Manuscript Type: Original Article

Title: Transcriptomic analysis of papillary thyroid cancer focused on immune-subtyping, oncogenic fusion, and recurrence.

Author's name: Seung-Jin Park\(^{1,2}\), Yea Eun Kang\(^{3}\), Jeong-Hwan Kim\(^{1}\), Jong-Lyul Park\(^{1}\), Seon-Kyu Kim\(^{1}\), Seung-Woo Baek\(^{1,2}\), In Sun Chu\(^{1,2}\), Shinae Yi\(^{3}\), Seong Eun Lee\(^{3}\), Young Joo Park\(^{4}\), Eun-Jae Jung\(^{5}\), Jin Man Kim\(^{6}\), Hye Mi Ko\(^{7}\), Je-Ryong Kim\(^{7}\), Seung-Nam Jung\(^{8}\), Hong Ryun Won\(^{8}\), Jae Won Chang\(^{8}\), Bon Seok Koo\(^{8}\), and Seon-Young Kim\(^{1,2,9}\)

Affiliation:

\(^{1}\)Korea Research Institute of Bioscience and Biotechnology, Daejeon 34131, Republic of Korea
\(^{2}\)Department of Bioscience, University of Science and Technology (UST), Daejeon 34113, Republic of Korea
\(^{3}\)Department of Internal Medicine, Chungnam National University College of Medicine, Daejeon, South Korea
\(^{4}\)Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Korea
\(^{5}\)Department of Otorhinolaryngology-Head and Neck Surgery, Seoul National University College of Medicine, Seoul, Korea
\(^{6}\)Department of Pathology, Chungnam National University College of Medicine, Daejeon, South Korea
\(^{7}\)Department of Surgery, Chungnam National University College of Medicine, Daejeon, Republic of Korea
\(^{8}\)Department of Otolaryngology-Head and Neck Surgery, Chungnam National University College of Medicine, Daejeon, South Korea
\(^{9}\)Korea Bioinformation Center (KOBIC), Korea Research Institute of Bioscience and Biotechnology, Daejeon 34131, Republic of Korea

\(\dagger\)These authors contributed equally to this work.
Correspondence to:

Seon-Young Kim, PhD
Principal Investigator, Korea Research Institute of Bioscience and Biotechnology
125, Gwahak-ro, Yuseong-gu, Daejeon, Republic of Korea
Tel: 82-42-879-8500
E-mail: kimsy@kribb.re.kr

Bon Seok Koo MD, PhD.
Professor, Department of Otolaryngology-Head and Neck Surgery,
Chungnam National University College of Medicine,
640 Daesa-Dong, Chung-Gu, Daejeon, 301-721, Republic of Korea
Tel: +82-42-280-7690 Fax: +82-42-253-4059
E-mail: bskoo515@cnuh.co.kr
ABSTRACT

Objectives.

Thyroid cancer is the most common endocrine tumor, with rapidly increasing incidence worldwide. However, its transcriptomic characteristics associated with immunological signatures, driver fusion, and recurrence markers are still unclear. We aimed to investigate the transcriptomic characteristics of advanced papillary thyroid cancer patients.

Methods

This study included 282 papillary thyroid cancer tumor samples, and 155 normal samples from Chungnam National University Hospital and Seoul National University Hospital. Transcriptomic quantification was determined by high-throughput RNA sequencing. We investigated the association of clinical parameters and molecular signatures using RNA sequencing. We validated the predictive biomarker using The Cancer Genome Atlas (TCGA) database.

Results

By comparing differentially expressed genes, gene sets, and pathways in papillary thyroid cancer than normal adjacent of tumor tissue, we found increased immune signaling associated with cytokine or T-cell and decreased thyroid hormone synthetic pathways. In addition, patients with recurrence represented increased CD8+ T-cell signature and Th1 cells signature, respectively. Interestingly, we found differentially overexpressed genes related to immune-escape signaling such as CTLA4, IDO1, LAG3, and PDCD1 in advanced PTC with low thyroid differentiation score (TDS). Fusion analysis showed that the PI3K and MAPK signaling pathways were regulated differently according to RET fusion partner genes (CCDC6 and NCOA4). Finally, we identified HOXD9 as a novel molecular biomarker that predicts the recurrence of thyroid cancer in addition to known risk factors (tumor size, lymph node...
Conclusion

We identified a high association with immune-escape signaling in the immune-hot group with aggressive clinical factors of Korean thyroid cancer patients. Moreover, RET fusion differentially regulated PI3K and MAPK signaling depending on the partner gene of RET, and HOXD9 was found to be a recurrence marker for advanced PTC patients.

Keywords: thyroid cancer, Korean thyroid cancer, Advanced papillary thyroid cancer, RNA sequencing, immune subtyping, immune-escape signaling, fusion outlier, predictive biomarker, RET, HOXD9
INTRODUCTION

The incidence of thyroid cancer is increasing worldwide (1, 2). Recently, the distinction between benign and malignant thyroid nodules is a very challenging problem, thus, methods of various clinical or experimental elements have been studied for increasing the accuracy (3).

Approximately 90% of thyroid cancers are differentiated thyroid carcinomas (DTC), including the papillary and follicular types. Although most DTCs have good prognosis, they often recur and in some cases become metastatic and aggressive, which makes treatment difficult. In addition, the dedifferentiation of DTC is known to be one of the mechanisms related to the development of anaplastic thyroid carcinoma (ATC), which has rapid disease progression and extremely poor survival rates (4). However, the major mechanism related to thyroid dedifferentiation in advanced PTC is still unknown.

Decades after RNA sequencing has been developed, attempts to use RNA sequencing continue to evolve clinically. In the past, RNA sequencing was used only to investigate the gene expression, but now it is being used to discover analysis of oncogenic fusions, tumor microenvironment from calculating the relative abundance of immune cells and oncogenic splice patterns. (5-7). While large-scale comprehensive multi-omics studies of thyroid cancer have produced the molecular landscape of thyroid cancer and surgical factors related to the recurrence of thyroid cancer were identified through meta-analysis (8), the molecular markers to predict the recurrence in advanced PTC were not fully understood (9-11).

RET rearrangement is a widespread gene fusion event observed in PTC and crucial in thyroid tumorigenesis (12), and sepcatinib, an FDA-approved drug for RET fusion-positive thyroid cancer, was developed (13). RET fusion in thyroid cancer occurs more frequently in children than adults, and it is known that RET fusion causes transformation in thyroid cells to cause hyperplasia or neoplasia (14). However, since approximately 30% of patients present with
treatment-ineffective cases, biomarkers to distinguish different subtypes of RET fusion patients are necessary. Especially study on the molecular mechanism of RET fusion in Korean thyroid cancer patients is still few. Additionally, common driver mutations (e.g., BRAF and KRAS) either through targeted sequencing or through whole exome/genome sequencing are insufficient to predict recurrence in Korean patients with racial characteristics with high iodine intake (15-17). Thus, it is necessary to study a molecular biomarker that predicts recurrence in Korean-specific advanced PTC patients.

Our study aimed to identify novel therapeutic targets and prognostic biomarkers through transcriptomic analyses of Korean advanced papillary thyroid cancer.
MATERIALS AND METHODS

Ethics statement
This study was approved by the Institutional Research and Ethics Committee at Chungnam National University Hospital (CNUH- 2020-11-004-001), and Institutional Research and Ethics Committee at Seoul National University Hospital (H-1508-147-700). 40 PTC patients were enrolled from Seoul National University Hospital, and the other patients were enrolled at Chungnam National University Hospital.

Patients
Fresh frozen tissues from 282 PTC patients who underwent thyroidectomy were analyzed using a massively parallel sequencing method. Normal thyroid tissues (n=155) were obtained from patients who underwent thyroidectomy, with confirmed cases of differentiated thyroid cancer. The clinicopathological characteristics of patients according to histology are shown in Supplementary Table 1.

RNA preparation, library construction, and sequencing
In addition, a 2100 Agilent Bioanalyzer (Agilent Technologies, Waldbroon, Germany) was used to estimate the RNA integrity number (RIN) score. After stranded total RNA-seq library construction, we used the TruSeq Stranded Total RNA Sample Preparation Kit (Illumina) based on the manufacturer's instructions. We sequenced the samples via the HiSeq 4000 sequencing system, yielding 2x100-bp sequencing reads. The raw data were deposited in the Korean Nucleotide Archive (KoNA, https://kobic.re.kr/kona) with the accession ID PRJKA210106.

RNA sequencing data analysis
Sequencing reads were trimmed by trim galore (available at https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with a threshold of average sequence quality over 30 and mapped to the human reference genome GRCh38 by STAR aligner (STAR-2.7.9a) (18). Ensemble gencode (ver. 27) gene annotation was used to profile gene expression from raw read counts, normalization, and quantification by cufflinks at the gene level (19).

**ssGSEA and GSEA**

We performed single-sample gene set enrichment analysis (ssGSEA) for functional characterization by samples (20). For the reference gene matrix transposed (GMT) file, we downloaded gmt files in MSigDB (21). Then, the gsva function in the GSVA package was used with the default option. We used enrichR for functional classification called gene-set enrichment analysis (GSEA) by differentially expressed genes with a threshold p-value under 0.01 from Fisher’s exact test (22).

**Fusion detection and outlier analysis**

Sequenced reads were mapped to the human reference genome GRCh38 for fusion detection using STAR fusion with default options (5). The result of fusion analysis was filtered out through the following three processes: 1) genes were eliminated, such as immunoglobulin genes, uncharacterized genes, and RNA genes; 2) fusions including the same gene or paralog gene were removed; and 3) fusions detected in normal samples were omitted. We performed outlier analysis for fusion candidates. The outlier means that the sample's level at a given gene was greater than the 75th percentile + 1.5 interquartile range (IQR). This trend should not be discovered in the normal group.
Thyroid differentiation score (TDS) analysis.

We used the TDS score from the TCGA-THCA study (9). TDS scores is from ssGSEA performed with 16 genes (DIO1, DIO2, DUOX1, DUOX2, FOXE1, GLIS3, NKX2-1, PAX8, SLC26A4, SLA5A5, SLC5A8, TG, THRA, THRB, TPO, TSHR) associated with thyroid function and metabolism from the TCGA study. As a result of calculating the correlation of the TDS scores, correlation values were highly significant (R=0.859, p < 2.2e-15).

Immune-subtyping and module analysis.

We carried out immune-subtyping from xCell (6). The abundances of 64 various cell types for papillary thyroid cancer samples were computed using FPKM. We divided the patients into three major clusters using unsupervised K-means clustering using pheatmap package from R 4.1.2. These groups have a significant difference of ImmuneScore from xCell. Thus, we named them immune-hot, -intermediate, and -cold. ImmuneScore is a sum of 10 immune cell types such as B-cells, CD4+ T-cells, CD8+ T-cells, dendritic cells, eosinophils, macrophages, monocytes, mast cells, neutrophils, and NK cells. It is an experimental score and known to better optimize tumor microenvironment. In three subtypes, we compared various immune abundances and RNA expression (Fisher’s exact test; p value < 0.01). To discover sets of coexpressed and hub genes in the network, we utilized CEMiTool (23). This tool integrates the identification and analysis of coexpression gene modules, evaluating which modules contain genes that are overrepresented by specific gene-set terms with p values under 0.01 from Fisher’s exact test.
RESULTS

Overview of molecular profiling in thyroid cancer cohort

We collected 282 papillary thyroid cancer (PTC) samples (Supplementary Table 1). Clinicopathological characteristics included capsular invasion (CI), extrathyroidal extension (ETE), central and lateral lymph node metastasis (C_ and L_LNM), lymphovascular invasion (LVI), recurrence, age, and sex. The average age of the PTC patients was 49 years. The mean tumor size of PTC samples was 1.83 cm.

We quantified fragment per kilobase of transcript per million (FPKM) values for 15,690 genes through a standard RNA-sequencing analysis pipeline. First, we performed identification of differentially expressed genes of tumors than normal from PTC at fold-change > 1.5 and p-value < 0.01, and we selected a total of 2,307 genes as DEGs (Figure 1A). There were 1,039 down-regulated genes and 1,268 up-regulated genes, respectively. Top five highly up-regulated genes in PTC were ZCCHC12, FN1, CHI3L1, SLC34A2, and DCSTAMP.

We performed gene set analysis with DEGs using Pathways-KEGG 2021 Human and MSigDB hallmark genesets in enrichR (Figure 1B). Up-regulated gene set terms in PTC vs. normal samples included cytokine receptor interaction, p53 signaling, and complement and coagulation in KEGG and EMT, inflammatory response, and allograft rejection in Hallmark. Down-regulated gene set terms included thyroid hormone synthesis and amino acid metabolism in KEGG and estrogen response, UV response, and hypoxia in Hallmark. Interestingly, many immune-associated gene-sets were significantly higher such as EMT including ECM1, SPARC, LOXL1 and inflammatory response including ICAM4, MMP14, TLR2.

We also found that immune-related signaling was up-regulated in PTC. When patients with and without LNM were compared, immune-related signaling, such as NF-κB, TNF-α, epithelial
cell score, WNT signaling, CD4+ T-cells, memory B-cells, etc. was increased exclusively. In addition, when patients with and without recurrence were compared, CD8+ T-cells score and Th1 cells score were increased specifically (Figure 1C).

We delineated pathway maps that show the expression of all genes belonging to the terms of KEGG (Supplementary Figure 1). In PTC tumors, we showed the cytokine receptor interaction pathway in which many cytokines including CCR1-2, CCR4-5, and CCR7 in CC-su family and TGFB1, BMPR2, and ACVR1 in TGF-β family are significantly up-regulated.

**Immunological characteristics of the thyroid cancer cohort**

Attempts to apply immunotherapy to thyroid cancer are continuing, and in addition to the discovery of many immune checkpoint inhibitors (ICI) such as PD-L1, CTLA4, IDO1, and HAVCR2, various immune signaling such as TNF-α signaling and IL-6/JAT/STAT3 pathway have been discovered (24, 25). To investigate the role of immune-escape signaling in advanced PTC, we investigated the expression of genes related to immunotherapy and immune subtyping using ImmuneScore. We performed an immune signature analysis using xCell from RNA sequencing data in PTC (Methods). We divided PTC into three groups using K-means clustering with 64 immune-related scores. Based on the immune scores, we classified patients into immune-hot (n=35), immune-intermediate (n=107) and immune-cold (n=140) groups. We described significantly different immune-related scores and expression of the top 50 genes across the three immune groups (Figure 2A). In the Immune-hot group, B-cells, CD4+ or CD8+ T-cells, dendritic cells, ImmuneScore, etc., are increased. In the Immune-intermediate group, Astrocytes, Fibroblasts, Hepatocytes, Preadipocytes, etc., are enriched. In the Immune-cold group, Thyroid differentiation score (TDS) score, common lymphoid progenitor, myocytes,
osteoblasts are elevated (Figure 2A top panel). The association between clinical data and immune-subtype showed significant differences in ETE status, C_LNM, and L_LNM (Figure 2A middle panel and Supplementary Table 2). Interestingly, the immune-hot group revealed the characteristics of poor prognostic indicators (tumor size, ETE, c_LNM, and L_LNM). We selected significantly different genes from each group and performed gene-set enrichment analysis. In the Immune-hot group, genes belonging to interferon-γ, such as ISG20, CASP8, and PTPN6, represent the significant difference (p<0.001). In the Immune-intermediate group, genes belonging to EMT such as LAMA2, MMP2, VEGFA, and BMP1 (p<0.001) and in the Immune-cold group, genes belonging to the Myc target, such as XRCC6, XPOT, NPM1, and CDK4, were significantly (p<0.001) enriched at each group (Figure 2A bottom panel). Based on the distribution of TDS scores across three immune groups, it was found that the TDS score of the immune-cold group was significantly higher and the TDS score of the immune-hot group was significantly lower than other groups (Figure 2B). In addition, we extracted the distribution of TDS scores across immune-subtypes and genes with a high correlation with TDS scores (Figure 2C). As the TDS score decreased, the distribution of immune-hot increased and the distribution of immune-cold decreased. Conversely, as the TDS score increased, the distribution of immune-hot decreased, and the distribution of immune-cold increased. Moreover, genes such as B3GNT9 (R=0.80), TSHR (R=0.77), MMP15 (R=0.77), LIPG (R=0.76), and CD24 (R=0.76) showed very a high correlation with the TDS score.

In the immunogenic hot group with a low TDS score, the expression of genes related to immune-escape signaling such as CTLA4, IDO1, LAG3, and PDCD1 was significantly higher than that of the other groups (Figure 2D). Finally, we performed module and co-expression analysis across three immune groups (Figure 2E). In the immune-hot group, genes such as ITGA4, UBC, NFKB1, and PRKCA belonging to the interferon-γ module were acting as hub
genes with many interactions, in the immune-intermediate group, MMP2, in the cold group, XPOT, TCEA1, IARS, etc. of genes were presented as hub genes. These data suggested that immune subtyping, including immune-hot, -intermediate, and -cold groups significantly associated with TDS and prognostic clinic parameters associated with the immunogenic-hot signature.

**Fusion effects of RET in the papillary thyroid cancer cohort**

We performed fusion analysis from RNA-sequencing data with STAR fusion (Methods), identified 3,160 fusions, filtered out low-significance fusions, and finally obtained 1,129 fusions. All significant fusions were found in PTC samples. Each fusion count is shown in order of frequency (Figure 3A). FBXO25 fusion was found in 22 samples, followed by RET, CA13, NTRK1, and MET fusion. We performed fusion outlier analysis with three fusion genes (FBXO25, RET, and CA13) to determine driver fusions (Methods). No significant upregulation was observed for CA13 and FBXO25 fusion samples (Supplementary Figure 2A). The expression of RET in all fusion-positive tumor samples was higher than that in fusion-negative samples, and CCDC6, NCOA4, and DLG5 were partner genes of RET fusion (Figure 3B). We investigated how the mitogen-activated protein kinase (MAPK) and phosphatidylinositol 3-kinase (PI3K) pathways, known as downstream signaling of RET fusion, are differentially regulated (26-28) by different fusion partners of RET. DLG5 was excluded from further analysis, as it was found in only one sample. We divided samples into RET-CCDC6, RET-NCOA4, and RET-WT and selected DEGs (Fisher’s exact test; p < 0.05). In the RET-CCDC6 and RET-NCOA4 fusions, 213 and 1260 genes were significant, respectively, and among them, genes belonging to the MAPK and PI3K signaling pathways were selected and described (Figure 3C). In the RET-CCDC6 fusion group, CREB1 and PIK3R1 were up-regulated, and
PPP2R2D and MAP2K7 were downregulated. In the RET-NCOA4 group, more genes belonging to the PI3K and MAPK signaling pathways were differentially expressed than in the RET-CCDC6 fusion group. When the association with clinical factors according to two partner genes was analyzed, ETE and L_LNM were significantly frequent in RET-NCOA4 patients, unlike RET-CCDC6 group (Supplementary Figure 2B). Then, we selected marker genes that can distinguish RET-CCDC6 and RET-NCOA4 and discovered FGF13 in RET-CCDC6 and AKT1 and CCND2 in RET-NCOA4 (Supplementary Figure 3A-B and Figure 3D). The three genes were validated in the TCGA-THCA cohort (Figure 3E). Thus, we found that RET fusion differentially regulated the PI3K and MAPK signaling pathways according to the partner genes of RET (29) (Figure 3F).

**Discovery of HOXD9 as a molecular biomarker for the recurrence of papillary thyroid cancer**

One of the unmet medical needs for treating thyroid cancer is the identification of biomarkers that predict recurrence or distant metastasis (30). Clinical factors such as age, sex, tumor size, ECI, ETE, C_ and L_LNM, and L-LVI are currently used to predict recurrence or distant metastasis (31, 32). By Fisher’s exact test for each clinical factor, we found that recurrence occurred more frequently in the group with large tumor size and the group with ETE and L_LNM (Figure 4A). For the three significant risk factors, we divided samples into two groups: recurrence and non-recurrence groups. Most recurrence samples had all three risk factors (n=24, 59%), and most non-recurrence samples did not have any of the three risk factors (n=88, 36%) (Figure 4B). However, interestingly, we found 3 no-risk-factor samples in the recurrence group and 42 all-risk-factor samples in the non-recurrence group. We named these two groups No_RF and All_RF to distinguish them from other recurrences. We then performed statistical tests to
identify genes that can predict recurrence in the No_RF group and identified the homeobox d9 (HOXD9) gene, which is significant in disease-free survival in The Cancer Genome Atlas Database (TCGA) using GEPIA (http://gepia.cancer-pku.cn/) (Figure 4C-D).

To understand the relationship between HOXD9 expression and recurrence, we divided the patients into HOXD9-high and HOXD9-low groups by the median value of HOXD9 expression, compared the ratio of recurrence, and confirmed that the HOXD9-high group had significantly more frequent recurrence than the HOXD9-low group (p=0.0009, Figure 4D).

Also, in the HOXD9 high group, the tumor size was larger and ETE was more frequent than HOXD9 low group (Supplementary Figure 4A-B). We found that HOXD9 is significantly related to recurrence as well as factors known to cause recurrence. Additionally, gene-set enrichment analysis revealed that the NF-kappa B signaling pathway was the most significant gene set among the selected genes at the threshold of p < 0.05 from Fisher’s exact test and fold change > 1.5 (Supplementary Figure 4C and Figure 4E).

We then analyzed transcription factors (TFs) regulating the expression of HOXD9 using a TF database (TFDB) (Methods) and validated the genes in the TCGA-THCA cohort (p < 0.0001) (Supplementary Figure 4D). Transcription Factor 3 (TCF3) and enhancer of zeste 2 polycomb repressive complex 2 subunit (EZH2) are known to bind to the promoter region of HOXD9, and the correlation with HOXD9 was also significant (Figure 4G). Thus, we suggest that TCF3 and EZH2 may promote overexpression of HOXD9 associated with recurrence. Therefore, we suggest HOXD9 as a novel prognostic marker of recurrent papillary thyroid cancer that acts through the NF-kappa B signaling pathway (Figure 4H).
DISCUSSION

In this study, we report the transcriptomic characterization of 282 Korean papillary thyroid cancer samples. Although RNA sequencing was performed on samples from two centers, we discovered a few novel transcriptomic characteristics. In addition, we made efforts to increase the accuracy by validating TCGA-THCA.

Our data revealed that ZCHHC12, an oncogene in thyroid cancer, is overexpressed explicitly in PTC tumors. Previous studies reported that zinc finger CCHC-type containing 12 (ZCCHC12) is a novel oncogene in papillary thyroid cancer and has been particularly associated with lymph node metastasis (33). We also showed how genes belonging to a particular gene set are expressed by analyzing genes, gene sets, and pathways. We found that THRA and PAX8 were repressed as previously reported from PTC patients (11). Efforts to apply immunotherapy to thyroid cancer are continuing (25). We subtyped 282 samples into immune-hot, immune-intermediate, and immune-cold groups using xCell. Based on this, we found immune subtype-specific genes in each subtype, such as PTPN22, PODN, and CASC4.

In particular, PTPN22, a physiological regulator of the T cell receptor known as a target for immunotherapy, was up-regulated in many human cancers, and we discovered an association between PTPN22 and immune score in the immune-hot group (35, 36). PODN, which encodes a small proteoglycan family protein that is a type of ECM protein, was up-regulated in gastric tumors, associated with cell proliferation and a gene member of 6 genes to predict the prognosis of gastric cancer (37). CASC4, an associated splicing event, was up-regulated in human breast cancer or ovarian cancer. Its overexpression is associated with poor prognosis in breast cancer patients with accumulating aberrantly spliced forms (38).

The overexpression of the well-known immune inhibitor genes in the immune-hot subtype was also confirmed in our dataset and the TCGA-THCA dataset. The overexpression of both
cytotoxic T lymphocyte antigen-4 (CTLA4) and programmed cell death protein 1 (PD1) suggests that combination therapy is expected to be effective for the treatment of aggressive immune-hot thyroid cancer (39). In addition, lymphocyte activating 3 (LAG3) is expressed on exhausted T cells and is known as an immune checkpoint inhibitor (40). Also, since IDO1 was associated with immune escape and inflammatory neovascularization from previous report, we speculated IDO1 might be a key player of the immune-escape mechanism in the thyroid cancer cell (41). It is also generally known that thyroid cancer is not immunogenic, but thyroid cancer has immunogenic properties during dedifferentiation (24). Based on this, we found IDO1 may be one of those genes inducing immune escapes for a better understanding of thyroid cancer and immunogenicity. In addition, therapeutic targets in immune-intermediate and immune-cold groups were presented through module analysis. XPOT, IARS, TRIB3, and MMP2 were candidate molecular targets. Isoleucyl-tRNA synthetase (IARS), whose function in cancer remains unclear, and exportin for tRNA (XPOT), a member of the RAN GTPase exportin family for exporting tRNA and associated with tumor development, were discovered (42). The elevated expression of MMP2 is involved in angiogenesis and development in cancer cells (43).

We also found RET-CCDC6 and RET-NCOA4 to be driver fusions in thyroid cancer. In particular, we revealed that the two fusions regulate the MAPK and PI3K signaling pathways differently depending on the partner gene. We discovered FGF13 in RET-CCDC6 and AKT1 and CCND2 in RET-NCOA4. Fibroblast growth factor 13 (FGF13), AKT serine/threonine kinase 1 (AKT1), and cyclin d2 (CCND2) are known to promote the survival of cancer cells, are involved in angiogenesis as a cancer hallmark gene, or are associated with tumorigenesis (44-46). We first found that the various RET fusions have different oncogenic impacts depending on the partner genes in thyroid cancer. This finding will be important in developing therapeutic targets for RET fusions.
Finally, HOXD9 was identified as a novel prognostic biomarker for recurrence in PTC samples. HOXD9, a well-known oncogene, is up-regulated in various cancers, including gastric cancer, hepatocellular carcinoma, and thyroid cancer (47, 48). Our data also revealed that the NF-kappa B signaling pathway was the most significant gene set in HOXD9-high enriched signaling pathways. NF-kappa B signaling is known to cause oncogenesis and disease recurrence in malignant cells and is also known to cause therapy resistance (49). NF-kappa B signaling induces the MAPK signaling pathway downstream of RET fusion (50). In addition, we showed that HOXD9 regulation by TCF3 and EZH2 impacted immune signaling pathways. Overexpression of HOXD9 shortened disease-free survival in TCGA-THCA data.

There are inherent limitations to this study. The experiment was performed only with RNA sequencing, and samples from various races and centers could not be collected. However, we tried to overcome this limitation by analyzing TCGA-THCA. Further functional studies are necessary to validate our findings.

In conclusion, our study showed a high association with immune-escape signaling in the immune-hot group with aggressive clinical factors of thyroid cancer patients. Moreover, RET fusion differentially regulated PI3K and MAPK signaling depending on the partner gene of RET, and HOXD9 was found to be a recurrence marker for advanced PTC patients.
This work was supported by Systemic Industrial Infrastructure Projects through the Ministry of Trade, Industry, and Energy (MOTIE) [P0009796, 2019 to S.-Y.K and Y.E.K.].

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.
FIGURE LEGENDS

Figure 1. Transcriptomic overview and immuno-clinical association from Korean papillary thyroid cancer patients. (A) Volcano plots showing DEGs across groups. Red and gray dots represent significance (Fisher’s exact test; \( p < 0.01 \) and over 1.5-fold) and non-significance, respectively. (B) Results of gene-set enrichment analysis using KEGG and Hallmark database from enrichR. Light red and light blue indicate up-regulated and downregulated terms in tumors compared to normal tissues. (C) Increased immune-related signaling when LNM and recurrence occur exclusively. NF-κB and TNF-α are only enriched when LNM occurs and CD8+ T cells and Th1 cells are only enriched when recurrence occurs.

Figure 2. Association of immune signatures with thyroid differentiation score in advanced PTC. (A) Immune landscape of the PTC using the immune signature of xCell. The immune subtype based on the immune score was represented as immune-hot (n=35), immune-intermediate (n=107), and immune-cold (n=140). (B) Thyroid differentiation score (TDS) across three immune subtypes. (C) Heatmap sorted by TDS and expression of top 20 genes with high correlation with TDS score. (D) Boxplots showing the major immune-checkpoint inhibitors, including CTLA4, IDO1, LAG3, and PDCD1. (E) Hub gene discovery through network module analysis at each immune-subtypes.

Figure 3. Different ways to regulate PI3K and MAPK signaling pathways according to partner genes of RET fusion. (A) Barplot showing fusion count across PTC tumors. (B) Outlier analysis of RET fusion. Blue and red dots indicate the expression of the RET gene in samples without and with fusions, respectively. (C) Impacts of two partners of RET fusion on the MAPK and PI3K pathways. “Common” indicates overlapping genes in both pathways. Circles and squares
indicate cis- and trans-acting genes, and their sizes indicate the significance of the p value. (D-E) Genes regulated by RET fusion of CCDC6 and NCOA4 may bridge the MAPK signaling pathway. Figure 3D represents PTC in the KTC cohort, and Figure 3E illustrates PTC in the TCGA cohort. (F) Two ways that RET fusion regulates MAPK signaling in different partners.

Figure 4. HOXD9 is a candidate gene associated with the recurrence of papillary thyroid cancer. (A) Discovery of recurrence-related factors in PTC samples. Fisher’s exact tests were performed with a threshold under p value 0.01. (B) Barplots showing the spectrum of samples in the PTC group on well-known risk factors such as tumor size, extrathyroidal extension (ETE), and lateral lymph node metastasis (L_LNM). ECI; extracapsular invasion; C_LNM; central lymph node metastasis; L_LVI; lymphovascular invasion. (C) Workflow for selecting HOXD9. (D) Kaplan–Meier plot of thyroid cancer patients based on the expression of HOXD9 (high and low based on the median value from GEPIA database. (E) Barplots showing the recurrence rate based on the expression of HOXD9. (F) Enriched signaling pathways among HOXD9 high group. (G) Correlation of gene expression between HOXD9 and transcription factors such as TCF3 and EZH2 that regulate HOXD9 in the KTC and TCGA-THCA cohorts. (H) Summary of two routes to recurrence from Korean PTC samples.

Supplementary Figure 1. The most altered pathway of PTC.

Supplementary Figure 2. (A) Expression patterns of non-driver fusion. Boxplots showing non-driver oncogenic fusions, such as CA13 and FBXO25. (B) The ratio of ETE and L_LNM according to the partner gene of RET fusion compared to WT.
Supplementary Figure 3. Differently regulated genes PI3K and MAPK signaling according to partner genes in Fusion analysis. We found RET-CCDC6-specific expression of FGF13 (A) and RET-NCOA4-specific expression of AKT1 and CCND2 (B).

Supplementary Figure 4. Characteristics of HOXD9 high group. (A-B) Association between HOXD9 and clinical factors, including tumor size and ETE. (C) NF-kB signaling from gene set enrichment analysis in HOXD9 high group. (D) TFs such as TCF3 and EZH2 are known to regulate the expression of HOXD9 as binding promoter regions of HOXD9.

Supplementary Table 1. Clinical characteristics of Korean papillary thyroid cancer. Clinical characteristics table from the Korean papillary thyroid cancer cohort.

Supplementary Table 2. Association between clinical characteristics and immune subtypes. All statistical tests were performed from Fisher’s exact test.
References


15. Kim HJ, Park HK, Byun DW, Suh K, Yoo MH, Min YK, et al. Iodine intake as a risk factor for...


32. Randolph GW, Duh QY, Heller KS, LiVolsi VA, Mandel SJ, Steward DL, et al. The prognostic significance of nodal metastases from papillary thyroid carcinoma can be stratified based on the size and number of metastatic lymph nodes, as well as the presence of extranodal extension. Thyroid. 2012 Nov;22(11):1144-52.


Supplementary Table 1.

<table>
<thead>
<tr>
<th>Korean papillary thyroid cancer cohort Variables</th>
<th>PTC tumor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>282</td>
</tr>
<tr>
<td>Age (years) (mean ± SD)</td>
<td>49 (± 15.5)</td>
</tr>
<tr>
<td>Female sex (n, %)</td>
<td>206 (73.0)</td>
</tr>
<tr>
<td>Tumor size, cm (mean ± SD)</td>
<td>1.83 (± 1.79)</td>
</tr>
<tr>
<td>Extracapsular invasion (n, %)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>40 (14.2)</td>
</tr>
<tr>
<td>No</td>
<td>242 (85.8)</td>
</tr>
<tr>
<td>Extrathyroidal extension (n, %)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>124 (44.0)</td>
</tr>
<tr>
<td>No</td>
<td>158 (56.0)</td>
</tr>
<tr>
<td>Central lymphnode metastasis (n, %)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>151 (53.5)</td>
</tr>
<tr>
<td>No</td>
<td>131 (46.5)</td>
</tr>
<tr>
<td>Lateral lymphnode metastasis (n, %)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>118 (41.8)</td>
</tr>
<tr>
<td>No</td>
<td>164 (58.2)</td>
</tr>
<tr>
<td>Lymphovascular invasion (n, %)</td>
<td></td>
</tr>
<tr>
<td>Recurrence or persistent disease</td>
<td>Yes</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Recurrence or persistent disease</th>
<th>Yes</th>
<th>40 (14.2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
<td>242 (85.8)</td>
</tr>
</tbody>
</table>
Supplementary Table 2.

<table>
<thead>
<tr>
<th>Immune-subtypes variables</th>
<th>Number of patients</th>
<th>Immune-subgroups</th>
<th></th>
<th></th>
<th></th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Hot (n=35)</td>
<td>Intermediate (n=107)</td>
<td>Cold (n=140)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumor size (cm) (mean ± SD)</td>
<td></td>
<td>2.42 ± 1.8</td>
<td>2.01 ± 2.1</td>
<td>1.55 ± 1.4</td>
<td>0.02*</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td>F 206</td>
<td>22(62.9)</td>
<td>73(68.2)</td>
<td>111(79.3)</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M 76</td>
<td>13(37.1)</td>
<td>34(31.8)</td>
<td>29(20.7)</td>
<td></td>
</tr>
<tr>
<td>ECI</td>
<td></td>
<td>Yes 147</td>
<td>16(45.7)</td>
<td>49(45.8)</td>
<td>82(58.6)</td>
<td>0.099</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No 135</td>
<td>19(54.3)</td>
<td>58(54.2)</td>
<td>58(41.4)</td>
<td></td>
</tr>
<tr>
<td>ETE</td>
<td></td>
<td>Yes 124</td>
<td>22(62.9)</td>
<td>61(57.0)</td>
<td>41(29.3)</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No 158</td>
<td>13(37.1)</td>
<td>46(43.0)</td>
<td>99(70.7)</td>
<td></td>
</tr>
<tr>
<td>C_LNM</td>
<td></td>
<td>Yes 151</td>
<td>26(74.3)</td>
<td>75(70.1)</td>
<td>50(35.7)</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No 131</td>
<td>9(25.7)</td>
<td>32(29.9)</td>
<td>90(64.3)</td>
<td></td>
</tr>
<tr>
<td>L_LNM</td>
<td></td>
<td>Yes 118</td>
<td>24(68.6)</td>
<td>52(48.6)</td>
<td>42(30.0)</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No 164</td>
<td>11(31.4)</td>
<td>55(51.4)</td>
<td>98(70.0)</td>
<td></td>
</tr>
<tr>
<td>LVI</td>
<td></td>
<td>Yes 231</td>
<td>29(82.9)</td>
<td>91(85.0)</td>
<td>111(79.3)</td>
<td>0.501</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No 51</td>
<td>6(17.1)</td>
<td>16(15.0)</td>
<td>29(29.7)</td>
<td></td>
</tr>
<tr>
<td>Recurrence or persistent disease</td>
<td>Yes</td>
<td>40</td>
<td>8(22.9)</td>
<td>15(14.0)</td>
<td>17(12.1)</td>
<td></td>
</tr>
<tr>
<td>---------------------------------</td>
<td>------</td>
<td>------</td>
<td>---------</td>
<td>----------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>242</td>
<td>27(77.1)</td>
<td>92(86.0)</td>
<td>123(87.9)</td>
<td>0.267</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. Transcriptomic overview and immuno-clinical association from Korean papillary thyroid cancer patients

A. Transcriptomic overview

B. KEGG
- Cytokine receptor interaction
- p53 signaling
- Complement and coagulation
- Pathways in cancer
- Viral protein interaction
- Thyroid hormone synthesis
- S, S and T metabolism
- Mineral absorption
- A and P metabolism
- Cortisol synthesis and secretion

C. Hallmark
- EMT
- Inflammatory response
- Allograft rejection
- TNF-α signaling via NF-κB
- KRAS signaling UP
- Estrogen response early
- UV response down
- Hypoxia
- Reptogen response late
- Myogenesis
Figure 2. Association of immune signatures with thyroid differentiation score in advanced PTC
Figure 3. Different way to regulate PI3K and MAPK signaling pathways according to partner genes of RET fusion
Figure 4. HOXD9 is a candidate gene associated with the recurrence of papillary thyroid cancer.

A. Recurrence vs non-recurrence (n=40 vs 242)

<table>
<thead>
<tr>
<th></th>
<th>PTC (n=282)</th>
<th>Age</th>
<th>Gender</th>
<th>Tumor Size</th>
<th>EOI</th>
<th>ETE</th>
<th>C_LNM</th>
<th>L_LNM</th>
<th>L_IWI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recurrence (n=40)</td>
<td>N.S.</td>
<td>N.S.</td>
<td>p &lt; 0.01</td>
<td>Larger</td>
<td>N.S.</td>
<td>p &lt; 0.01</td>
<td>Frequent</td>
<td>N.S.</td>
<td>p &lt; 0.01</td>
</tr>
</tbody>
</table>

p-value < 0.01 (Fisher's exact test)

B. Recurrence (n=40)

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HOXD9</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No_RF</td>
<td>(8)</td>
<td>(13)</td>
<td>(10)</td>
<td>(5)</td>
<td>(5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ETE+L_LNM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Larger+L_LNM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Larger+ETE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Larger+F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L_LNM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C. 15690 Genes

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>high enriched in No_RF + All_RF (p &lt; 0.05)</td>
<td>1737</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>high enriched in Others group (p &lt; 0.05)</td>
<td>118</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DFS of TCGA-THCA from GEPIA DB (p &lt; 0.01)</td>
<td>HOXD9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

D. Disease Free Survival

E. HOXD9-high (n=141)

n=30 n=111

HOXD9-low (n=141)

n=10 n=131

F. HOXD9-high enriched signaling pathways

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NF-κ B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T-cell receptor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-cell receptor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemokine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

G. KTC cohort (n=282) vs TCGA-THCA cohort (n=482)

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HOXD9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E2F2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

H. 3 risk factors (tumor size, ETE, L_LNM)

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>TFs (TCF3, EZH2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOXD9 overexpression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NF-κ B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Recurrence