and PTC2, as well as between PTC1 and paired normal tissues. We designed MSP primers targeting the genomic regions encompassing these top-tier CpG sites (Supplementary Fig. 8), utilizing MethPrimer software for primer design²⁸. To assess the effectiveness of our MSP primers in distinguishing the methylation status, we conducted MSP on both unmethylated and methylated control DNA samples treated with bisulfite. The signal ratio between the primers for methylated and unmethylated DNA detected using our MSP primers quantitatively reflected the level of DNA methylation, corresponding to the proportion of methylated to unmethylated control DNA (Fig. 6A). To further validate the detection efficacy of the MSP primers, we applied them to the genomic DNA of BCPAP cells treated with 5-Aza. Consistent with our expectations, a noticeable reduction in the methylation signal was observed in the candidate regions of BCPAP cells following 5-Aza treatment (Fig. 6B). Subsequently, we assessed the ability of MSP primers to distinguish between DNA methylation levels in PTC1 and PTC2. We introduced a scoring system to categorize DNA methylation subgroups in



patients with PTC. The scores were calculated as the sum of the seven biomarker methylation signals normalized to the methylation signal generated from paired normal tissues. The results confirmed that the MSP system was highly effective in distinguishing between PTC1 and PTC2 ($p < 0.0001$) (Fig. 6C). Furthermore, we validated our biomarker gene expression in the two subgroups using quantitative RT-PCR, confirming an

increase in the expression of these genes in patients belonging to PTC1, as expected (Fig. 6D).

For methylation and expression validation, 2 ng of nucleic acid was required for a single qMSP, and 5 ng was required for a single qRT-PCR; these quantities are readily obtainable from FNAB samples. Previous studies have demonstrated the adequacy of these quantities, even with a limited

Fig. 6 | Methylation and expression profiling and diagnostic potential of candidate biomarkers through PTC tissues. **A** Quantification of methylation signals using methylation control DNA with each qMSP primer set. Each qMSP was performed with three technical replicates. **B** Quantification of DNA methylation signal with BCPAP genomic DNA following 5-Aza treatment using each qMSP primer set. Three independent experiments were conducted. **C** DNA methylation signal quantification across genomic DNA from PTC subgroups using surgical thyroid specimens. The average diagnostic score of each subgroup was calculated. **D** Validation of gene expression via quantitative PCR in the discovery cohort, using surgical thyroid specimens. PTC tissues were categorized according to the

methylation status of 329 DMRs. **E** Assessment of biomarker methylation levels in 30 FNAB samples from patients with PTC using real-time MSP. **F** Assessment of biomarker expression levels in 30 FNAB samples from patients with PTC using real-time PCR. **G** PCA plot of 30 patients' qMSP scores (PC1, representing methylation levels) and qRT-PCR scores (PC2, representing gene expression levels). (**A**, **C**, **D**) Two-sided *t*-tests were used for comparisons. (**B**) One-sided *t*-test was applied due to the anticipated decrease in methylation signal following 5-Aza treatment. (**E**, **F**) Unpaired Student's *t* test was used to compare groups. ****P* < 0.001; ***P* < 0.01; **P* < 0.05.

number of cancer cells present in FNAB samples^{29,30}. We aimed to assess whether our sensitive approach could detect DNA methylation variations in candidate DMRs in minimally invasive biopsy samples. We conducted qMSP using thirty FNAB samples from a new prospective cohort with PTC (Supplementary Table 3). We successfully identified the subgroups divided by the DNA methylation score of seven candidate biomarkers using a trace amount of DNA extracted from tumor cells in FNAB samples (Fig. 6E). Specifically, we classified 13 patients as PTC1 and 17 patients as PTC2 based on DNA methylation profiles. Moreover, we found that the expression levels of the candidate genes from the corresponding FNAB samples classified as PTC1 are higher than that from PTC2 using qRT-PCR (Fig. 6F), confirming the correlation between DNA methylation and expression level among DNA methylation subgroups by FNAB samples. To visualize the methylation and expression patterns of each biomarker, we generated heatmaps displaying individual methylation levels and expression values for the seven genes in FNAB samples (Supplementary Fig. 9). Furthermore, PCA plot of combined qMSP and qRT-PCR data clearly demonstrated that patients were divided into two distinct subgroups (Fig. 6G). Notably, the qMSP method yielded a more distinct and accurate separation of patient groups compared to qRT-PCR, emphasizing the utility of DNA methylation as a highly specific marker for patient classification. Moreover, clinicopathological analysis revealed that PTC1 patients exhibited more aggressive tumor characteristics, including larger tumor size and higher frequency of bilateral disease, compared to PTC2 (Supplementary Table 3). Additionally, a significantly higher proportion of PTC1 patients underwent total thyroidectomy, indicating that these tumors were associated with a greater disease burden and were more likely to require extensive surgical intervention. These results demonstrate that qMSP-based methylation profiling identifies clinically distinct patient subgroups, highlighting its potential for precise preoperative risk stratification.

The TNM staging system serves as a fundamental framework for classifying malignancies³¹ and is widely used to predict patient prognosis and aid in cancer diagnosis. We aimed to evaluate the potential enhancement of survival prediction by integrating our classification system with the established TNM staging system, as in a previous study³². Intriguingly, our survival analysis, utilizing TNM staging data from TCGA combined with DNA methylation subgroup status (PTC1 and PTC2), revealed a significant disparity in survival rates. Patients classified under TNM stages III–IV and subgroup PTC1 exhibited notably lower survival rates than those in the same TNM stages but within subgroup PTC2 (Supplementary Fig. 10). This finding suggests that DNA methylation subgrouping may serve as an independent prognostic factor, augmenting the traditional TNM classification. Therefore, differentiating between PTC1 and PTC2 is critical for determining the surveillance of thyroid cancer.

Discussion

In this study, we identified 7217 CpG islands of interest from DNA methylation data and isolated 333 DMRs from our cohort using bisulfite sequencing. Utilizing K-means clustering and principal component analysis, we delineated two distinct patient subgroups, which were validated through correlation analysis with TCGA data. Notably, the subgroup designated as PTC1 exhibited worse overall survival compared to PTC2, with 245 hypermethylated and 84 hypomethylated regions primarily located in promoter and distal intergenic regions. Functional analysis revealed

that hypermethylated DMRs in PTC1 were enriched with genes involved in development and Polycomb Repressive Complex 1 and 2 (PRC1/PRC2) binding motifs, which are associated with tumor progression and poorer prognosis. By integrating bisulfite sequencing, RNA sequencing, and ATAC-seq data, we identified 35 genomic regions and 28 genes as potential biomarkers. Seven genes were ultimately established as bona fide biomarkers, confirmed through qMSP and qPCR analyses. Importantly for clinical application, our highly sensitive quantification system enabled the detection of methylation variations in FNAB samples—a minimally invasive method routinely used in preoperative evaluation.

Among these seven biomarkers, *RIN3* and *AGAP2* play pivotal roles in the Ras pathway and are critically involved in thyroid cancer progression. *RIN3* contains a Ras-association domain, indicating its involvement in the Ras signaling pathway. *AGAP2* exerts anti-apoptotic effects through the activation of nuclear phosphoinositide 3-kinase (*PI3K*), a process frequently initiated by Ras activation. Furthermore, *AGAP2*, *EHBP1L1*, and *CARMIL2* contribute to cell migration and invasion, highlighting their significance in cancer metastasis. Conversely, *CD37* and *GPR84* are linked to the immune system, aligning with the PTC1 immune-related transcriptomic features and suggesting their potential role in modulating the immune response in thyroid cancer. These seven biomarkers represent underexplored areas in thyroid cancer research. Their identification not only provides a deeper understanding of the molecular intricacies of THCA but also highlights their novelty and potential as therapeutic targets or diagnostic markers in thyroid cancer, particularly given their diverse functions, ranging from signaling pathways to immune system interactions.

We observed enrichment of binding sites for *MAX* and *E2F6*, components of the PRC1 complex, as well as *SUZ12* and *EZH2*, components of the PRC2 complex, in hypermethylated DMRs. Previous studies have highlighted the interplay between the PRC1 complex and DNA methylation in regulating the expression of genome defense genes during mammalian development³³. These genes play critical roles in maintaining cellular genomic stability by detecting and repairing DNA damage, minimizing mutations, and suppressing the activity of selfish DNA elements, such as transposons, thereby preserving genomic integrity³⁴. Furthermore, the PRC2 complex, known for its role in mediating H3K27 methylation and consequently silencing the genome, recruits DNMTs³⁵. Abnormal activity of the PRC2 complex has been widely recognized to alter transcriptional regulation, thereby contributing to cancer development³⁶. However, further bioinformatics and experimental validation are imperative to ascertain whether these hypothesized mechanisms are characteristic of the PTC1 subgroup.

FNAB is a cornerstone in the preoperative evaluation of thyroid nodules, primarily used to determine malignancy and guide the necessity of surgical intervention¹⁰. According to the 2024 NCCN guidelines, active surveillance is sometimes recommended for small thyroid cancers (<1 cm) instead of immediate surgery. However, even in small tumors, if they exhibit aggressive characteristics, prompt thyroidectomy is necessary. The challenge lies in the fact that traditional cytological evaluation alone may not always provide sufficient information to determine tumor aggressiveness, particularly in indeterminate cases. Our study addresses this limitation by demonstrating that FNAB-based methylation profiling can serve as a complementary molecular tool, providing objective epigenetic insights into tumor behavior beyond cytological assessment. By integrating this approach

into existing clinical workflows, clinicians could refine risk stratification, allowing for the safe surveillance of indolent tumors while ensuring timely surgical intervention for high-risk cases. Given that active surveillance is increasingly considered for small thyroid cancers, incorporating methylation profiling into FNAB analysis could further enhance clinical decision-making, reducing unnecessary surgeries while ensuring aggressive cases are promptly treated.

Notably, qMSP demonstrated higher specificity and clearer subgroup separation compared to qRT-PCR, likely due to the inherent instability of RNA and variability in expression levels, which can compromise the consistency of qRT-PCR results. In contrast, qMSP detects the more stable methylation patterns in DNA, offering a more reliable distinction. This reliability is crucial in surgical planning, as it allows for confident stratification of patients even with minimal DNA obtained from FNAB samples. Our primers successfully classified subgroups under these conditions, reinforcing qMSP as a valuable tool for distinguishing tumor subtypes in thyroid cancer.

For surgeons, this means that beyond diagnosing malignancy, FNAB provides valuable prognostic information preoperatively, assisting in decisions regarding the extent of surgery and the need for additional treatments like radioactive iodine (RAI) therapy. According to the 2024 NCCN guidelines, total thyroidectomy and RAI therapy are not routinely recommended for T1b/T2 (1–4 cm) N0/NX or small-volume N1a disease. However, our findings suggest that high-risk PTC2 patients, who exhibit poorer prognosis, may benefit from more aggressive interventions, such as total thyroidectomy instead of lobectomy, and consideration of postoperative RAI therapy to reduce recurrence risk. The ability to accurately classify tumor subtypes preoperatively can help surgeons decide between total thyroidectomy and lobectomy, determine the necessity for lymph node dissection, and plan for more aggressive treatments in patients identified as having the high-risk PTC1 subtype.

Beyond surgical considerations, the molecular distinction between PTC1 and PTC2 could also provide insights into novel therapeutic targets. If PTC2 exhibits specific epigenetic alterations or signaling pathway activations that drive its aggressive behavior, targeted therapies could be explored to mitigate progression and improve patient outcomes. Further investigation into the functional consequences of these methylation patterns may uncover actionable targets for personalized treatment approaches. To further explore the prognostic value of these subtypes, we conducted a multivariate Cox regression analysis, incorporating key clinicopathological factors, including sex, TNM stage, TCGA methylation subgroups, and TCGA-defined PTC1 and PTC2 clusters. The results demonstrated a trend toward worse survival outcomes in the PTC1 group compared to PTC2 (HR = 3.6, 95% CI: 0.91–14.1). Although the *p*-value (0.068) did not reach conventional statistical significance, the hazard ratio suggests a potential survival difference between the two subgroups, highlighting the need for further investigation with additional clinical and molecular factors to refine their prognostic distinction. The relatively small cohort size (*N* = 55) in our study limits our ability to independently validate the survival differences observed in TCGA. Furthermore, as our cohort is prospectively designed, the long-term survival impact of these molecular subtypes remains to be fully elucidated. We mitigated this limitation by incorporating TCGA data to infer clinical significance, but further large-scale, long-term studies are needed to validate these biomarkers and fully integrate them into clinical practice. In future studies, we plan to incorporate a broader range of clinicopathological and molecular factors into survival analysis to improve risk stratification, while larger, long-term cohorts will be essential to confirm these findings and support their clinical application.

Nonetheless, our findings have significant implications for surgical practice by providing a molecular tool to predict tumor behavior, allowing for personalized treatment plans, improved prognostic accuracy, and enhanced patient counseling regarding surgical options and expected outcomes. This approach aligns with the principles of precision medicine, aiming to tailor interventions based on individual tumor biology.

In conclusion, our study presents a novel, minimally invasive method for preoperative risk stratification in papillary thyroid carcinoma using epigenetic biomarkers detectable in FNAB samples. By detecting specific DNA methylation patterns associated with tumor aggressiveness, qMSP enables accurate patient stratification, even with minimal DNA input. Through these insights, we offer surgeons a valuable tool to inform surgical decision-making. This approach enables more precise tailoring of surgical interventions, potentially improving patient outcomes and optimizing resource utilization in thyroid cancer care. Future large-scale studies are warranted to validate these biomarkers and fully integrate them into clinical practice, advancing personalized medicine in surgical oncology.

Methods

Human tissue collection and processing

Fifty-five papillary thyroid tumor and adjacent normal tissue samples, along with thirty FNAB specimens, were collected from Gangnam Severance Hospital (Seoul, Korea). The use of these samples was approved by the Institutional Review Board of Gangnam Severance Hospital (IRB number: 3-2020-0309) in accordance with institutional ethical guidelines. All patients provided written informed consent for the research use of their tissues. All procedures performed in this study involving human participants comply with the principles of the Declaration of Helsinki.

Comprehensive demographic data including age, sex, TNM stage can be found in Supplementary Table 1. Those surgically excised tissues were immediately stored in Dulbecco's modified Eagle's medium (Cytiva, Marlborough, MA, USA) supplemented with 10% FBS (Cytiva, Marlborough, MA, USA) and transported to the laboratory. FNAB samples were collected separately from thirty patients. All FNAB were performed by an experienced clinicians and were reviewed by an experienced cytologist. FNAB was performed using a 20 mL plastic syringe, secured in a syringe holder (Cameco syringe pistol, Precision Dynamics, San Fernando, CA). The needle varied from 21 to 23 gauge, according to the sonographic pattern of the nodules. Upon arrival, the tissues were rinsed twice with PBS (Cytiva, Marlborough, MA, USA) and then rapidly snap-frozen using liquid nitrogen to preserve the integrity of the nucleic acids for subsequent use.

Selection of CpG islands to design a probe pool for targeted bisulfite sequencing

To select candidate CpG islands for bisulfite sequencing, we utilized Infinium HumanMethylation450 BeadChip data, as well as RRBS data obtained from thyroid cancer (THCA) data from TCGA¹² and a thyroid cancer dataset from the Gene Expression Omnibus (GSE107738). The beta values of each CpG site were averaged to represent the methylation value of its corresponding CpG islands, as per the human reference genome (NCBI GRCh37; hg19). Subsequently, the methylation values of CpG islands in normal tissue samples were averaged to generate representative methylation values for normal tissues. Differences in methylation between tumor samples and the average of normal tissues were calculated and tabulated. CpG islands displaying methylation differences greater than or equal to 10% in more than 10% of the patients were shortlisted. Based on these criteria, 5812, 2612, and 2484 CpG islands were selected by comparing normal and tumor samples from TCGA, normal and benign tumor samples from GSE107738, and normal and malignant samples from GSE107738, respectively. To eliminate nonspecific target regions in the genome, we used the probe design tool provided by Arbor Bioscience (MyBaits), resulting in the final selection of 7217 CpG islands as targets.

Targeted bisulfite library preparation and sequencing

Genomic DNA was extracted from the collected tissue samples using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany), following the manufacturer's recommendations and guidelines. Subsequently, DNA concentration and purity were determined using a UV spectrophotometer (Nanodrop 2000; Thermo Fisher, Carlsbad, CA, USA). A total of 500 ng of genomic DNA was fragmented using an M220 Focused Ultrasonicator (Covaris, Woburn, MA, USA) in low-EDTA TE buffer. The quality,

quantity, and fragment size (with a major peak in the range of 250–300 bp) of the sheared genomic DNA were confirmed using a 2100 Bioanalyzer system (Agilent Technologies, Santa Clara, CA, USA) before library preparation. The DNA library was bisulfite-converted using the EZ DNA Methylation-Gold Kit (Zymo Research, Irvine, CA, USA). Subsequent library preparation was performed using the Accel-NGS® Methyl-Seq DNA Library Kit (Swift Biosciences, Ann Arbor, MI, USA), along with other specified enzymes, buffers, and reagents, following the manufacturer's protocol. Finally, the eight libraries were pooled and incubated with a probe pool designed to target the regions of interest. After cleanup, the libraries were sequenced on a HiSeq2500 sequencer (Illumina, San Diego, CA, USA) with 2 × 100 bp paired-end reads featuring a unique dual index, generating 2 Gb of sequencing data from each sample.

ATAC library preparation and sequencing

The ATAC libraries for BCPAP were prepared as previously described³⁷. The procedure was as follows: The cells were lysed according to the preparation guidelines with fresh ATAC-seq lysis and wash buffers. Cells were pelleted, and the supernatant was removed before resuspension in the lysis buffer. After incubation, the samples were diluted, washed, and pelleted. A transposition mix was added, followed by incubation. After incubation, the DNA was cleaned and concentrated. Barcoding involved the assignment of unique adapters, PCR amplification, and cleanup. Library concentrations were then quantified for sequencing. The libraries were sequenced on a HiSeq2500 sequencer (Illumina, San Diego, CA, USA) with 2 × 100 bp paired-end reads featuring a unique dual index, generating 2 Gb of sequencing data from each sample.

Preprocessing of next-generation sequencing data

Targeted bisulfite sequencing data were quality-controlled using FastQC (version 0.11.9). Trim Galore (version 0.6.7) and Cutadapt³⁸ (version 2.8) were used to eliminate adaptor sequences and low-quality sequences. Bismark³⁹ (version 0.22.3), based on the human CpG island reference hg19, aligned the sequencing reads with Bowtie2 (version 2.4.4). The alignment quality of each dataset was assessed using Picard CollectHsMetrics, and all quality control data were summarized using multiQC⁴⁰. Methylated and unmethylated cytosines at each CpG site were identified from the post-indexed data using the Bismark methylation extractor, and only those 10X or higher were selected for downstream analysis. Finally, the methylation values of the CpG sites within the same CpG island were calculated by averaging the methylation values based on hg19. The RNA-seq data were aligned using HISAT2, and the sam files were sorted using Samtools. Read quantification of the bam files was conducted using HTSeq (version 0.11.1) to generate read count data, which were normalized using DESeq2 (version 3.12). Principal component analysis (PCA) of the gene expression patterns was performed using ggplot2 (version 3.3.3) to analyze sample correlations. Differentially expressed genes (DEGs) between samples were identified using DESeq2, selecting genes with at least a 2-fold expression difference and an adjusted *p*-value of < 0.05. ATAC sequencing (ATAC-seq) data from BCPAP cell lines and publicly available data (GSE162515) were processed using the PEPATAC pipeline. PEPATAC employs a RefGenie Asset Manager for consistent reference genome annotation, which is crucial for reproducibility. Adapters and mitochondrial DNA were trimmed using Trimmomatic, and read alignment and deduplication were performed using Bowtie2 and Picard, respectively.

Selecting the DMRs

Analyses were performed under the assumption that the average value represented each CpG island. Targeted bisulfite sequencing data were examined to identify targets where DNA methylation exhibited a change of more than 20% in tumors compared to paired normal tissues in over 20% of 55 patients with thyroid cancer. Additionally, DMRs were identified separately in patients with and without thyroiditis, enabling a comparative analysis of the epigenetic changes associated with this condition. In total, 333 DMRs were identified. These regions were further classified into

hypermethylated and hypomethylated groups based on their average methylation levels across DNA methylation subgroups.

Analysis of next-generation sequencing data

Before the analysis, we selected 399 patients with PTC from the THCA (TCGA, Firehose Legacy) dataset available on cBioPortal^{41,42}. The relative DNA methylation levels of CpG islands in TCGA samples were determined by subtracting the average methylation level of normal thyroid samples from that of each tumor sample. To divide patients in each cohort into two subgroups, we applied K-means clustering. To ensure the stability and reliability of the clustering, we ran the algorithm 1000 times with different random centroids, selecting the best outcome and allowing a maximum of 5000 iterations for convergence. Genomic regions of interest were annotated using the ChIPseeker⁴³ package with the TxDB.Hsapiens.UCSC.hg19.knownGene and org.HS.eg.db packages in R. We utilized i-cisTarget^{19,20} to explore the predictive consensus motifs in the 329 candidate DMRs. For gene ontology analysis and gene set enrichment analysis (GSEA), the gprofiler2⁴⁴ and fgsea packages were utilized with customized gene matrix transposed (GMT) files from the molecular signatures database (MSigDB)^{45,46}. Data visualization from the DNA methylation microarray, bisulfite sequencing, RNA-seq, and ATAC-seq was performed using the ComplexHeatmap package in R (version 4.2.2). Differentially accessible regions across conditions were identified using Diffbind, with a threshold of |Fold Change| > 2 for differential accessibility analysis and |Fold Change| > 0.25 for selecting candidate biomarkers. The initial set of 28 candidate genes was refined to select a core gene set and eliminate redundancy using the random forest algorithm implemented in the VSURF package in R.

Multivariable Cox proportional hazards regression analysis

Clinical and molecular data were obtained from The Cancer Genome Atlas (TCGA) PTC cohort, and patients with incomplete survival data or missing covariates were excluded from the analysis. The Cox proportional hazards model was constructed using the survival package (version 3.8-3) in R (version 4.1.0), with hazard ratios (HR) and 95% confidence intervals (CI) estimated for each covariate. Covariates included sex, TNM stage, TCGA patient clusters, and TCGA methylation subgroups. *P*-values for each covariate were calculated using the Wald test, while a global *p*-value for the model fit was obtained from the log-rank test. Model performance was assessed using the Akaike Information Criterion (AIC) and concordance index (C-index). A forest plot was generated using the ggforest() function from the ggplot2 package (version 3.4.0).

qMSP

Before quantifying the DNA methylation levels of the target regions, 500 ng of genomic DNA extracted from patients with PTC and DNA methylation control DNA (Takara Bio Inc., Japan) were treated with sodium bisulfite (EZ DNA Methylation-Lightning Kits, Zymo Research, Irvine, CA, USA). The concentration of bisulfite-converted genomic DNA was quantified using a UV spectrophotometer (Nanodrop 2000; Thermo Fisher Scientific, Carlsbad, CA, USA). For qMSP analysis, we used a master mixture (KAPA SYBR FAST qPCR Master Mix (2X), Kapa Biosystems), which facilitated GC-rich PCR, along with a PCR cyclor (LightCycler 480 II; Roche Diagnostics). For each qMSP run, 2 ng of bisulfite-converted genomic DNA was used. After 45 cycles of quantitative PCR (qPCR) reaction, the crossing point (Cp) was determined by directly adjusting the signal threshold. The DNA methylation

levels of each CpG island were calculated using the following equation Eq. (1):

$$2^{(Cp \text{ of Unmet primer}) - (Cp \text{ of Met primer})} \quad (1)$$

$$\frac{2^{-\Delta Cp(\text{Tumor DNA})} - 2^{-\Delta Cp(\text{Normal DNA})}}{2^{-\Delta Cp(\text{methylated DNA})} - 2^{-\Delta Cp(\text{unmethylated DNA})}} \quad (2)$$

To distinguish between samples based on MSP results, we established the following scoring system: First, we computed the individual DNA methylation scores for each biomarker using the following equation Eq. (2):

To normalize the background signal intrinsic to the normal thyroid tissue of each patient, we calculated a normalized score by subtracting the paired normal score from the score of the corresponding tumor. To assign PTC subgroups, we summed the cumulative DNA methylation scores of the seven biomarkers for each patient and integrated these into a comprehensive DNA methylation score.

qPCR

To check the expression of each candidate gene of patients with PTC, we synthesized cDNA from 1 µg of total RNA using reverse transcriptase (Invitrogen, Carlsbad, CA, USA). For qPCR, we used a master mixture (KAPA SYBR FAST qPCR Master Mix (2X), Kapa Biosystems, Wilmington, MA, USA) with a PCR cyclor (LightCycler 480 II; Roche Diagnostics, Basel, Switzerland). For each qPCR run, 5 ng of cDNA was used. After 45 cycles of qPCR reaction, the Cp was calculated by directly adjusting the signal threshold. The gene expression level of each target gene was determined using the comparative Cp method ($2^{-\Delta\Delta C_p}$), with GAPDH expression serving as the control. To classify PTC subgroups, we calculated the sum of the expression levels of the seven biomarkers for each patient and combined them into a comprehensive score.

Cell culture and transfection

BCPAP (Elabscience, Houston, Texas, USA) were cultured in RPMI-1640 medium (Gibco, Houston, Texas, USA) supplemented with 10% fetal bovine serum (Hyclone, Logan, UT, USA) supplemented with 1% penicillin-streptomycin (Gibco, Houston, Texas, USA), 293 T (ATCC, Manassas, VA, USA) cells were grown in Dulbecco's modified Eagle medium (Hyclone, Logan, UT, USA) supplemented with 10% FBS (Hyclone, Logan, UT, USA) and 1% penicillin (Gibco, Houston, Texas, USA). The cell lines were cultured at 37 °C in a humidified atmosphere containing 5% CO₂. Cells were sub-cultured every 3 days when they reached 80–90%. For CD37 gene overexpression, CD37 sequence was cloned into pcDNA3.1 V5/His A vector (Invitrogen, Waltham, MA, USA). pcDNA3.1 CD37 V5/His A plasmid (12 µg, 100 mm scale) was transfected into cells with Lipofectamine 2000 (Invitrogen, Waltham, MA, USA) as per manufacturer's instructions.

Proliferation assay

For the IncuCyte® cell proliferation assay, 1×10^4 293 T cells were seeded into a 96-well plate. Cell proliferation was monitored using the IncuCyte® Live-Cell Analysis System, with images captured over time to measure cell growth. The IncuCyte® S3 Live-Cell Analysis System (Sartorius, Göttingen, Germany) was used for automated cell count analysis.

Migration and invasion assays

Invasion assay was performed using 24-well transwell systems with 8 µm pore-sized inserts (Costar; Corning Incorporated, Corning, NY, USA). For these assays, 1×10^5 293 T cells were plated into Matrigel-coated upper chambers. The upper chamber was filled with serum-free RPMI medium, while the lower chamber contained RPMI medium with serum, acting as a chemoattractant. After 48 h of incubation, non-invaded or non-migrated cells remaining on the upper side of the membrane were removed with a cotton swab, and the invaded cells that had migrated through the membrane were stained and counted. Wound healing assays were performed in 24-well plates, where 2×10^5 cells were seeded per well. A wound was created using scratcher (SPL; Pocheon, Gyeonggi-do, South Korea). Images of the wound healing were captured at 0 h and 24 h. The width of the wound was measured using ImageJ software.

Colony forming assay

2.5×10^3 cells were plated in 6-well plates and cultured for 12 days. The cells were then fixed with 4% paraformaldehyde and stained with 1% crystal violet to facilitate colony counting.

Statistical analysis

To examine the correlation between DNA methylation subgroups in each cohort, we computed the average relative methylation level of tumor samples from each subgroup and calculated the Pearson correlation coefficient for each subgroup comparison using the 'ggpairs' function in the Gally package. To validate the statistical significance of any clinical analysis, we primarily used the chi-square test. However, in cases where the expected frequency in any cell was less than 5, Fisher's exact test was used to ensure statistical validity. Survival analysis was performed using the Kaplan–Meier method with the log-rank test.

Data availability

The bisulfite and ATAC sequencing data generated in this study are publicly available in the Gene Expression Omnibus (GEO) under the accession numbers GSE256293 and GSE262493, respectively. The GEO (GEO Accession Numbers: GSE107738 and GSE165212) and TCGA-THCA datasets were used for the analyses in this study. Additional data requests may be directed to the corresponding authors.

Received: 5 November 2024; Accepted: 30 April 2025;

Published online: 10 May 2025

References

- Sung, H. et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
- Mazzoni, F. L. A. R. C. A. T. Papillary Thyroid Carcinoma. in *StatPearls. Treasure Island (FL): StatPearls* (2023).
- LiVolsi, V. A. Papillary thyroid carcinoma: an update. *Mod. Pathol.* **24**, S1–S9 (2011).
- Haugen, B. R. et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid* **26**, 1–133 (2016).
- Abdullah, M. I. et al. Papillary Thyroid Cancer: Genetic Alterations and Molecular Biomarker Investigations. *Int. J. Med. Sci.* **16**, 450–460 (2019).
- Biswas, S. & Rao, C. M. Epigenetic tools (The Writers, The Readers and The Erasers) and their implications in cancer therapy. *Eur. J. Pharm.* **837**, 8–24 (2018).
- Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* **20**, 590–607 (2019).
- Klein, E. A. et al. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. *Ann. Oncol.* **32**, 1167–1177 (2021).
- Lone, S. N. et al. Liquid biopsy: a step closer to transform diagnosis, prognosis and future of cancer treatments. *Mol. Cancer* **21**, 79 (2022).
- Yokozawa, T. et al. Thyroid Cancer Detected by Ultrasound-Guided Fine-Needle Aspiration Biopsy. *World J. Surg.* **20**, 848–853 (1996).
- Ugurluoglu, C., Dobur, F., Karabagli, P. & Celik, Z. E. Fine needle aspiration biopsy of thyroid nodules: cytologic and histopathologic correlation of 1096 patients. *Int J. Clin. Exp. Pathol.* **8**, 14800–14805 (2015).
- Cancer Genome Atlas Research, N. Integrated genomic characterization of papillary thyroid carcinoma. *Cell* **159**, 676–690 (2014).
- Yim, J. H. et al. Identification of Tissue-Specific DNA Methylation Signatures for Thyroid Nodule Diagnostics. *Clin. Cancer Res.* **25**, 544–551 (2019).
- Sanghi, A. et al. Chromatin accessibility associates with protein-RNA correlation in human cancer. *Nat. Commun.* **12**, 5732 (2021).
- Lee, Y. et al. Hypermethylation of PDX1, EN2, and MSX1 predicts the prognosis of colorectal cancer. *Exp. Mol. Med.* **54**, 156–168 (2022).

16. Pratilas, C. A. et al. V600E)BRAF is associated with disabled feedback inhibition of RAF-MEK signaling and elevated transcriptional output of the pathway. *Proc. Natl. Acad. Sci. USA* **106**, 4519–4524 (2009).
17. Guo, Y. J. et al. ERK/MAPK signalling pathway and tumorigenesis (Review). *Exp. Ther. Med.* **19**, 1997–2007 (2020).
18. Zaballos, M. A. & Santisteban, P. Key signaling pathways in thyroid cancer. *J. Endocrinol.* **235**, R43–R61 (2017).
19. Herrmann, C., Van de Sande, B., Potier, D. & Aerts, S. i-cisTarget: an integrative genomics method for the prediction of regulatory features and cis-regulatory modules. *Nucleic Acids Res.* **40**, e114 (2012).
20. Imrichova, H., Hulselmans, G., Atak, Z. K., Potier, D. & Aerts, S. i-cisTarget 2015 update: generalized cis-regulatory enrichment analysis in human, mouse and fly. *Nucleic Acids Res.* **43**, W57–W64 (2015).
21. Chang, H., Shin, B. K., Kim, A., Kim, H. K. & Kim, B. H. DNA methylation analysis for the diagnosis of thyroid nodules - a pilot study with reference to BRAF(V) (600E) mutation and cytopathology results. *Cytopathology* **27**, 122–130 (2016).
22. Kikuchi, Y. et al. Aberrantly methylated genes in human papillary thyroid cancer and their association with BRAF/RAS mutation. *Front. Genet.* **4**, 271 (2013).
23. Reinert, T. et al. Comprehensive genome methylation analysis in bladder cancer: identification and validation of novel methylated genes and application of these as urinary tumor markers. *Clin. Cancer Res.* **17**, 5582–5592 (2011).
24. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
25. Delys, L. et al. Gene expression and the biological phenotype of papillary thyroid carcinomas. *Oncogene* **26**, 7894–7903 (2007).
26. Ross-Innes, C. S. et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393 (2012).
27. Herman, J. G., Graff, J. R., Myöhänen, S., Nelkin, B. D. & Baylin, S. B. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc. Natl. Acad. Sci.* **93**, 9821–9826 (1996).
28. Li, L. C. & Dahiya, R. MethPrimer: designing primers for methylation PCRs. *Bioinformatics* **18**, 1427–1431 (2002).
29. Titov, S. E. et al. Combined quantitation of HMGA2 mRNA, microRNAs, and mitochondrial-DNA content enables the identification and typing of thyroid tumors in fine-needle aspiration smears. *BMC Cancer* **19**, 1010 (2019).
30. Bu, D. et al. Identification of breast cancer DNA methylation markers optimized for fine-needle aspiration samples. *Cancer Epidemiol. Biomark. Prev.* **22**, 2212–2221 (2013).
31. Edge, S. B. & Cancer, A. J. C. O. *AJCC cancer staging manual*. Vol. 7 (Springer, 2010).
32. Chen, K. et al. Individualized tumor-informed circulating tumor DNA analysis for postoperative monitoring of non-small cell lung cancer. *Cancer Cell* **41**, 1749–1762.e1746 (2023).
33. Mochizuki, K. et al. Repression of germline genes by PRC1.6 and SETDB1 in the early embryo precedes DNA methylation-mediated silencing. *Nat. Commun.* **12**, 7020 (2021).
34. Ying, H. et al. The Role of DNA Methylation in Genome Defense in Cnidaria and Other Invertebrates. *Mol. Biol. Evol.* **39**. <https://doi.org/10.1093/molbev/msac018> (2022).
35. Vire, E. et al. The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**, 871–874 (2006).
36. Laugesen, A., Højfeldt, J. W. & Helin, K. Role of the Polycomb Repressive Complex 2 (PRC2) in Transcriptional Regulation and Cancer. *Cold Spring Harb Perspect. Med.* **6**. <https://doi.org/10.1101/cshperspect.a026575> (2016).
37. Grandi, F. C., Modi, H., Kampman, L. & Corces, M. R. Chromatin accessibility profiling by ATAC-seq. *Nat. Protoc.* **17**, 1518–1552 (2022).
38. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. **17**, 3. <https://doi.org/10.14806/ej.17.1.200> (2011).
39. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
40. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
41. de Bruijn, I. et al. Analysis and Visualization of Longitudinal Genomic and Clinical Data from the AACR Project GENIE Biopharma Collaborative in cBioPortal. *Cancer Res.* **83**, 3861–3867 (2023).
42. Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
43. Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).
44. Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J. & Peterson, H. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Res* **9**. <https://doi.org/10.12688/f1000research.24956.2> (2020).
45. Subramanian, A. et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545–15550 (2005).
46. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).

Acknowledgements

We thank all the members of Prof. Lark Kyun Kim's laboratory for helping with the critical reading of the manuscript. This work was supported by National Research Foundation of Korea (NRF) grants funded by the Korean government (Ministry of Science and ICT, MSIT) (RS-2023-NR076689 and RS-2024-00408822). This work was also supported by a grant funded by Optolane, Inc. (2021-31-0782).

Author contributions

Y.L. and H.J.P. collected the data, performed the experiments, analyzed the data, and wrote the manuscript. J.S.L. and H.J.J. provided clinical samples and information. S.M.K. and S.H.D. analyzed the data. M.C. and W.W. performed the experiments. S-M.K. conceived the study and provided the clinical samples and information. L.K.K. conceived the study, analyzed the data, and wrote the manuscript. All the authors have read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41698-025-00932-7>.

Correspondence and requests for materials should be addressed to Seok-Mo Kim or Lark Kyun Kim.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025