DNA Methylation-Based Risk Stratification and Classification of Pediatric Thyroid Carcinoma

Jenny Z. Li[1], Julio C. Ricarte-Filho[2], Amber R. Isaza[2], Kyle Hinkle[2], Feng Xu[3], Marilyn M. Li[3,4,5], Andrew J. Bauer[2,4,6], Aime T. Franco[2,4,6,7], and Wanding Zhou[1,5,7]

ABSTRACT

Purpose: Accurate assessment of invasiveness in pediatric thyroid carcinomas is essential to prevent unnecessary surgery and avoid surgery-associated complications. DNA methylation, a proven molecular biomarker for cancer classification, holds promise for stratifying thyroid cancer risk. The objectives were to determine the epigenetic hallmarks of pediatric thyroid carcinomas and investigate whether DNA methylome profiling is a feasible approach for preoperative risk stratification of this pediatric disease.

Experimental Design: We interrogated genome-wide DNA methylation profiles from two separately processed cohorts of pediatric thyroid carcinoma. The reference cohort included 100 samples, consisting of 87 well-differentiated primary tumors—77 papillary and 10 follicular thyroid carcinomas—and 13 matched lymph node metastases. To predict oncogenic drivers and tumor invasiveness, defined by the presence of nodal metastasis, we trained two classifiers on the reference cohort and then evaluated their performance on a second validation cohort of 84 samples, including 83 primary tumors and one lymph node metastasis.

Results: We identified distinct methylation patterns associated with tumor invasiveness and key driver mutations, including *BRAF* p.V600E, *RAS*-like mutations, kinase fusions, and *DICER1* mutations. The differentially methylated regions reflect inflammatory stress and disrupted thyroid development and function, implicating androgen receptor, Hippo, and AP-1 signaling. Leveraging these epigenetic signatures, we developed and validated two methylation-based classifiers that accurately predict tumor invasiveness and oncogenic mutation subgroups.

Conclusions: In patients with pediatric thyroid carcinoma, DNA methylation assays accurately predict tumor invasiveness and driver mutations. Our findings highlight the clinical value of DNA methylation profiling for risk stratification and classification of pediatric thyroid cancers.

Introduction

Thyroid carcinoma is the most common endocrine cancer in children, with increasing incidence over the past 2 decades (1, 2). Similar to adults (3), papillary thyroid carcinoma (PTC) represents approximately 90% of all thyroid cancer cases, with follicular thyroid carcinoma (FTC) and medullary thyroid carcinoma accounting for the remaining 10% of all cases (3). Compared with adult cases, pediatric thyroid malignancies exhibit distinct clinical, pathologic, and molecular features, often presenting as more advanced diseases with higher recurrence rates (4). Rates of extracapsular extension have been reported in up to 50% of children vs. 30% of adults, regional nodal involvement in up to 80% vs. 50%, and distant metastasis in up to 30% versus 5% (5, 6).

The preoperative diagnosis of pediatric thyroid tumors relies on thyroid and neck ultrasound as well as cytopathology; both tests are highly subjective with broad inter- and intra-observer variability (7). Current guidelines recommend total thyroidectomy in nearly all pediatric patients with PTC due to the increased risk of multifocality and associated risk of recurrence and persistent disease (8–10). Though total thyroidectomy lowers the risk of recurrence and persistent disease, it carries higher complication rates in children than in adults, with hypoparathyroidism reported in up to 15% of pediatric patients and recurrent laryngeal nerve injury in 6% (10, 11). For tumors with low invasive potential, total thyroidectomy exposes the patients to an unnecessary risk of surgical complications and a lifelong need for levothyroxine replacement. However, the overall mortality rate of pediatric thyroid carcinoma is low. The incidence-based mortality rates ranged from 0 to 6.35×10^{-4} per 100,000 person-years and did not vary by tumor size or extent of disease (12).

DNA methylation, an essential epigenetic modification involving adding a methyl group to the 5-carbon of cytosine residues in CpG dinucleotides, plays a crucial role in regulating gene expression and maintaining genomic stability (13, 14). Aberrant DNA methylation patterns in thyroid cancer, including global hypomethylation and

[1]Center for Computational and Genomic Medicine, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania. [2]The Thyroid Center, Division of Endocrinology and Diabetes, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania. [3]Genomic Diagnostic Laboratory, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania. [4]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania. [5]Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania. [6]Division of Endocrinology and Diabetes, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania. [7]Abramson Cancer Center, University of Pennsylvania, Philadelphia, Pennsylvania.

J.Z. Li and J.C. Ricarte-Filho equally contributed to this article.

Corresponding Authors: Andrew J. Bauer, Department of Pediatrics, Children's Hospital of Philadelphia, 3500 Civic Center Boulevard, Buerger Center, 12-149, Philadelphia, PA 19104. E-mail: BAUERA@chop.edu; Aime T. Franco, Department of Pediatrics, Children's Hospital of Philadelphia, 3615 Civic Center Boulevard, Philadelphia, PA 19104. E-mail: FRANCOA1@chop.edu; and Wanding Zhou, Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, 3501 Civic Center Boulevard, CTRB 9050, Philadelphia, PA 19104. E-mail: wanding.zhou@pennmedicine.upenn.edu

Clin Cancer Res 2026;XX:XX–XX

doi: 10.1158/1078-0432.CCR-25-2109

Translational Relevance

Thyroid carcinoma is the most common endocrine malignancy in children, and current guidelines recommend total thyroidectomy for nearly all pediatric cases. Although effective, the procedure carries higher complication risks in children, including hypoparathyroidism and nerve injury. Improved preoperative diagnostics could reduce unnecessary surgeries and lifelong hormone dependence. Existing imaging-based approaches are subjective and variable. In this study, we demonstrate that genome-wide DNA methylation profiling robustly captures molecular features of pediatric thyroid carcinoma, including invasiveness and driver mutations. These findings support the potential of DNA methylation as a preoperative prognostic tool to inform treatment decisions and minimize surgical risk.

site-specific hypermethylation, have been linked to various oncogenic processes (15). The exploration of DNA methylation in thyroid carcinoma has largely focused on analyzing changes in CpG island and genic methylation, as well as their effects on specific gene expression (16). Notably, The Cancer Genome Atlas (TCGA) project in 2014 provided comprehensive methylation profiles for a large cohort of PTC cases using the Illumina Infinium HM450 array, suggesting a classification of PTC subtypes based on molecular features and identifying potential biomarkers that could inform disease management (17). Machine learning methods have classified thyroid cancers into subtypes based on their methylation profiles, and a prognostic classifier based on 21 methylation sites was developed to predict recurrence in well-differentiated thyroid cancers (18, 19).

The disrupted mitogen-activated protein kinase (MAPK) signaling pathway is a hallmark of thyroid cancers (1). *BRAF* mutations, particularly *BRAF* p.V600E, are the most prevalent genetic alteration in adult PTC (20), but their prevalence is lower in adolescents and rare in prepubertal pediatric PTC (21). Their associations with the extent of metastases and radioiodine avidity are variable (22–24). *RAS* mutations are the second most prevalent alterations found in adult thyroid tumors (25). Although *RAS* mutations are less common in children, certain *RAS*-like mutations, as well as *PTEN* and *DICER1* mutations, are more prevalent in pediatric tumors and associated with a lower risk for metastasis (26–29). Kinase fusions (KF) involving *RET, NTRK, ALK*, or *BRAF* (30) are more common in pediatric patients. They are associated with an increased rate of regional and distant metastasis in children and adolescents (21, 26, 27, 31, 32).

Despite extensive research on adult thyroid cancers, the potential of epigenome profiles for risk stratification in pediatric thyroid cancer, which exhibits distinct oncogenic signatures, remains underexplored, highlighting the need for tailored classification approaches specific to pediatric patients. To better understand the potential of DNA methylation for stratifying clinical risk in pediatric carcinoma, we conducted a methylome study of a cohort of pediatric malignant thyroid lesions and identified clinically significant DNA methylation biomarkers. These markers revealed known and novel oncogenic mechanisms that contribute to thyroid cancer invasiveness. We also developed a methylation-based classifier to stratify the risk of invasive disease behavior to guide the extent of surgical approaches of pediatric thyroid carcinoma and validated our model on a second cohort.

Materials and Methods

Cohort and clinical data

A total of 184 samples across 169 patients were studied, split into reference and validation cohorts (**Fig. 1A**). The reference cohort ($n = 100$) consisted entirely of fresh-frozen tissue samples used for model training and primary analysis: 87 primary tumors (77 PTCs, 10 FTCs) and 13 paired lymph node (LN) metastases, all of which were paired with PTCs (**Fig. 1A**). The reference patient cohort consisted of 69 females (79%) and 18 males (21%), reflecting the higher prevalence of thyroid cancer in females compared with males. Patient ages ranged from 6.24 to 21.43 years [mean 15.3, standard deviation (SD) 2.80; Supplementary Tables S1 and S2]. The overall sex and age distributions in our pediatric PTC cohort match US distribution (Supplementary Table S3).

The validation cohort ($n = 84$) consisted of 83 primary tumors (78 PTC, five FTC) and one LN metastasis (FTC). The tissue samples were subjected to DNA extraction in a different laboratory before being processed following the same procedures as the reference cohort. Sixty-four (76%) samples were fresh-frozen tissues, and 20 (24%) samples were formalin-fixed paraffin-embedded (FFPE; Supplementary Fig. S1A). The patient cohort consisted of 66 females (80%) and 17 males (20%). Patient ages ranged from 3.64 to 21 years (mean 15.3, SD 3.12). Across both cohorts, 14 matched primary LN metastasis were available for analysis; 13 PTC LNs from the reference cohort were matched to 10 primary tumors in the reference cohort and two in the validation cohort, with case 0212 having two matched LN metastases (one axial). The FTC LN metastasis sample from the validation cohort was matched to its primary in the same cohort (Supplementary Table S1).

The study protocol was approved by the Institutional Review Board of Children's Hospital of Philadelphia (IRB #20-018240), and informed consent was signed by all patients included in the cohort according to the Helsinki Declaration. Tumor stages were categorized according to the eighth edition of the American Joint Committee on Cancer tumor–node–metastasis staging manual. DNA methylation data were generated at the Children's Hospital of Philadelphia from frozen tissue and FFPE samples. Illumina Infinium HumanMethylationEPICv2 (EPICv2) BeadChip microarrays (RRID: SCR_010233; ref. 33) were performed for all thyroid samples. Existing methylation studies conducted on thyroid carcinomas used the Infinium HumanMethylation450 K (HM450) and EPIC BeadChip due to their compatibility with variable DNA input (34), simple workflow (35), and fast processing time. This technology is being replaced by its successor, the EPICv2 array, enabling extensive mapping of more than 937,000 CpG sites (33). The EPICv2 array encompasses promoter regions and extends to CpG island shores that showed significant differential methylation in tissues.

Primary tumors were classified by nodal involvement: low-invasive (N0, no LN metastasis) and high-invasive (N1a/N1b, regional LN metastasis). High-invasive cases with <5 positive LN were designated "low-confidence" to acknowledge borderline clinical staging. In the reference cohort, 32 were low-invasive, and 55 were high-invasive (16 of which were of low confidence). In the validation cohort, 25 were low-invasive, and 58 were high-invasive (14 of which were of low confidence). Distant metastasis (M1) was present in 11 reference cases and nine validation cases, all with N1b staging (Supplementary Tables S1 and S2).

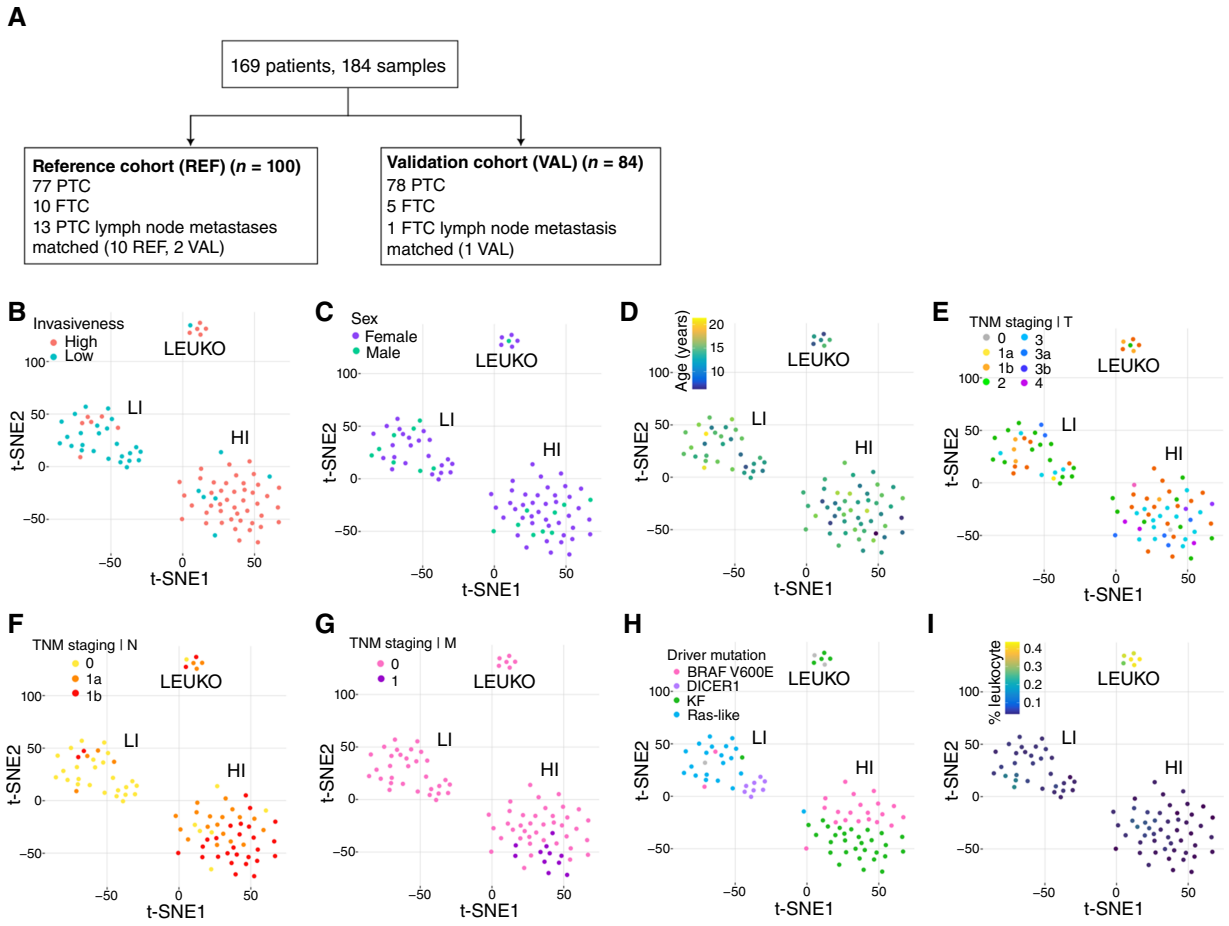

Figure 1.

Molecular subgroups of pediatric thyroid carcinoma defined by invasiveness and driver mutations. **A,** CONSORT diagram of this study, split into reference cohort for primary analysis and training of the models, and validation cohort. **B–I,** t-SNE dimensionality reduction of pediatric primary tumors ($n = 87$) in the reference cohort, color-coded by (**B**) clinical invasiveness, (**C**) sex, (**D**) age at diagnosis, (**E–G**) TNM stage, (**H**) genetic driver mutation group, and (**I**) estimated leukocyte fraction. TNM, tumor–node–metastasis.

Genetic drivers were identified using the CHOP Solid Tumor Panel and CHOP Cancer Fusion Panel or RNA sequencing (RNA-seq) analysis. Driver alterations were successfully determined for all except for 10 indeterminate samples and grouped into four mutational subgroups based on their genetic drivers: *BRAF* p.V600E, KFs (primarily involving *RET*, *NTRK1/3,* and *ALK*), *DICER1* (RNase IIIb hotspot mutations), and *RAS*-like mutations (mostly involving *N-H-KRAS*, *BRAF* non-V600E, *PTEN,* and *TSHR*). The reference cohort comprised 22 patients (25.2%) with *BRAF* p.V600E, 34 patients (39.1%) with KFs, 20 patients (23.0%) with *RAS*-like mutations, eight patients (9.2%) with *DICER1* mutations, and three patients (3.4%) with indeterminate driver mutations. The validation cohort comprised 20 patients (24.1%) with *BRAF* p.V600E, 45 patients (54.2%) with KFs, nine patients (10.8%) with *RAS*-like mutations, two patients (2.4%) with *DICER1* mutations, and seven patients (8.4%) with indeterminate driver mutations.

For comparison, adult thyroid cancer data (TCGA-THCA) were downloaded using TCGAbiolink (RRID:SCR_017683; ref. 36), consisting of 499 PTC samples profiled on the HM450 array, including primary ($n = 449$) and paired LN metastases ($n = 50$). The study cohort consisted of 325 females (72%) and 124 (28%) males. Patient ages ranged from 15 to 89 years (mean 47.15, standard deviation 15.59). Primary tumors were similarly classified by nodal involvement, with 225 (50%) low- and 224 (49%) high-invasive samples. The cohort was also classified by genetic driver mutations and comprised of 265 patients (59%) with *BRAF* p.V600E, 60 (13.4%) patients with KFs, 84 patients (18.7%) with *RAS*-like mutations, two patients (0.4%) with *DICER1*, and 38 patients (8.5%) with indeterminate driver mutations, representing the dominance of *BRAF* v.600E drivers in adult PTC (Supplementary Table S4; ref. 20).

Data preprocessing

Raw methylation data from the reference and validation cohorts were obtained from IDAT files, normalized, and methylation values were jointly calculated using the R package *SeSAMe's* standard preprocessing pipeline (35). Samples with discordant inferred and clinical sex were excluded. After excluding sex chromosome probes ($n = 24{,}953$) and retaining only CpG-targeting probes ($n = 908{,}400$), methylation values were collapsed to probe prefixes, and 902,052 probes were retained for downstream analysis.

Samples with probe success rates (PSR) < 0.70 were excluded (Supplementary Table S2). The validation cohort [median: 0.892 (IQR: 0.863–0.9190)] showed significantly lower PSR compared with the reference cohort [median: 0.948 (IQR: 0.932–0.957); Wilcoxon test, $P = 2.2e{-}16$], which may be attributable to DNA damage from higher proportion of FFPE samples, known to degrade DNA through formaldehyde cross-linking, fragmentation, and other processing artifacts (Supplementary Fig. S1B; ref. 37). Thus, the validation cohort was retained to represent the technical and biological variation representative of real-world conditions.

For the joint analysis of pediatric and adult thyroid carcinoma samples, preprocessing of the adult methylation data was performed similarly as described above, excluding sex chromosomes ($n = 11{,}661$) and retaining only CpG-targeting probes ($n = 470{,}865$). Pediatric (EPICv2) and adult (HM450) methylation array platform data were subset to overlapping probes ($n = 382{,}015$), after which they were merged into a unified matrix of 683 samples across all primary and LN samples. Subsequent integrative analysis was performed on this integrated dataset.

Unsupervised analysis and consensus clustering

Principal component analysis (PCA) was performed on the top 30,000 most variable CpG sites for each cohort. The basic assumption is that probes with higher variability contribute more to the clustering process. t-distributed stochastic neighbor embedding (t-SNE) visualization was performed on PCA-transformed data.

To identify stable methylation-based clusters, consensus clustering was performed on all reference and validation cohort samples using the *ConsensusClusterPlus* R package (RRID: SCR_016954; ref. 38). For every candidate cluster number tested ($k = 2$–10), hierarchical clustering was performed and pruned into k clusters (80% of features, 1,000 iterations, Euclidean distance). Pairwise consensus values, the proportion of runs in which two samples clustered together, were calculated and stored in consensus matrices for each run. To determine the optimal number of clusters, the cumulative distribution function (CDF) for each candidate cluster number k was plotted against the consensus value index, which represents the fraction of sample pairs that have consensus values greater than or equal to that particular consensus value. Change in area under the CDF was plotted against each candidate k. The final consensus cluster assignments were determined by hierarchical clustering of the consensus matrix distance (1 - consensus values) and pruning to the optimal k clusters.

Cell type deconvolution

Cell type deconvolution was performed on methylation data to infer the stromal composition using the *EpiDISH* R package (39). Solid tissue-type inference was performed to deconvolve the samples into broad cell type components: epithelial, fibroblast, and immune cell (reference matrix: centEpiFibIC.m), and the immune cell component was further deconvolved into seven immune cell subtypes (centBloodSub.m). The immune cell fraction was denoted as the estimated leukocyte fraction for subsequent analyses. Hierarchical clustering was performed on the inferred cell type fractions. microRNA 200c (*miR*-200c) expression was estimated using the *CytoMethIC* R package.

Differential methylation and enrichment analysis

CpG loci associated with tumor invasiveness were identified through differential methylation using the *sesame* modeling pipeline (35). DNA methylation β values were fitted to a linear model, and corresponding slope tests and goodness-of-fit tests (F-tests holding out each contrast variable) were performed to evaluate the significance of differences in DNA methylation levels. Differentially methylated loci (DML) were selected based on β coefficient (>0.2 or <−0.2) and Benjamini–Hochberg (BH)-adjusted *P* value (<0.05). The *knowYourCG* pipeline (40) was used to test the enrichment of DML across curated biological and technical databases, including chromatin states, gene association, transcription factor binding sites (TFBS), and more. Fischer exact test was conducted to determine whether a set of CpGs was enriched in certain categories or features. Significantly enriched databases were plotted.

Integrated DNA methylation and RNA expression analysis

RNA-seq data ($\log_2$CPM) from primary reference cohort samples were analyzed with the *limma* R package (three samples missing data, $n = 84$, RRID:SCR_010943; ref. 41). Differential expression between high- and low-invasive samples was modeled with a linear model adjusted by sex and leukocyte fraction, followed by Empirical Bayes and BH correction. DML were annotated to genes (hg38) and summarized per gene by median invasiveness-associated β coefficient. Gene-level methylation and expression results were integrated to identify activated (hypomethylation with increased expression) or silenced (hypermethylation with decreased expression). Gene set enrichment analysis was performed using the *fgsea* R package (RRID: SCR_020938) using signed metric ($-\log_{10}(\text{pval}) \times \log_2$ fold change).

Epigenetic clock analysis

Epigenetic clock analysis has become a valuable method for estimating biological age by examining patterns of DNA methylation. This study applied age regression techniques to samples using the Horvath multi-tissue model from the *methylClock* R package (42). The data were processed and verified for compatibility with the pediatric (EPICv2) and adult (HM450) arrays (43). Epigenetic age acceleration (EAA) was estimated as the residual value from a linear regression model of epigenetic age on chronologic age (age at DNA sampling). Models were adjusted for sex and estimated leukocyte fraction. Statistical comparison between groups was performed using Wilcoxon rank-sum tests to assess pairwise differences in methylation distributions.

Random Forest classifier development and evaluation

For the reference cohort, a leave-one-out cross-validation (LOOCV) approach was implemented using primary tumor samples to train classifiers predicting tumor invasiveness as previously defined by nodal involvement ($n = 87$, classes: high-, low-invasive) and driver mutation group excluding indeterminate samples ($n = 84$, classes: *BRAF* p.V600E, *RAS*-like, KF, *DICER1*). In each fold, one training sample was held out for testing, whereas the remaining samples were used for training. Feature selection was performed by training multiple Random Forest (RF) classifiers on subsets of 10,000 probes (n_trees = 500). Feature importance was determined by the permutation-based variable importance measure from the *randomForest* R package. Importance scores from all classifiers were aggregated, and the top 3,000 most important features were used to train the final RF model using all training samples for that fold, which was used to predict the label of the one held-out test sample. Prediction confidence was defined as the RF prediction probability associated with the predicted label. Performance metrics, including error, accuracy, and area under the receiver operating curve (AUC), were calculated from the LOOCV

predictions. For the multiclass driver mutation classifier, AUC was calculated in a one-vs-all approach for each class. SHapley Additive exPlanations (SHAP) values were calculated from the final model using the *TreeSHAP* R package and visualized with *shapviz* (44). The final classifiers were then applied to the validation cohort and LN metastases, and performance metrics and predictions were recorded.

Results

DNA methylation distinctly encodes the invasiveness of pediatric thyroid carcinoma

Unsupervised clustering analysis of the tumor methylation profiles was performed to explore the global methylome landscape of pediatric thyroid samples (**Fig. 1B–I**). An exploratory t-SNE visualization of the reference cohort's 87 primary samples identified three distinct clusters. To validate the robustness of the cluster assignments, we performed consensus clustering and determined the optimal number of clusters by examining the cumulative distribution plot (CDF) (Supplementary Fig. S1C–S1F). The peak in the delta area under the CDF at $k = 3$ indicated diminishing returns with additional clusters (Supplementary Fig. S1D), and the pairwise consensus matrix showed strong intracluster and minimum cross-cluster agreement (Supplementary Fig. S1E).

The clusters were denoted (i) LI: a cluster mainly comprised of low-invasive samples ($n = 26$); (ii) HI: a cluster mainly comprised of high-invasive samples ($n = 54$); (iii) leukocyte-infiltrated (LEUKO): a smaller cluster comprised of samples with high leukocyte infiltration ($n = 7$; **Fig. 1B**). Clusters showed little segregation by sex or age, as sex chromosome CpGs were excluded from analysis (**Fig. 1C** and **D**). Although clustering generally aligned with clinical invasiveness, some samples diverged from their assigned clusters, likely due to limitations in surgical assessment, lymphocyte infiltration, and subclonality (see discussion below).

We reviewed the available histopathologic and molecular data associated with clusters as defined by methylation profiles. Pathologically, the LI cluster includes mainly T1a, T1b, T2, and T3/3a samples, with a predominance of T2 tumors. The LEUKO cluster consists of T1a, T1b, and T2 samples. The HI cluster spans all stages from T0 to T4/4a, with T4/T4a and M1 cases exclusive to this group (**Fig. 1E–G**). Molecularly, two distinct subclusters emerged within the HI cluster, defined by the *BRAF* p.V600E and KF mutational groups, respectively (**Fig. 1H**). Similarly, the LI cluster is further divided into two subclusters, each defined by *RAS*-like and *DICER1* mutations. The LEUKO cluster is composed primarily of KF tumors and two indeterminate driver mutation samples, suggesting a decoupling from mutational signatures. Furthermore, our reference cohort showed a relatively copy-number–quiet profile with few focal alterations, lacking recurrent *CDKN2A/B* deletions or focal amplifications of receptor tyrosine kinases and immune-evasion genes reported in adult anaplastic thyroid cancers (45), consistent with differences in patient age and disease stage.

We denoted the LEUKO cluster due to its high leukocyte fraction (**Fig. 1I**; see discussion below; Supplementary Fig. S2A), as evidenced by the expression of mesenchymal markers such as *miR-200c* (Supplementary Fig. S1G–S1I; ref. 46). Indeed, all seven high-invasive PTC cases within this cluster had chronic lymphocytic thyroiditis, characterized by marked infiltration of immune cells. Four of these harbor KF mutations (0149, 0334, 0201, and 0164), one carries a *BRAF* p.V600E mutation (0261), and the remaining two have unknown drivers: one is a tall-cell variant (0138) with low invasiveness, and the other is a papillary thyroid microcarcinoma (0170B) with high-invasive behavior.

Differential methylation links invasiveness to disrupted thyroid development and function

To define the epigenetic hallmark of invasiveness in the pediatric thyroid carcinoma, we first performed differential methylation analysis of the three previously defined clusters in the reference cohort. Differentially methylated CpG loci (DMLs) associated with invasiveness were identified using the following criteria: (i) BH FDR-corrected P value <0.05 and (ii) β coefficient >0.2 (hypermethylated) or < -0.2 (hypomethylated).

LEUKO cluster samples exhibit higher global methylation and display extensive differential methylation, with 83,726 DMLs relative to the HI cluster and 60,290 DMLs relative to the LI cluster (Supplementary Fig. S2A and S2B). These differences largely reflect leukocyte-derived methylation signatures. Accordingly, LEUKO-associated DMLs were enriched for cell-type signatures of lymphocytes, including T and B cells (Supplementary Fig. S2C and S2D), as well as TFBSs involved in lymphoid development and hematopoiesis, such as RUNX3, LYL1, LMO2, ETV6, PBX1, TLX1, ZBTB16, and LMO1 (Supplementary Fig. S2E and S2F). Finally, cell-type deconvolution confirmed that LEUKO samples cluster together due to substantial immune infiltration, particularly CD4 T cells and B cells (**Fig. 2A**). Together, these findings indicate that immune cell–driven epigenetic alterations dominate the LEUKO methylome, dwarfing methylation differences associated with tumor invasiveness observed between the HI and LI groups (19,125 DMLs; Supplementary Fig. S2A).

To dissect invasiveness-associated changes from immune cell–driven alterations, we then performed a regression analysis in which DNA methylation was treated as the response variable and invasiveness as the primary predictor, adjusted for sex and leukocyte fraction. Low-invasive samples have higher global methylation levels than high-invasive samples (Wilcoxon test, $P = 2.2e{-}3$). Compared with low-invasive cases, high-invasive tumors have 4,706 significantly hypomethylated and 150 hypermethylated CpGs (**Fig. 2B** and **C**). This is primarily driven by *RAS*-like tumors, as *DICER1* tumors have a comparable global methylation level with the high-invasive tumors (Supplementary Fig. S2B).

In high-invasive tumors, *hypomethylated* loci were enriched for binding sites of thyroid lineage-determining and oncogenic TFs, including NKX2-1/TTF1, YAP, and TEAD1 (Hippo pathway) and AP-1 components FOS and JUN (**Fig. 2D**; Supplementary Fig. S2E and S2F). The enrichment of NKX2-1 binding sites links invasive behavior to disruption of thyroid tissue development and lineage regulation, a common mechanism in tumorigenesis (47). Concurrent enrichment of YAP and TEAD1 sites implicates activation of Hippo signaling in invasive progression, consistent with its role in promoting epithelial plasticity and growth (48, 49). Similarly, hypomethylation at AP-1 binding sites suggests engagement of stress-responsive transcriptional programs that facilitate invasion (50).

In contrast, *hypermethylated* loci in high-invasive tumors showed minimal enrichment of TFBSs overall, with the androgen receptor (AR) binding representing the only notable signal (**Fig. 2D**; Supplementary Fig. S2E and S2F). The enrichment at AR-associated sites is consistent with prior evidence that hormone signaling can modulate tumor behavior (51) and that preserved AR activity may constrain the invasive potential of carcinomas (52, 53). Together, these epigenetic alteration patterns associated with invasive progression reflect the disruption of lineage programming and

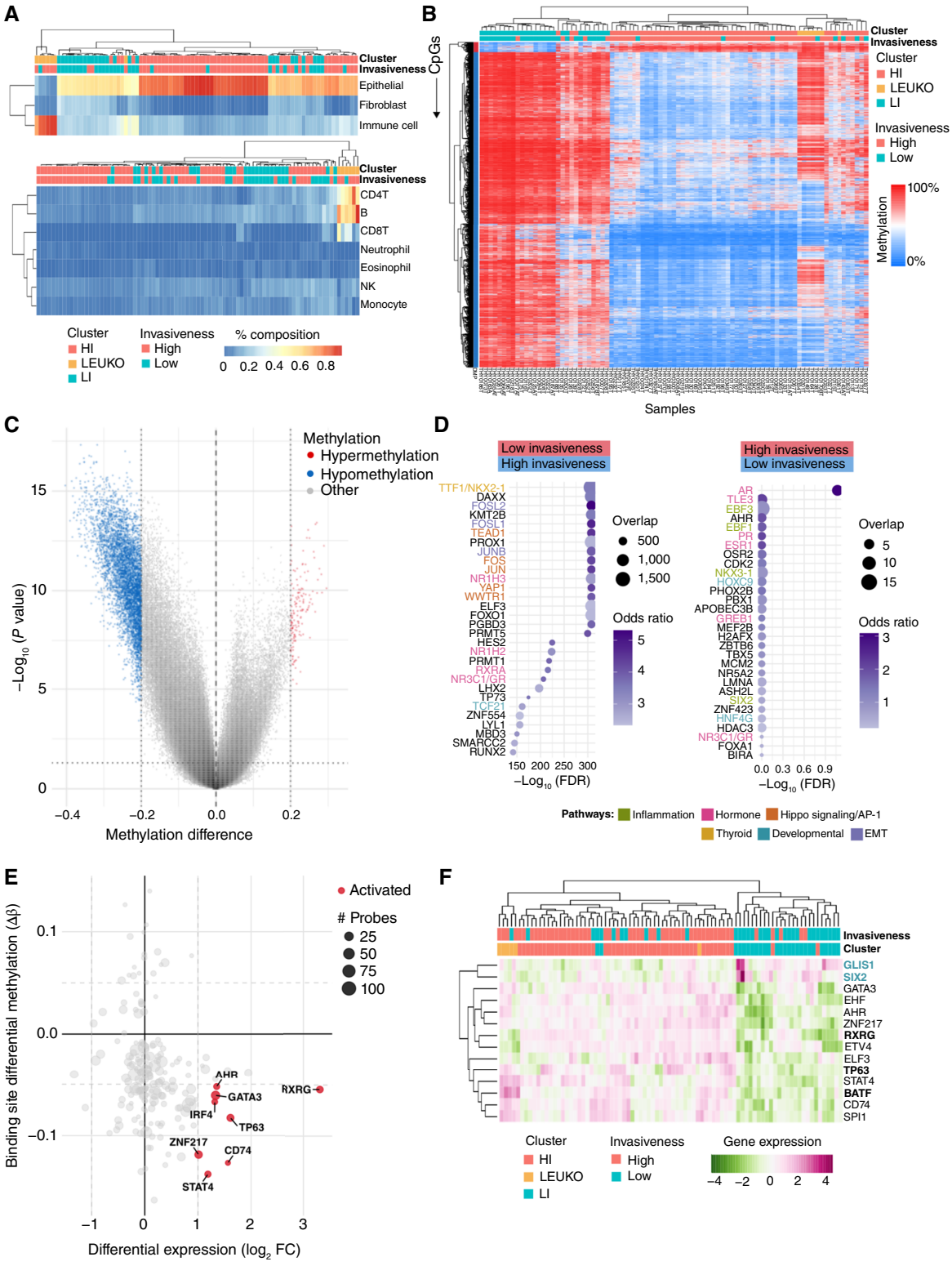

Figure 2.

Epigenome and transcriptome show distinct immune-related signatures in high-invasive thyroid carcinoma. **A,** Unsupervised hierarchical clustering of deconvoluted cell type fractions in reference cohort primary tumors ($n = 87$). Top, major cell type fractions (epithelial, fibroblast, and immune cell). Bottom, immune cell infiltration across seven immune subtypes. **B,** Heatmap of 4,856 significant DMLs comparing methylation levels (β) (*Continued on the following page.*)

activation of oncogenic transcriptional programs, alongside the repression of hormone-associated regulatory elements that may otherwise restrain invasion.

DNA methylation alteration at the binding site of TFs is often observed in conjunction with the upregulation of the TF's own expression, as shown by integrated methylation expression analysis (**Fig. 2E**) and expression heatmaps (**Fig. 2F**). High-invasive PTC-specific TFs include inflammatory and immune-associated TFs, such as *GATA3*, *STAT4*, *SPI1*, and *CD74*, together with stress- and invasion-linked epithelial regulators, including *TP63*, *RXRG*, *AHR*, *ETV4,* and *ELF3*, again suggesting an inflammation-coupled epithelial plasticity program (Supplementary Fig. 2H), consistent with the literature (54–56) and more recent evidence (57). On the contrary, low-invasive tumors retain higher expression of differentiation-associated TFs (GLIS2 and SIX2), consistent with a more lineage-constrained state; GLIS1 is broadly downregulated in adult cancers, supporting its association with less aggressive biology (58). The gene-level integration of DNA methylation and RNA transcription revealed coordinated regulation and function of TFs linked to thyroid carcinoma invasiveness.

Pediatric and adult thyroid carcinomas exhibit distinct epigenetic aging signatures

Pediatric and adult PTCs differ in mutation spectra and likely tumorigenic mechanisms (4), but their epigenomic differences remain less understood. A joint t-SNE embedding of methylation profiles from pediatric and adult cohorts revealed that global methylation profiles show varying degrees of separation based on invasiveness (**Fig. 3A**), genetic subtypes (**Fig. 3B**), age groups (**Fig. 3A–D**), and leukocyte infiltration (**Fig. 3C**). The frequency of molecular alterations differed by age, with *BRAF* p.V600E mutations observed more commonly in adult thyroid cancer (**Fig. 3B**). Whereas pediatric *BRAF* p.V600E cases are all highly invasive, adult *BRAF* p.V600E cases are mixed in invasiveness. *RAS*-like and *DICER1* mutation–driven samples clustered separately, exhibiting relatively less separation between pediatric and adult groups, suggesting that the epigenome changes driven by these mutations predominate over age-related differences. In contrast, adult KFs and *BRAF* p.V600E samples showed a more dispersed distribution, suggesting greater epigenetic heterogeneity than pediatric cases, potentially due to stochastic epigenetic drift with age (**Fig. 3B**).

The epigenome can be used to inform estimates of chronologic and biological aging, particularly the accelerated aging observed in cancer (59, 60). Using established multi-tissue epigenetic clocks (42), we inferred epigenetic ages of thyroid carcinomas (**Fig. 3E**). In the pediatric cohort, the median inferred epigenetic age was 22.9 years [IQR: (18–31.6)], significantly higher than the actual median age at surgery of 15.3 years [IQR: (14–17.3)]. Similarly, in the adult cohort, the median inferred epigenetic age was 57.4 years [IQR: (45.3–67.9)], compared with an actual median age of 46 years [IQR: (35–58); Supplementary Table S2]. These findings highlight significant EAA in both groups, consistent with cancer's extended proliferative history (60).

Despite a younger age at surgery, pediatric and adult groups show similar absolute EAA, measured by regression residuals, suggesting a disproportionately greater acceleration in pediatric cases relative to age at diagnosis (**Fig. 3E** and **F**). Interestingly, in pediatric cases, high-invasive tumors exhibit slightly greater age acceleration than low-invasive tumors (Wilcoxon test: $P = 0.039$); a similar trend is observed in adult thyroid tumors but does not reach statistical significance (**Fig. 3F**). Our data indicate that age acceleration is likely reflective of intrinsic tumor biology, e.g., cell proliferation and invasiveness, relatively independent of the age of carcinoma initiation.

Machine learning models to stratify high- and low-invasive thyroid carcinomas

Having established the methylome differences linked to clinical invasiveness and driver mutations, we next evaluated whether DNA methylation profiles can inform clinical stratification. We developed RF classifiers to predict clinical invasiveness, as defined by LN metastasis, and driver mutation classes from genome-wide CpG methylation profiles of primary thyroid carcinoma samples of the reference cohort, excluding indeterminate samples from driver classifier (invasiveness: $n = 87$, driver: $n = 84$; **Fig. 4A**). Applying a LOOCV approach, the invasiveness classifier achieved an overall accuracy of 84% (73/87), whereas the driver mutation classifier attained an overall accuracy of 95% (80/84), correctly predicting 95.5% (21/22) of *BRAF* p.V600E, 100% (8/8) of *DICER1*, 94.1% (32/34) of KFs, and 95% (19/20) of *RAS*-like cases (**Fig. 4B**). Applying the final models on the validation cohort (invasiveness: $n = 83$; driver: $n = 76$, excluding indeterminate) yielded slightly lower performance with 77.1% (64/83; **Fig. 4C**) for the invasiveness classifier. The driver mutation classifier attained an overall accuracy of 82.9% accuracy (63/76), correctly predicting 60% (12/20) of *BRAF* p.V600E, 50% (1/2) of *DICER1*, 86.7% (39/45) of KFs, and 100% (9/9) of *RAS*-like cases (**Fig. 4C**). Receiver operating characteristic analyses confirmed strong discriminatory power across both cohorts, with AUCs exceeding 0.81 for invasiveness and ranging from 0.92 to 1 for driver mutation categories (Supplementary Fig. S4A and S4B).

Analysis of misclassifications across both cohorts revealed biologically meaningful patterns that could improve risk stratification. The two classifiers often misclassified the same cases with predicted oncogenic driver status aligned with genomic characterization, potentially offering clinically relevant leads for follow-up. The majority of invasiveness misclassifications involved clinically low-invasive tumors predicted as high-invasive, which often harbored aggressive oncogenic drivers such as *BRAF* p.V600E or KFs. In the reference cohort, six of nine low-to-high misclassifications carried KFs ($n = 4$) or *BRAF* p.V600E ($n = 2$) mutations (**Fig. 4D** and **E**). Several validation samples also demonstrated this pattern, with ten of sixteen misclassifications harboring *BRAF* p.V600E ($n = 3$) or KF

(*Continued*.) between low-invasive and high-invasive primary tumors (P adj <0.05). **C,** Volcano plot of DML showing hypomethylated (blue, $\Delta\beta < -0.2$) and hypermethylated (red, $\Delta\beta > 0.2$) loci in high-invasive tumors (P adj <0.05). **D,** TFBS enrichment analysis of DMLs showing hypomethylated sites (left) and hypermethylated sites (right) in high-invasive samples, ordered by significance [$-\log_{10}$(FDR)]. Dot size indicates size of database overlap; color intensity indicates enrichment strength (estimate). EMT, epithelial–mesenchymal transition. **E,** Integrated methylation–expression analysis plotting median differential methylation ($\Delta\beta$) against differentially expression of corresponding TF in high-invasive samples (red: activated TF with hypomethylated binding sites; gray: not significant; FDR <0.05). Size indicates the number of aggregated CpG probes per TFBS. FC, fold change. **F,** Heatmap of expression levels for the 14 most differentially expressed TFBSs in high-invasive samples with unsupervised hierarchical clustering.

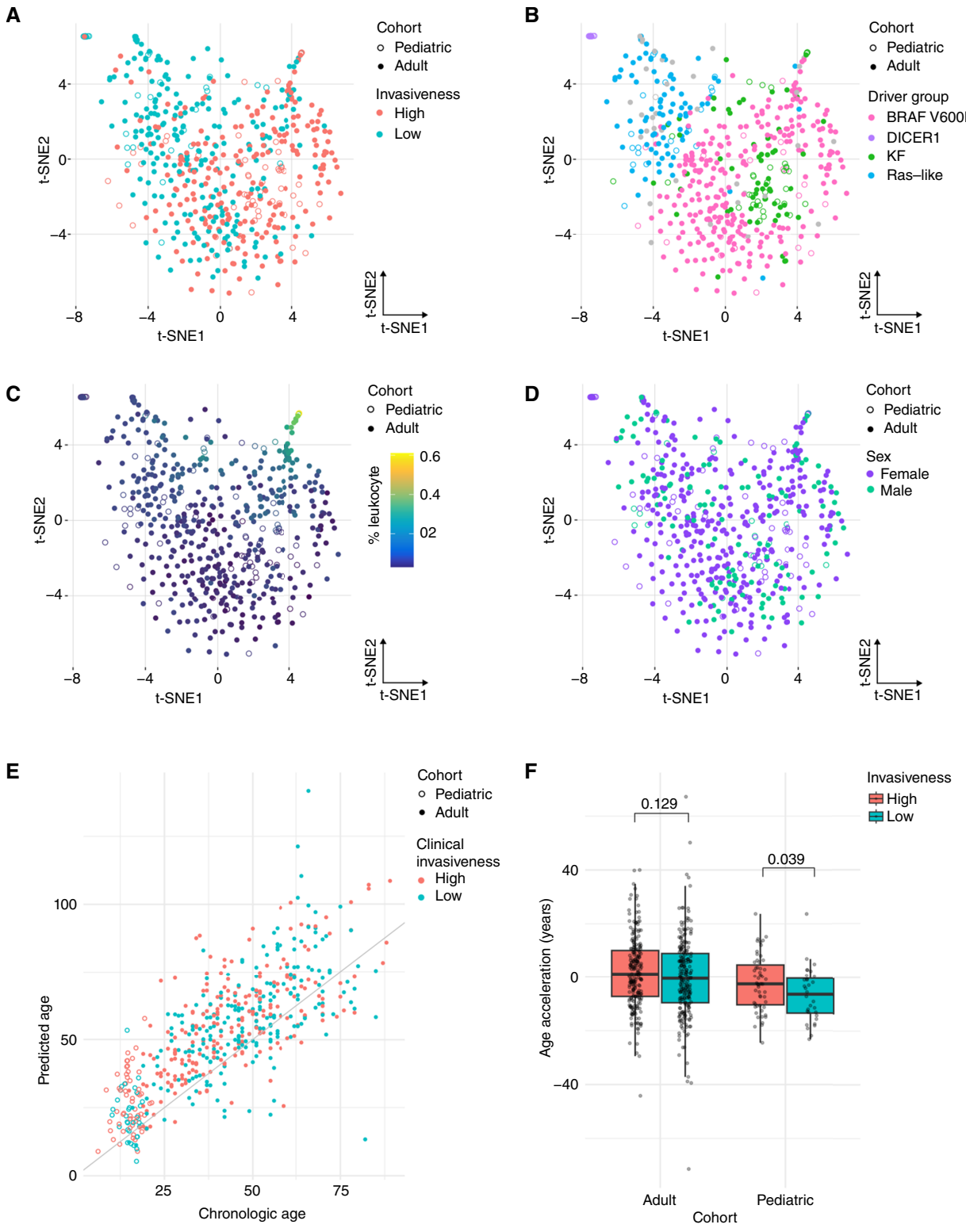

Figure 3.

Integrative analysis of pediatric and adult thyroid carcinoma epigenomes. **A–D,** t-SNE dimensionality reduction of primary samples from the pediatric reference cohort ($n = 87$; open circles) and adult cohort ($n = 449$; filled circles) colored by (**A**) clinical invasiveness, (**B**) genetic driver mutations, (**C**) estimated leukocyte fraction, and (**D**) sex. **E,** Correlation between inferred epigenetic age and chronologic age, colored by clinical invasiveness. The diagonal line indicates where predicted age equals chronologic age. **F,** EAA comparison between pediatric and adult cohorts, stratified by clinical invasiveness (Wilcoxon test: pediatric $P = 0.039$, adult $P = 0.129$).

mutations ($n = 7$; Supplementary Fig. S4C and S4D). Given the congruence between the methylation classifications and oncogenic driver status, the low invasiveness clinical designation may reflect undetected or future metastatic potential, warranting heightened clinical surveillance for such cases. The remaining misclassifications often involved a second comutation. For example, the *RAS*-like, low-invasive case 0259 in the validation cohort is misclassified as high-invasive. This case harbored a cooperating KRAS and inactivating KEAP1 mutation, associated with the upregulation of KEAP1/NRF2 target genes. Such comutations, known to promote tumor progression in other cancers (61), may explain its high-invasive classification despite being clinically low-invasive.

Similarly, among the clinically reported high-invasive tumors that were misclassified as low-invasive, three in five misclassifications in the reference cohort (0135, 0172, and 0208) carried *RAS*-like mutations and clustered with low-invasive profiles, despite having high-invasive clinical annotations. Interestingly, several of these misclassifications were clinically N1b but involved fewer than five LNs in the central neck, suggesting limited spread and representing borderline cases in our classification threshold (reference: 0208; validation: 0372, 0868; **Fig. 4D** and **E**; Supplementary Fig. S4C and S4D).

Besides classification performance, prediction confidence may be informative. For invasiveness classification, prediction confidence was significantly higher for correct predictions in both the reference and validation cohorts (one-sided Wilcoxon test, invasiveness: $P = 0.019$, $P = 0.003$; Supplementary Fig. S4E), with many misclassified samples located at cluster boundaries (**Fig. 4D** and **E**; Supplementary Fig. S4C and S4D), suggesting ambiguous or transitional methylation states. Furthermore, the LEUKO cluster exhibits lower confidence scores (median = 0.679) compared with the HI (median = 0.924) and LI (median = 0.928) clusters. This suggests that leukocyte contamination does affect the performance of our classifiers, as expected.

DNA methylation refines the classification of pediatric thyroid carcinoma invasiveness beyond driver mutations

Despite the overall concordance between methylation and mutation-based classification, methylation classifiers may refine invasiveness predictions beyond conventional driver mutation associations. For example, case 0033, harboring *RAS*-like mutations typically linked to low-invasiveness, was clinically characterized as high-invasive. It clustered with HI and was predicted accordingly (**Fig. 4D** and **E**; Supplementary Fig. S1F). Further examination revealed that this case carried complex mutations of *NRAS* p.Q61 K and *TP53* p.R273C, suggesting a more aggressive profile from comutations. On the other hand, the KF case 0077A, despite carrying a TG::IGF1R fusion, is correctly predicted to be of low-invasiveness by the methylation classifier and placed in the LI cluster (**Fig. 4D** and **E**; Supplementary Fig. S1F). This discrepancy between the driver mutation and DNA methylation classification suggests a potential distinction between this fusion and other KFs. Indeed, studies in thyroid cells indicate that *IGF1* signals preferentially via the PI3K pathway, as opposed to the usual MAPK pathway activation observed in other KFs (62).

Interestingly, the methylation classifier can also resolve some inconsistencies between clinical annotation and global methylation cluster assignment. For example, case 0126, carrying a subclonal *BRAF* p.V600E mutation with limited LN involvement (2/15 positive LN from central neck), was clinically labeled high-invasive (N1a) but showed an overall methylome profile resembling less invasive tumors. Despite this, the methylation classifier correctly identified it as high-invasive, indicating the classifier's sensitivity to aggressive subclones. Another case (0171) with a *BRAF* p.V600E mutation was reported clinically as low-invasive and located in the periphery of the HI cluster. It was correctly classified as low-invasive, though with borderline confidence (0.54). In the LEUKO cluster, sample 0138, harboring an indeterminate driver mutation and tall cell histology, was clinically reported as low-invasive but classified as high-invasive (**Fig. 4D** and **E**).

Together, the above case studies suggest that methylation profiles can detect subtle differences in invasiveness that are potentially overlooked by driver mutation groups or similarities in global methylation profiles.

DNA methylation classifier robustly predicts driver mutations from LN metastasis

To test the generalizability of our classifiers, we analyzed the DNA methylome profiles of matched LN metastases across both reference and validation cohorts**.** Thirteen patients had matched LN metastases samples, with case 0212 having two matched LN metastases (one axial; Supplementary Table S1). In the global methylome clustering, LN metastases maintain a close relationship with their matched primary tumors (**Fig. 4F**). Consistently, all had the same genetic driver predicted as was reported in the primary tumor. The driver classifier accurately predicted 13/14 (92.8%) driver mutations in the matched LN metastasis samples based on their methylome profiles. The sole misclassification occurred in case 0868, a high-invasive, FTC *DICER1* case, which was incorrectly classified as *RAS*-like. Generally, the methylome of LN metastases clustered with their matched primary tumors, with the exception of case 0212, which deviated due to significant lymphocyte infiltration (0.57 leukocyte fraction; **Fig. 4F**; Supplementary Fig. S2A). The classifier correctly predicted the driver mutation of this case despite the deviation.

Together, these results underscore the utility of methylation-based classifiers in predicting both tumor invasiveness and oncogenic genetic alterations. Methylation profiles can not only recapitulate known molecular subtypes but may also resolve discrepancies between genotype and phenotype, supporting their potential clinical utility for triaging tumor aggressiveness.

DNA methylation classifier interpretation reveals divergent epigenetic signatures

Finally, we interpreted the features used in the machine learning models for biological insights. First, feature importance analysis revealed that invasiveness and driver classifiers relied on distinct sets of CpG sites, though some overlapping features contributed to both tasks with moderate importance (Supplementary Fig. S4F). Enrichment of the top 3,000 predictive CpG features ranked by mean decrease in accuracy index uncovered Hippo signaling and thyroid tissue development, recapitulating findings from differential methylation analysis (**Fig. 4G**). The results aligned closely with previous enrichment results of hypermethylated sites from differential methylation analysis between low-invasive and high-invasive tumors in both adult (Supplementary Fig. S3F) and pediatric (**Fig. 2D**) cohorts, suggesting conservation of these regulatory mechanisms across age groups.

To provide a more interpretable, sample-level view of feature influence, we performed a SHAP analysis to explore heterogeneity in the contribution of methylation features across different clinical and mutation groups. Targeting final models trained on all primary

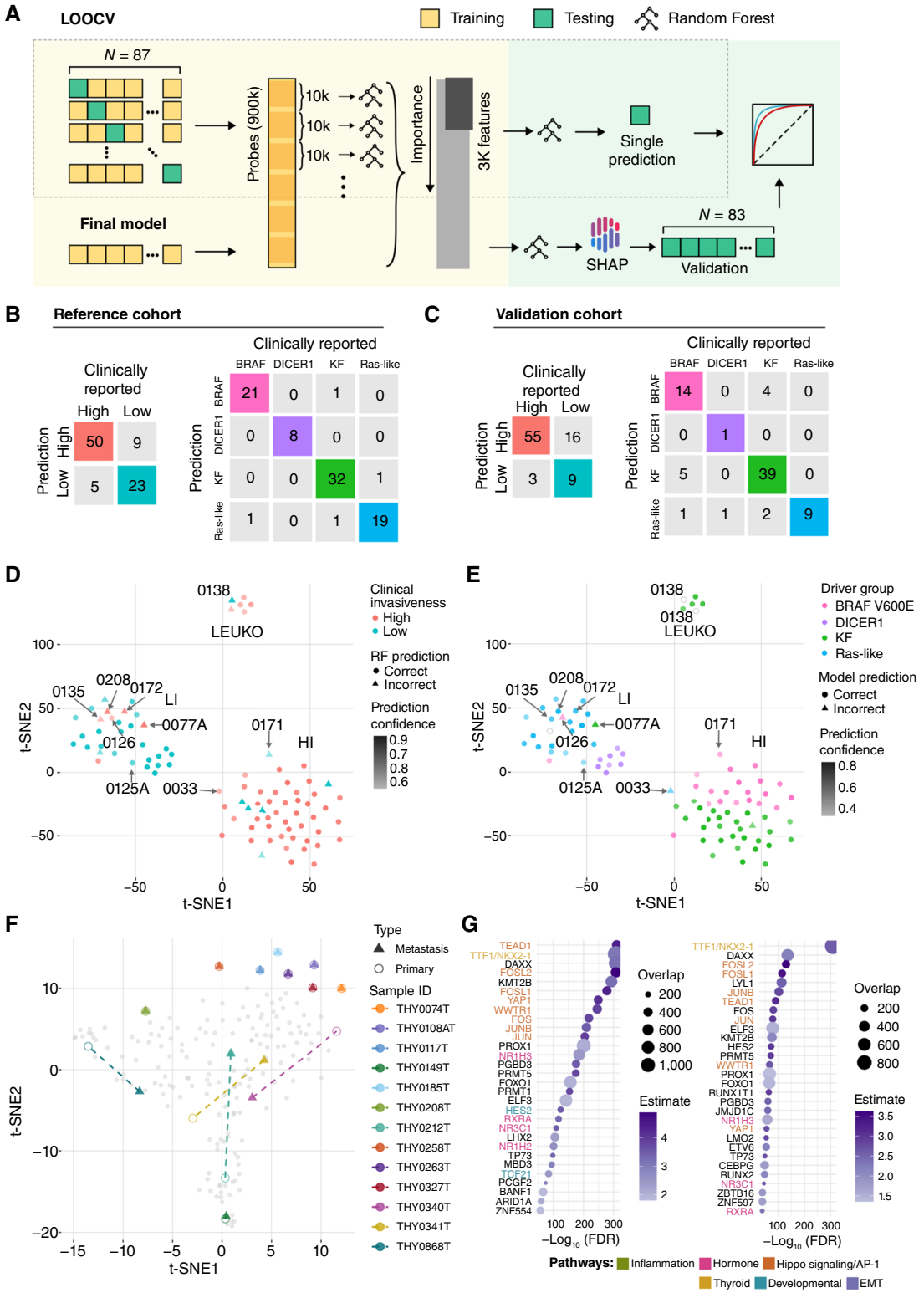

Figure 4.

Construction of classification models for pediatric thyroid carcinoma invasiveness. **A,** Schematic of the RF classifier development and validation pipeline for predicting clinical invasiveness and driver mutation groups. Top, LOOCV was performed on primary samples from the pediatric reference cohort using iterative training on subsets with feature selection. Bottom, final models were trained on the reference cohort, with SHAP analysis (*Continued on the following page.*)

tumor samples from the reference cohort, we showed that high-invasive predictions were primarily driven by hypomethylation features, consistent with global hypomethylation patterns observed in aggressive cases (Supplementary Fig. S4G). In contrast, prediction patterns for driver mutations were more complex, as revealed by analyzing one-vs-rest classifiers. Both hypomethylation and hypermethylation features contributed to *BRAF* p.V600E and *DICER1* predictions (**Fig. 4G**), whereas KF and *RAS*-like mutation predictions were more consistently driven by hypomethylation and hypermethylation, respectively (Supplementary Fig. S4G). Methylation alterations that contribute to *DICER1* predictions are more limited and involve gene-specific methylation alterations, such as in genes regulated by miRNAs, consistent with DICER1's function as a pre-miRNA processor (63). Overall, driver classifiers exhibit more complex patterns, suggesting stronger individual feature contributions.

Discussion

We presented and validated DNA methylation–based classifiers of tumor invasiveness and oncogenic driver in pediatric thyroid carcinoma. Pediatric tumors differ fundamentally from adult PTC in their mutational profiles, clinical behavior, and treatment challenges; yet, they have been largely excluded from prior methylome studies. Rather than focusing on diagnosis or histologic subtype, our model targets invasiveness, a clinically actionable feature directly informing surgical decision-making and outcomes.

Our study revealed that pediatric thyroid tumors exhibit a diverse methylation landscape. High-invasive tumors are characterized by widespread hypomethylation and focal increases in methylation, which may reflect differences in the degree of developmental disruption, proliferation, and inflammatory stress. Invasiveness-associated DMLs were enriched at TFBSs for nuclear receptors (AR), thyroid-specific TF such as *TTF1/NKX2-1*, and regulators of Hippo (64) and AP1 signaling (**Fig. 2D**; ref. 65). Notably, methylation gain at the *TTF1/NKX2-1* locus has been linked to TF binding loss and poor prognosis (66), and our findings align with the suppression of the androgen–AR axis as a methylation-mediated mechanism of aggressive PTC (67). Moreover, AR transactivation drives metastasis via the WNT pathway in prostate cancer (68), pointing to common epigenetic mechanisms across hormone receptor–involved carcinomas. The epigenetic separation by invasiveness may also reflect differences in differentiation, as more invasive tumors tend to exhibit lower thyroid differentiation scores, lower BRAF-RAS scores (BRS, reflecting a more BRAF-like state), and higher ERK scores (indicating elevated MAPK pathway activity), consistent with HI behavior (Supplementary Fig. S3C–S3E).

Driver mutations are a primary force in shaping global methylation profiles, reflecting distinct tumorigenic trajectories and regulatory programs. *RAS*-like samples, associated with low-invasiveness, exhibit significantly higher methylation levels than other mutational groups (69, 70). In contrast, more invasive subtypes such as *BRAF* p.V600E and KF exhibit widespread hypomethylation (69, 70). These observed methylation differences reflect downstream effects of the *BRAF*-like and *RAS*-like gene expression subtypes, driven by MAPK/ERK and PI3K/AKT signaling, respectively (17). Despite sharing lower global methylation averages with the high-invasive tumors, *DICER1* samples are more similar to the *RAS*-like samples in the whole methylation profile and distinct from the high-invasive tumors. This is consistent with their shared low invasiveness and a previous report of *DICER1* samples exhibiting *RAS*-like transcriptomic patterns (71). Notably, a case of FTC (0125A) harboring the *DGCR8* hotspot mutation p.E518 K clustered closely with other *DICER1* tumors in terms of DNA methylation (**Fig. 4E**; refs. 72, 73), suggesting convergent disruptions in miRNA regulation.

Our methylation-based invasiveness classifier achieved 84% accuracy in the reference cohort and 77% concordance in the validation cohort with clinically reported invasiveness, confirming the expected clinical–molecular correlation in most cases. Of the samples whose classifications did not match their clinically reported invasiveness labels, 75% of these cases were supported by driver mutations, whereas the remaining 25% reflected borderline cases. Furthermore, the model correctly classified driver mutations in all LN metastasis samples, supporting prior evidence that most carcinoma-related methylation changes often precede metastasis (74). Collectively, these findings suggest that methylomes have high predictive value for both molecular characterization and clinical presentation at the time of surgery.

As mentioned, methylation profiling resolves biologically meaningful heterogeneity among genotypically similar tumors. In both the primary and validation cohorts, cases clinically labeled as low-invasive and RAS-like were predicted as high-invasive by the methylation classifier and were found to harbor cooperating driver alterations (e.g., *TP53* and *KRAS/KEAP1*), consistent with higher invasiveness. These examples illustrate how the methylome captures the nuanced consequence of driver mutations, such as the combined effects of comutations and different KF subclasses. One KF-to-BRAF misclassification in the validation cohort involved a *BRAF* fusion rather than a point mutation (0462: RAB3GAP2-BRAF), which may represent intermediate molecular subtypes (Supplementary Fig. S4D). The few discordant cases in which driver mutation did not match expected clinical invasiveness also showed features suggestive of intermediate states, including low variant allele fractions (e.g., 0870 has a *BRAF* p.V600E variant allele frequency = 0.23).

We identified an immune-infiltrated methylation group (LEUKO cluster) enriched for TFs involved in hematopoiesis and vascular maintenance (Supplementary Fig. S2E and S2F). Because immune infiltration, such as lymphocytic thyroiditis, can substantially alter DNA methylation profiles, samples in the LEUKO cluster showed reduced classification confidence and biased predictions (toward high-invasive tumors), confounding tumor-intrinsic risk stratification. These findings underscore the need for classifiers that distinguish between tumor-intrinsic and microenvironment-driven methylation signals, including approaches such as cell-type deconvolution, expanded sampling of immune-rich, low-invasive tumors, or subgroup-specific modeling. Alternatively, excluding highly leukocyte-infiltrated samples using predefined thresholds may be necessary.

(*Continued*.) for feature interpretation and comprehensive validation. **B** and **C,** Confusion matrices showing prediction performance for clinical invasiveness (left) and driver group (right) from the (**B**) reference cohort and (**C**) validation cohort. **D** and **E,** t-SNE dimensionality reduction of the reference cohort showing misclassifications for (**D**) clinical invasiveness and (**E**) driver group classifiers. Point transparency indicates prediction confidence, and shapes denote accuracy. **F,** t-SNE visualization of 14 paired primary samples (circle) and LN metastases (triangle), color-coded by patient ID with dashed lines connecting paired samples. **G,** TFBS enrichment analysis of 3,000 most important probes from the clinical invasiveness (left) and driver group (right) classifiers, ranked by significance [−log10(FDR)]. Dot size indicates database overlap; color intensity indicates enrichment strength (estimate). EMT, epithelial–mesenchymal transition.

Age-related and cancer-associated DNA methylation changes are often intertwined, and it has been proposed that age-associated epigenetic alterations, particularly at Polycomb-regulated regions, accumulate over time and predispose cells to malignant epigenetic silencing (75). Under this model, adult cancers would be expected to exhibit greater EAA than pediatric tumors, reflecting a longer premalignant mitotic and epigenetic history. In contrast, we observed comparable levels of methylation age acceleration in both pediatric and adult thyroid carcinomas (**Fig. 3E** and **F**), indicating that cancer-associated epigenetic aging is largely independent of chronologic age at diagnosis. This finding suggests that accelerated epigenetic aging primarily reflects tumor clonal expansion and oncogene-driven proliferation rather than the accumulation of preexisting age-related epigenetic changes. Together, these results support a model in which genetically driven tumor initiation rapidly imposes a cancer-specific epigenetic aging signature, even in pediatric disease.

As one of the first efforts to apply methylome-based stratification in pediatric carcinomas, this study is limited by its single-institution cohort. Validation of the classifier from external institutions will be critical to establish the generalizability and robustness of our model. Technical considerations such as batch effects, platform-specific biases, and tissue preservation methods must be addressed, especially when transitioning from laboratory experiments to clinical assays. Limited representation of certain histologic and molecular subtypes in the training cohort, such as *FTC*, *RAS*-like, *DICER1*, and *APC* mutation samples, reduced classification confidence. The association between driver mutation groups and clinical invasiveness also constrains within-group diversity, limiting driver-specific differential methylation analyses. Furthermore, cooperating oncogenic mutations remain undersampled, which complicates the classification of driver groups. Although the classifier demonstrated high accuracy, borderline cases underscore its value as a decision-support tool, complementing clinical, histologic, and radiologic data. Finally, prospective validation evaluating the classifier's impact on surgical planning, patient counseling, and long-term clinical outcomes will be critical to assess real-world clinical utility. Our retrospective study cannot address how this classifier would perform in preoperative decision-making contexts or whether it would meaningfully alter clinical management.

Conclusion

This study demonstrates the feasibility and clinical potential of using DNA methylation profiling to stratify tumor invasiveness and driver mutation status in pediatric thyroid carcinomas. By identifying robust epigenetic signatures associated with both invasive behavior and specific genetic alterations, we showed that methylation-based classifiers can enhance existing diagnostic approaches and support risk-adapted management. Notably, methylation profiling outperformed genotype alone in resolving phenotypic heterogeneity, identifying outlier cases, and predicting metastatic potential, such as in LN samples. These findings lay the groundwork for developing clinically actionable, epigenetic-based risk stratification tools in pediatric oncology.

Data Availability

The generated thyroid methylome profiles are available in the Gene Expression Omnibus (RRID: SCR_005012) with superset accession GSE312914. Informatics for array data preprocessing and functional analysis is available in the R/Bioconductor package *SeSAMe* (version 3.22+): https://bioconductor.org/packages/release/bioc/html/sesame.html. Trained models and training and testing scripts are available at https://github.com/jennyznli/2025_TC. Additional raw data are available upon request to the corresponding author.

Authors' Disclosures

M.M. Li reports personal fees from TD Securities LLC and other support from Oxford Nanopore Technologies outside the submitted work. W. Zhou reports grants from the National Institute of Health during the conduct of the study and grants and personal fees from Illumina outside the submitted work. No disclosures were reported by the other authors.

Authors' Contributions

J.Z. Li: Data curation, formal analysis, validation, investigation, visualization, methodology, writing–original draft, writing–review and editing. **J.C. Ricarte-Filho:** Conceptualization, data curation, validation, investigation, writing–review and editing. **A.R. Isaza:** Data curation, project administration. **K. Hinkle:** Data curation. **F. Xu:** Resources, data curation. **M.M. Li:** Resources, funding acquisition, writing–review and editing. **A.J. Bauer:** Conceptualization, resources, formal analysis, supervision, funding acquisition, investigation, writing–review and editing. **A.T. Franco:** Conceptualization, resources, formal analysis, supervision, funding acquisition, validation, investigation, writing–review and editing. **W. Zhou:** Conceptualization, formal analysis, supervision, funding acquisition, investigation, methodology, writing–original draft, writing–review and editing.

Acknowledgments

The authors thank the Center for Applied Genomics Genotyping Core at the Children's Hospital of Philadelphia for their help with array processing. This work was supported by the Children's Hospital of Philadelphia's Thyroid Center Frontier Program (A.J. Bauer, A.T. Franco, and J.C. Ricarte-Filho); National Institutes of Health R35-GM146978 (W. Zhou); National Institutes of Health R01-CA214511 (A.T. Franco); and Department of Defense: W81XWH2210655 (A.T. Franco).

Note

Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

Received June 7, 2025; revised December 29, 2025; accepted February 13, 2026; posted first February 17, 2026.

References

1. Guleria P, Srinivasan R, Rana C, Agarwal S. Molecular landscape of pediatric thyroid cancer: a review. Diagnostics (Basel) 2022;12:3136.
2. Bernier M-O, Withrow DR, Berrington de Gonzalez A, Lam CJK, Linet MS, Kitahara CM, et al. Trends in pediatric thyroid cancer incidence in the United States, 1998-2013. Cancer 2019;125:2497–505.
3. Hogan AR, Zhuge Y, Perez EA, Koniaris LG, Lew JI, Sola JE. Pediatric thyroid carcinoma: incidence and outcomes in 1753 patients. J Surg Res 2009;156:167–72.
4. Jarzab B, Handkiewicz-Junak D. Differentiated thyroid cancer in children and adults: same or distinct disease? Hormones (Athens) 2007;6:200–9.
5. Alzahrani AS, Alkhafaji D, Tuli M, Al-Hindi H, Sadiq BB. Comparison of differentiated thyroid cancer in children and adolescents (≤20 years) with young adults. Clin Endocrinol (Oxf) 2016;84:571–7.
6. Dinauer CA, Breuer C, Rivkees SA. Differentiated thyroid cancer in children: diagnosis and management. Curr Opin Oncol 2008;20:59–65.
7. Lai S-TT, Bauer AJ. Approach to the pediatric patient with thyroid nodules. J Clin Endocrinol Metab 2025;110:2339–52.
8. Handkiewicz-Junak D, Niedziela M, Lewiński A, Bosowski A, Chmielik E, Czarniecka A, et al. Diagnostics and treatment of differentiated thyroid carcinoma in children - guidelines of the polish national scientific societies, 2024 update. Endokrynol Pol 2024;75:565–91.

9. Lebbink CA, Links TP, Czarniecka A, Dias RP, Elisei R, Izatt L, et al European Thyroid Association Guidelines for the management of pediatric thyroid nodules and differentiated thyroid carcinoma. Eur Thyroid J 2022;11:e220146.
10. Francis GL, Waguespack SG, Bauer AJ, Angelos P, Benvenga S, Cerutti JM, et al. Management guidelines for children with thyroid nodules and differentiated thyroid cancer. Thyroid 2015;25:716–59.
11. Stack BC, Twining C, Rastatter J, Angelos P, Baloch Z, Diercks G, et al. Consensus statement by the American Association Of Clinical Endocrinology (AACE) and the American Head And Neck Society Endocrine Surgery Section (AHNS-ES) on Pediatric Benign And Malignant Thyroid Surgery. Head Neck 2021;43:1027–42.
12. Qian ZJ, Jin MC, Meister KD, Megwalu UC. Pediatric thyroid cancer incidence and mortality trends in the United States, 1973-2013. JAMA Otolaryngol Head Neck Surg 2019;145:617–23.
13. Pfeifer GP. Defining driver DNA methylation changes in human cancer. Int J Mol Sci 2018;19:1166.
14. Zingg JM, Jones PA. Genetic and epigenetic aspects of DNA methylation on genome expression, evolution, mutation and carcinogenesis. Carcinogenesis 1997;18:869–82.
15. Barros-Filho MC, Dos Reis MB, Beltrami CM, de Mello JBH, Marchi FA, Kuasne H, et al. DNA methylation-based method to differentiate malignant from benign thyroid lesions. Thyroid 2019;29:1244–54.
16. Nonaka D. A study of FoxA1 expression in thyroid tumors. Hum Pathol 2017;65:217–24.
17. Cancer Genome Atlas Research Network. Integrated genomic characterization of papillary thyroid carcinoma. Cell 2014;159:676–90.
18. Marczyk VR, Recamonde-Mendoza M, Maia AL, Goemann IM. Classification of thyroid tumors based on DNA methylation patterns. Thyroid 2023;33:1090–9.
19. Bisarro Dos Reis M, Barros-Filho MC, Marchi FA, Beltrami CM, Kuasne H, Pinto CAL, et al. Prognostic classifier based on genome-wide DNA methylation profiling in well-differentiated thyroid tumors. J Clin Endocrinol Metab 2017;102:4089–99.
20. Henke LE, Perkins SM, Pfeifer JD, Ma C, Chen Y, DeWees T, et al. BRAF V600E mutational status in pediatric thyroid cancer. Pediatr Blood Cancer 2014;61:1168–72.
21. de Sousa MSA, Nunes IN, Christiano YP, Sisdelli L, Cerutti JM. Genetic alterations landscape in paediatric thyroid tumours and/or differentiated thyroid cancer: systematic review. Rev Endocr Metab Disord 2024;25:35–51.
22. Li Y, Wang Y, Li L, Qiu X. The clinical significance of BRAFV600E mutations in pediatric papillary thyroid carcinomas. Sci Rep 2022;12:12674.
23. Givens DJ, Buchmann LO, Agarwal AM, Grimmer JF, Hunt JP. BRAF V600E does not predict aggressive features of pediatric papillary thyroid carcinoma. Laryngoscope 2014;124:E389–93.
24. Lupi C, Giannini R, Ugolini C, Proietti A, Berti P, Minuto M, et al. Association of BRAF V600E mutation with poor clinicopathological outcomes in 500 consecutive cases of papillary thyroid carcinoma. J Clin Endocrinol Metab 2007;92:4085–90.
25. Howell GM, Hodak SP, Yip L. RAS mutations in thyroid cancer. Oncologist 2013;18:926–32.
26. Yang AT, Lai S-TT, Laetsch TW, Bhatti T, Baloch Z, Surrey LF, et al. Molecular landscape and therapeutic strategies in pediatric differentiated thyroid carcinoma. Endocr Rev 2025;46:397–417.
27. Franco AT, Ricarte-Filho JC, Isaza A, Jones Z, Jain N, Mostoufi-Moab S, et al. Fusion oncogenes are associated with increased metastatic capacity and persistent disease in pediatric thyroid cancers. J Clin Oncol 2022;40:1081–90.
28. Suchy B, Waldmann V, Klugbauer S, Rabes HM. Absence of RAS and p53 mutations in thyroid carcinomas of children after Chernobyl in contrast to adult thyroid tumours. Br J Cancer 1998;77:952–5.
29. Nikiforov YE, Nikiforova MN, Gnepp DR, Fagin JA. Prevalence of mutations of ras and p53 in benign and malignant thyroid tumors from children exposed to radiation after the Chernobyl nuclear accident. Oncogene 1996;13:687–93.
30. Nikiforov YE. RET/PTC rearrangement in thyroid tumors. Endocr Pathol 2002;13:3–16.
31. Nikiforov YE, Nikiforova MN. Molecular genetics and diagnosis of thyroid cancer. Nat Rev Endocrinol 2011;7:569–80.
32. Nikiforov YE. Thyroid carcinoma: molecular pathways and therapeutic targets. Mod Pathol 2008;21(Suppl 2):S37–43.
33. Kaur D, Lee SM, Goldberg D, Spix NJ, Hinoue T, Li H-T, et al. Comprehensive evaluation of the infinium human methylationepic v2 beadchip. Epigenetics Commun 2023;3:6.
34. Lee SM, Loo CE, Prasasya RD, Bartolomei MS, Kohli RM, Zhou W. Low-input and single-cell methods for Infinium DNA methylation BeadChips. Nucleic Acids Res 2024;52:e38.
35. Zhou W, Triche TJ, Laird PW, Shen H. SeSAMe: reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions. Nucleic Acids Res 2018;46:e123.
36. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res 2016;44:e71.
37. Steiert TA, Parra G, Gut M, Arnold N, Trotta J-R, Tonda R, et al. A critical spotlight on the paradigms of FFPE-DNA sequencing. Nucleic Acids Res 2023;51:7143–62.
38. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics 2010;26:1572–3.
39. Teschendorff AE, Breeze CE, Zheng SC, Beck S. A comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies. BMC Bioinformatics 2017;18:105.
40. Goldberg DC, Fu H, Atkins D, Moyer E, Lee CN, Deng Y, et al. KnowYourCG: facilitating base-level sparse methylome interpretation. Sci Adv 2025;11:eadw3027.
41. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.
42. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol 2013;14:R115.
43. Chen BH, Zhou W. mLiftOver: harmonizing data across Infinium DNA methylation platforms. Bioinformatics 2024;40:btae423.
44. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2020;2:56–67.
45. Pozdeyev N, Gay LM, Sokol ES, Hartmaier R, Deaver KE, Davis S, et al. Genetic analysis of 779 advanced differentiated and anaplastic thyroid cancers. Clin Cancer Res 2018;24:3059–68.
46. Li F, Liang A, Lv Y, Liu G, Jiang A, Liu P. MicroRNA-200c inhibits epithelial-mesenchymal transition by targeting the BMI-1 gene through the phospho-AKT pathway in endometrial carcinoma cells in vitro. Med Sci Monit 2017;23:5139–49.
47. Flavahan WA, Gaskell E, Bernstein BE. Epigenetic plasticity and the hallmarks of cancer. Science 2017;357:eaal2380.
48. Cunningham R, Hansen CG. The Hippo pathway in cancer: YAP/TAZ and TEAD as therapeutic targets in cancer. Clin Sci 2022;136:197–222.
49. Zhang L, Yang S, Chen X, Stauffer S, Yu F, Lele SM, et al. The hippo pathway effector YAP regulates motility, invasion, and castration-resistant growth of prostate cancer cells. Mol Cell Biol 2015;35:1350–62.
50. Liu X, Li H, Rajurkar M, Li Q, Cotton JL, Ou J, et al. Tead and AP1 coordinate transcription and motility. Cell Rep 2016;14:1169–80.
51. Chen L-H, Xie T, Lei Q, Gu Y-R, Sun C-Z. A review of complex hormone regulation in thyroid cancer: novel insights beyond the hypothalamus-pituitary-thyroid axis. Front Endocrinol (Lausanne) 2024;15:1419913.
52. Hoffmann S, Maschuw K, Hassan I, Wunderlich A, Lingelbach S, Ramaswamy A, et al. Functional thyrotropin receptor attenuates malignant phenotype of follicular thyroid cancer cells. Endocrine 2006;30:129–38.
53. Platet N, Cunat S, Chalbos D, Rochefort H, Garcia M. Unliganded and liganded estrogen receptors protect against cancer invasion via different mechanisms. Mol Endocrinol 2000;14:999–1009.
54. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144:646–74.
55. Meng F, Dai L. Transcription factors TP63 facilitates malignant progression of thyroid cancer by upregulating KRT17 expression and inducing epithelial-mesenchymal transition. Growth Factors 2023;41:71–81.
56. Liu J, Lin PC, Zhou BP. Inflammation fuels tumor progress and metastasis. Curr Pharm Des 2015;21:3032–40.
57. Li P, Zhang W, Wu Q, Zhang X, Zheng Z. Retinoid X receptor γ regulates epithelial-mesenchymal transition and tumor immune infiltration in papillary thyroid cancer tumorigenesis: an experimental and in silico study. Endocr Connect 2025;14:e250015.
58. Peng Q, Xie T, Wang Y, Ho VW-S, Teoh JY-C, Chiu PK-F, et al. GLIS1, correlated with immune infiltrates, is a potential prognostic biomarker in prostate cancer. Int J Mol Sci 2023;25:489.

59. Zhou W, Reizel Y. On correlative and causal links of replicative epimutations. Trends Genet 2025;41:60–75.
60. Johnstone SE, Gladyshev VN, Aryee MJ, Bernstein BE. Epigenetic clocks, aging, and cancer. Science 2022;378:1276–7.
61. Romero R, Sayin VI, Davidson SM, Bauer MR, Singh SX, LeBoeuf SE, et al. Keap1 loss promotes Kras-driven lung cancer and results in dependence on glutaminolysis. Nat Med 2017;23:1362–8.
62. Saito J, Kohn AD, Roth RA, Noguchi Y, Tatsumo I, Hirai A, et al. Regulation of FRTL-5 thyroid cell growth by phosphatidylinositol (OH) 3 kinase-dependent Akt-mediated signaling. Thyroid 2001;11:339–51.
63. Bernstein E, Caudy AA, Hammond SM, Hannon GJ. Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 2001;409:363–6.
64. Harvey KF, Tang TT. Targeting the Hippo pathway in cancer. Nat Rev Drug Discov 2025;24:852–69.
65. Eferl R, Wagner EF. AP-1: a double-edged sword in tumorigenesis. Nat Rev Cancer 2003;3:859–68.
66. Kondo T, Nakazawa T, Ma D, Niu D, Mochizuki K, Kawasaki T, et al. Epigenetic silencing of TTF-1/NKX2-1 through DNA hypermethylation and histone H3 modulation in thyroid carcinomas. Lab Invest 2009;89:791–9.
67. Chou C-K, Chi S-Y, Chou F-F, Huang S-C, Wang J-H, Chen C-C, et al. Aberrant expression of androgen receptor associated with high cancer risk and extrathyroidal extension in papillary thyroid carcinoma. Cancers (Basel) 2020;12:1109.
68. Parolia A, Cieslik M, Chu S-C, Xiao L, Ouchi T, Zhang Y, et al. Distinct structural classes of activating FOXA1 alterations in advanced prostate cancer. Nature 2019;571:413–8.
69. Zafon C, Gil J, Pérez-González B, Jordà M. DNA methylation in thyroid cancer. Endocr Relat Cancer 2019;26:R415–39.
70. Ellis RJ, Wang Y, Stevenson HS, Boufraqech M, Patel D, Nilubol N, et al. Genome-wide methylation patterns in papillary thyroid cancer are distinct based on histological subtype and tumor genotype. J Clin Endocrinol Metab 2014;99:E329–37.
71. Minna E, Devecchi A, Pistore F, Paolini B, Mauro G, Penso DA, et al. Genomic and transcriptomic analyses of thyroid cancers identify DICER1 somatic mutations in adult follicular-patterned RAS-like tumors. Front Endocrinol (Lausanne) 2023;14:1267499.
72. Rodrigues L, Da Cruz Paula A, Soares P, Vinagre J. Unraveling the significance of DGCR8 and miRNAs in thyroid carcinoma. Cells 2024;13:561.
73. Kommoss FKF, Chong A-S, Chong A-L, Pfaff E, Jones DTW, Hiemcke-Jiwa LS, et al. Genomic characterization of DICER1-associated neoplasms uncovers molecular classes. Nat Commun 2023;14:1677.
74. Konishi K, Watanabe Y, Shen L, Guo Y, Castoro RJ, Kondo K, et al. DNA methylation profiles of primary colorectal carcinoma and matched liver metastasis. PLoS One 2011;6:e27889.
75. Vaidya H, Jelinek J, Issa J-PJ. DNA methylation, aging, and cancer. Epigenomes 2025;9:18.