scientific reports

Check for updates

OPEN

Characterization of the genomic alterations in poorly differentiated thyroid cancer

Yeeun Lee[1,6], SeongRyeol Moon[2,6], Jae Yeon Seok[3], Joon-Hyop Lee[4], Seungyoon Nam[1,5]✉ & Yoo Seung Chung[4]✉

Poorly differentiated thyroid carcinoma (PDTC) is a subtype of thyroid cancer that has a high rate of metastasis or recurrence and a relatively poor prognosis. However, there are few studies that have been conducted on PDTC at the whole protein-coding gene scale. Here, we performed genomic profiling of 15 patients with PDTC originated from follicular thyroid carcinoma using whole exome sequencing and also performed gene functional enrichment analysis of differentially expressed genes (DEGs) for three patients. Further, we investigated genetic variants associated with PDTC progression and the characteristics of clinical pathology. We revealed somatic genomic alterations in the *RAF1*, *MAP2K2*, and *AKT2* genes that were not reported in previous studies. We confirmed frequent occurrences in the *RAS* gene in patients with PDTC; the genetic alterations were associated with the RAS-RAF-MEK-ERK/JNK, PI3K-AKT-mTOR signaling pathways, and the cell cycle. DEG analysis showed that immune response was lower in cancer tissues than in normal tissues. Through the association analysis of somatic mutations and the characteristics of clinical pathology from patients with PDTC, the somatic mutations of *ABCA12*, *CLIP1*, and *ATP13A3* were significantly associated with a vascular invasion phenotype. By providing molecular genetic insight on PDTC, this study may contribute to the discovery of novel therapeutic target candidates.

Thyroid cancer is the most common type of endocrine tumor, and its incidence is continuously increasing, and it is the most rapidly increasing cancer globally[1]. More than 90% of thyroid cancer is differentiated thyroid cancer (DTC), which originates from the thyroid follicular epithelial cell; approximately 5% is medullary thyroid cancer (MTC), which originates from the nearby parafollicular C cells[2]. Compared to DTC such as papillary thyroid carcinoma (PTC) and follicular thyroid carcinoma (FTC), which differentiate well by pathological standards and have good prognosis, poorly differentiated thyroid carcinoma (PDTC) and anaplastic thyroid carcinoma (ATC) have a relatively poor prognosis and a higher risk of metastasis or recurrence[1].

PDTC originates from follicular cells, and its morphology and clinical behavior resemble both DTC and ATC[3]. The term PDTC was previously used to collectively represent various forms of cancer, including solid, trabecular, insular, poorly differentiated, intermediate-type, and primordial cell cancer[4]. However, it has been classified as a new type of cancer by the World Health Organization[4]. Although PDTC is rare, it is very aggressive, and because of its relatively poor prognosis, it is clinically important[3]. Since the diagnosis of PDTC is difficult with the fine needle aspiration biopsy (FNAB) performed prior to surgery, preoperative diagnosis is often delayed[5,6]. It has many similarities to FTC, histologically, and surgery may often be performed twice for treatment[7]. Moreover, because of the characteristics that appear throughout the stepwise dedifferentiation of PDTC[4,8,9], it shares genomic features with other thyroid cancer subtypes. Therefore, it is difficult to study the genomic characteristics inherent to PDTC[3].

[1]Department of Genome Medicine and Science, AI Convergence Center for Medical Science, Gachon Institute of Genome Medicine and Science, Gachon University Gil Medical Center, Gachon University College of Medicine, Dokjeom-ro 3Beon-gil, 38-13, Namdong-gu, Incheon 21565, Republic of Korea. [2]A.I. Structural Design Team, Division of Biodrug Analysis, New Drug Development Center, OSONG Medical Innovation Foundation, Cheongju-si 28160, Chungcheongbuk-do, Korea. [3]Department of Pathology, Yongin Severance Hospital, Yonsei University College of Medicine, Yongin Severance Hospital 363, Dongbaekjukjeon-daero, Giheung-gu, Yongin-si, Gyeonggi-do 16995, Korea. [4]Department of Surgery, Gachon University Gil Medical Center, Gachon University College of Medicine, Dokjeom-ro 3Beon-gil, 38-13, Namdong-gu, Incheon 21565, Republic of Korea. [5]Department of Health Sciences and Technology, Gachon Advanced Institute for Health Sciences and Technology (GAIHST), Gachon University, Incheon 21999, Korea. [6]These authors contributed equally: Yeeun Lee and SeongRyeol Moon ✉email: nams@gachon.ac.kr; dryooseung@gilhospital.com

As the incidence of thyroid cancer increases, there have been active studies on the genomic profiling of patients with thyroid cancer to develop more effective treatment methods[1,10]. With the introduction of next-generation sequencing analysis, the understanding of the genomic landscape of PTC, FTC, and general thyroid cancer has improved[1,10]. In PTC, which is the most common subtype, the *BRAF* alteration occurs at approximately 70% of the time[10]. In FTC, it is known that *RAS (N/H/KRAS)* gene alteration is the most common. However, due to its rare occurrence, sample collection and studies on the analysis of genetic information for PDTC are limited[3]. Despite this, a few studies have attempted to check genomic alterations of PDTC through targeted gene panel sequencing[9,11–13]. There have been a few studies that have examined the genomic landscape of PDTC by whole exome sequencing (WES), but the analyses were performed in few patients[9,14,15].

In this study, we aimed to identify somatic genomic alterations specific to PDTC and analyzed the WES data on tumor tissue from patients with PDTC against matched normal thyroid tissue to investigate genetic events that impact the known oncogenic signaling pathways. We also identified differentially expressed genes (DEGs) from tumor tissues and the normal tissues nearby through RNA sequencing (RNA-Seq) in three patients; for biological function analysis of DEGs, we performed gene ontology (GO) enrichment analysis and gene set enrichment analysis (GSEA). Lastly, we discovered somatic mutations relevant to tumor progression and clinicopathophysiologic characteristics.

Results

WES reveals the genomic landscape of 15 Korean patients with PDTC

We performed WES on tumor tissues from 15 patients with PDTC originated from FTC and the matched, normal thyroid tissues. We identified 3,446 somatic coding gene mutations (single nucleotide variant (SNV): 3,100; insertion and deletion (INDEL): 346) and copy number alterations (CNAs) in 15 patients with PDTC through WES. The range of mutations per patient was 45-13 (SNV: 32-543, INDEL: 2-108), and the median was 260 (SNV: 184, INDEL: 6; Fig. 1A, Supplementary Tables S1 and S2). For the *RAS* (*H/K/NRAS*) genes, mutations were found in eight patients (53.3%) (Fig. 1B), and this is more frequent than previously reported (28%)[12]. The *BRAF* mutation, whose frequency was reported as 33% in a previous study[12], was not found in patients in this study (Fig. 1B). The *RAS (H/K/NRAS)*'s copy number gain or amplification was found in four patients (Fig. 1B; Supplementary Figure S1). The WES in our patients also revealed somatic genomic alterations in the *RAF1, MAP2K2* and *AKT2* genes, which were not reported in previous targeted gene panel sequencing in PDTC[9,11–17] (Fig. 1B). Also, we have examined the predominant patterns in 15 PDTC patients. Among our patients, a solid pattern was observed in 13 patients, an insular pattern in 5 patients, and a trabecular pattern in 9 patients (Fig. 1C).

Mutalisk[18] was used to confirm the somatic mutation signatures in 15 patients with PDTC at Gil Medical Center (GMC). There was a correlation between age at diagnosis of carcinoma in 12 patients, and the single base substitution (SBS) 1 signature[19] related to endogenous mutational processes generated from the spontaneous or enzymatic deamination of 5-methylcytosine was identified. Of these, we confirmed that there was a high contribution of > 50% in two patients (Fig. 2A). We confirmed the association with homologous, recombination-based DNA damage repair in 13 patients; SBS3 signatures[19,20] commonly found in breast, pancreatic cancer, and ovarian cancers were found. The contribution was 50–70% in nine of the 13 patients (Fig. 2A). We noted that SBS5 signature[21] related to aging, smoking, and nucleotide excision repair (NER) deficiency contributed 15–60% in five patients (Fig. 2A). The SBS12 signature[21], which contributed a small portion (5–30%) in seven patients, had an unknown cause, and has a low contribution (< 20%) in liver cancer. Of the doublet base substitution (DBS) signatures, DBS3 signature, which is known to be related to the polymerase epsilon exonuclease domain, was found to contribute to all 15 patients (Fig. 2B). Of these, it had a high contribution rate (50–100%) in 11 patients. In addition, the DBS9 signature[22], which had an unknown cause, contributed some portion (10–55%) in seven patients; DBS11[22], also with unknown cause, contributed > 50% in 13 patients (Fig. 2B). Most of the insertion and deletion (ID) signatures had unknown causes, and the ID5, ID10, and ID16 signatures[22], which had major contributions in this cohort of patients, also had no known cause. The ID5 signature that contributed to 12 patients is known to be related to the age at diagnosis. The ID16 signature, which contributed to 14 patients, was identified to contribute to ovarian cancer lacking SBS signatures related to DNA mismatch repair deficiency[22] (Fig. 2C).

Comparison between previously known thyroid cancer pathways and the somatic alterations of the patients with PDTC revealed oncogenic signaling pathways

To confirm the oncogenic pathways relating to the patients with PDTC at GMC, we annotated the somatic genomic alterations (i.e., mutations and CNAs) of the patients in 13 canonical cancer-associated pathways[23] (Fig. 3A). A mutated pathway that was found in most of the patients was 'RAS-RAF-MEK-ERK/JNK signaling', and somatic genomic alterations in the pathway were confirmed in 12 of the 15 patients. In addition, somatic genomic alterations were found in more than 60% of patients for 'cell cycle', 'PI3K-AKT-MTOR signaling', and 'RTK signaling family' pathways. Interestingly, two patients (P03 and P04) did not have any somatic genomic alterations of any of the 13 canonical cancer-associated pathways. The 'RAS-RAF-MEK-ERK/JNK signaling' and 'PI3K-AKT-MTOR signaling' pathways are typical oncogenic signaling pathways in thyroid cancer[24]. In order to compare known thyroid cancer oncogenic pathways with somatic genomic events of the patients with PDTC, we mapped mutations (SNVs and INDELs) and CNAs from 15 patients with PDTC to the oncogenic pathways known for thyroid cancer (Fig. 3B). Thyroid cancer is known to affect the MAPK and PIK3CA/AKT pathways, which contribute to cell proliferation and growth through the RAS signaling pathway[25]. Upon analyzing the pathways, 13 of the 15 patients were found to be affected by the known thyroid cancer oncogenic pathways, and 10 of these patients had somatic mutations, copy number gain or copy number amplification of the *RAS* (*H/K/NRAS*) gene, which are upstream of the MAPK and PIK3CA/AKT pathways (Fig. 3B).

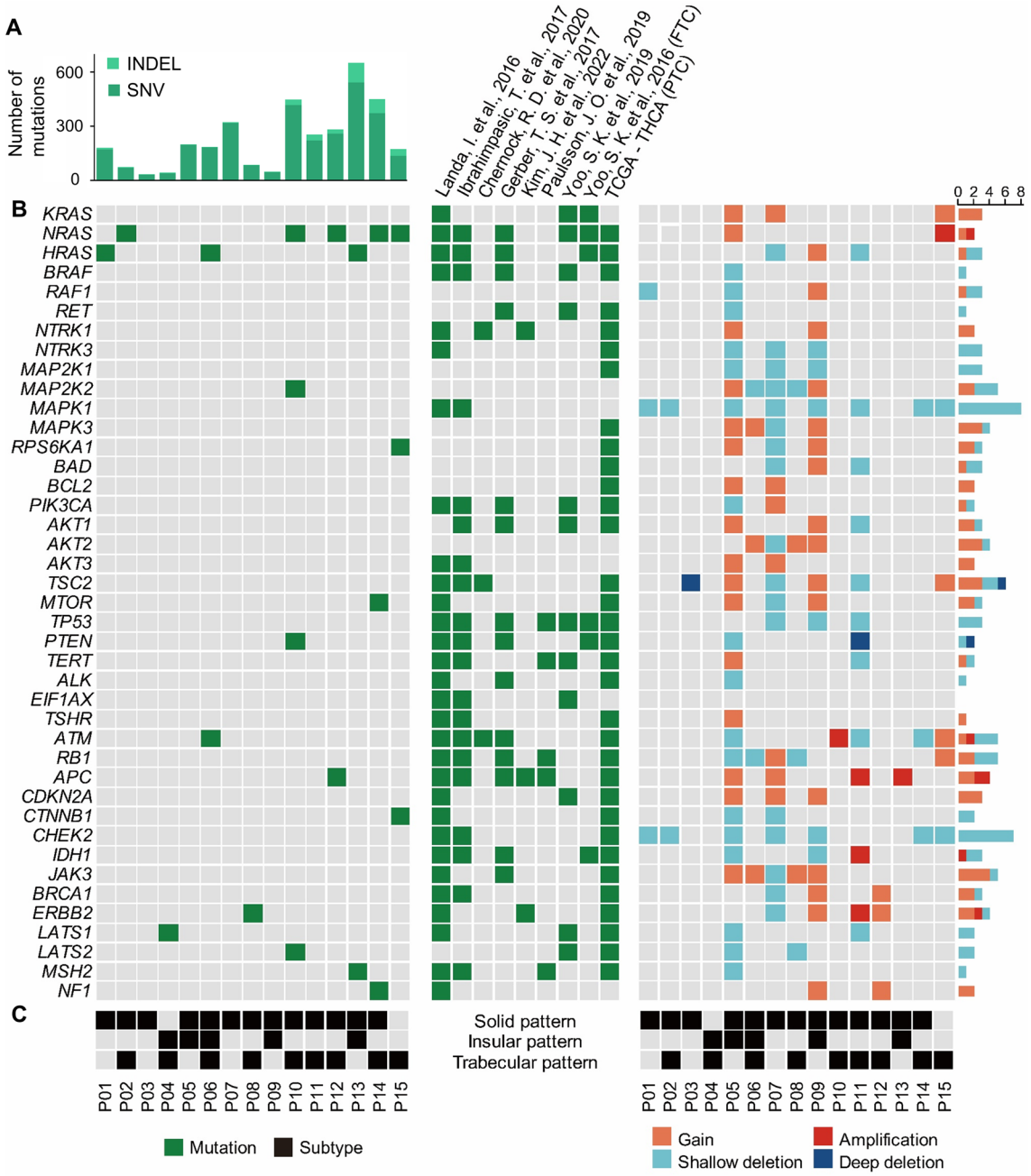


Figure 1. WES reveals the genomic landscape of 15 patients with PDTC and compares them with somatic mutation profiles from published studies. (**A**) Number of somatic mutations in our patients with PDTC. The numbers of SNVs and INDELs for 15 patients with PDTC were depicted. (**B**) Somatic mutation profiles and CNAs of our 15 patients with PDTC were in the left and right panels, respectively. Somatic mutation profiles of previously published studies in PDTC, FTC and PTC were in the middle panel. Genes on the Y-axis belongs to a curated gene set that is reportedly involved in major cancer pathways in thyroid cancer. Columns in the left and right panels refer to individual patients and those in the middle panel individual publications. (**C**) The clinical information on the three subtypes for our 15 patients with PDTC is indicated.

Functional analysis of DEGs for PDTC tumor tissue and the nearby normal tissues show changes in functions related to cell adhesion and immune response in PDTC

To analyze the function of DEGs in patients with PDTC, we performed GO enrichment analysis and GSEA. Compared to the matched, normal thyroid tissue, 4,668 genes (2,224 genes upregulated, 2,444 genes downregulated) from tumor tissues in three of the patients with PDTC at GMC were significantly differentially expressed. We used g: Profiler tools[26] to separate the GO of the DEGs (Fig. 4 and Supplementary Figure S2). According to the biological process (BP) term analysis, DEGs upregulated in the tumor tissues (compared to the matched, normal thyroid tissues) were enriched in terms that are related to cell adhesion, such as 'homophilic cell adhesion via plasma membrane adhesion molecules', 'cell–cell adhesion via plasma-membrane adhesion molecules' or in the

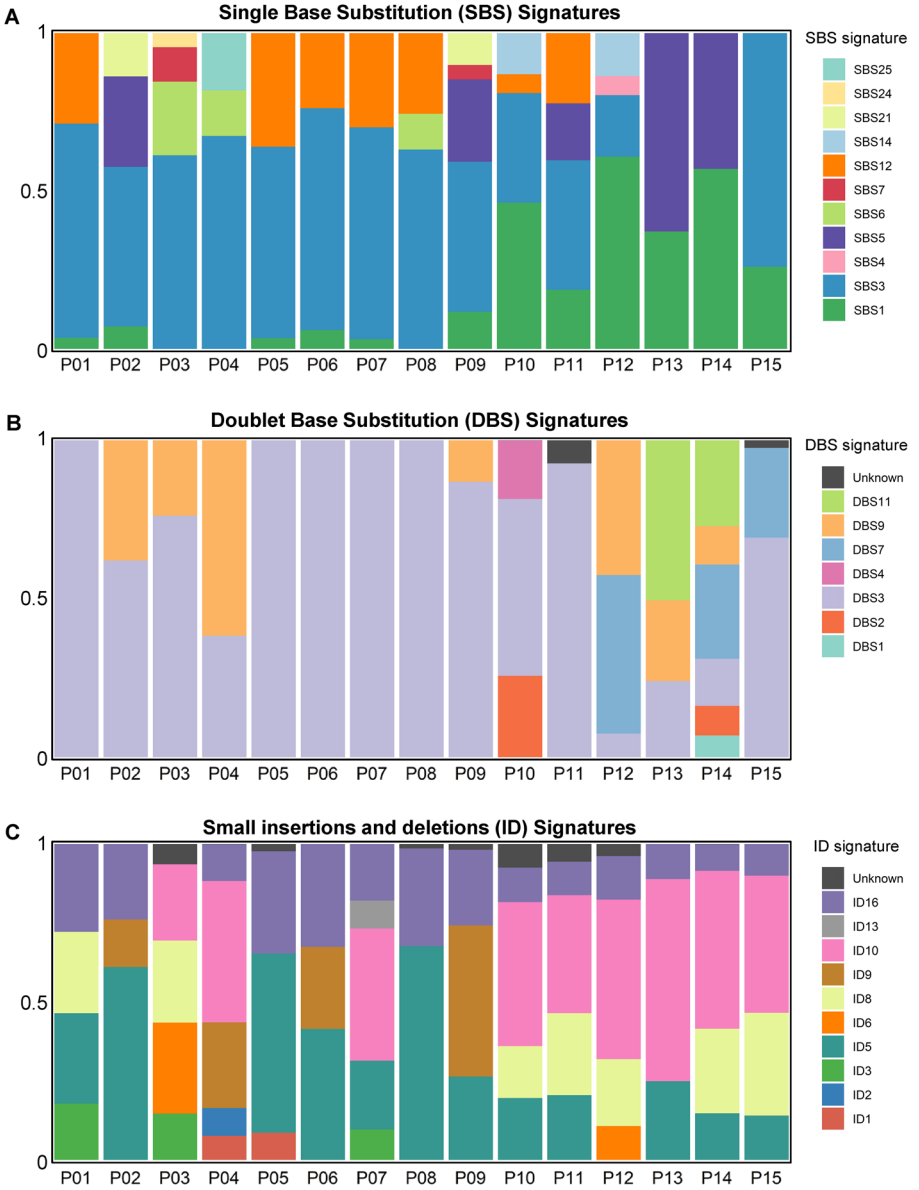


Figure 2. Relative contribution of somatic mutation signatures from 15 patients with PDTC. (**A**) Relative contribution of SBS signatures. SBS1, SBS3, SBS5, and SBS12 were mainly observed in the patients. (**B**) Relative contribution of DBS signatures. DBS3 was contributed to PDTC. **(C)** Relative contribution of ID signatures. ID5, ID10, and ID16 were involved in PDTC.

terms related to normal cell activity and development, such as 'regulation of cellular process' and 'developmental process' (Fig. 4A). Compared to the matched, normal thyroid tissue, DEGs downregulated in the tumor tissues were mainly enriched in the terms related to immune responses, such as 'immune system process', 'regulation of immune system process', and 'T cell activation' (Fig. 4B). We performed GSEA to evaluate the difference in the transcription profiles between tumor tissues in patients with PDTC and the matched, normal thyroid tissues. Compared to the matched, normal thyroid tissue, gene sets upregulated in tumor tissues were not found; 15 gene sets, including 'DNA repair' and 'apoptosis', were downregulated in tumor tissues compared to the matched, normal thyroid tissue (Fig. 4C). Furthermore, we evaluated potential immune checkpoint blockade (ICB) responses in the three patients using the Tumor Immune Dysfunction and Exclusion (TIDE) algorithm[27,28] (Supplementary Table S3). All three patients were predicted as non-responders.

Neoantigens were predicted from somatic mutations in PDTC

Tumor neoantigen burden (TNB) can affect the response to immune checkpoint inhibitor (ICI) treatment in cancer patients[29]. Tumor mutation burden (TMB) has a correlation with TNB[30], and it is related to the prognosis of PTC[31]. The average TMB (somatic nonsynonymous mutation counts per Mb) in our patients with PDTC was 6.05 (median 4.89, range: 0.95–17.13; Fig. 5A), and it was higher than TCGA PTC (average 0.41)[1]. In order to

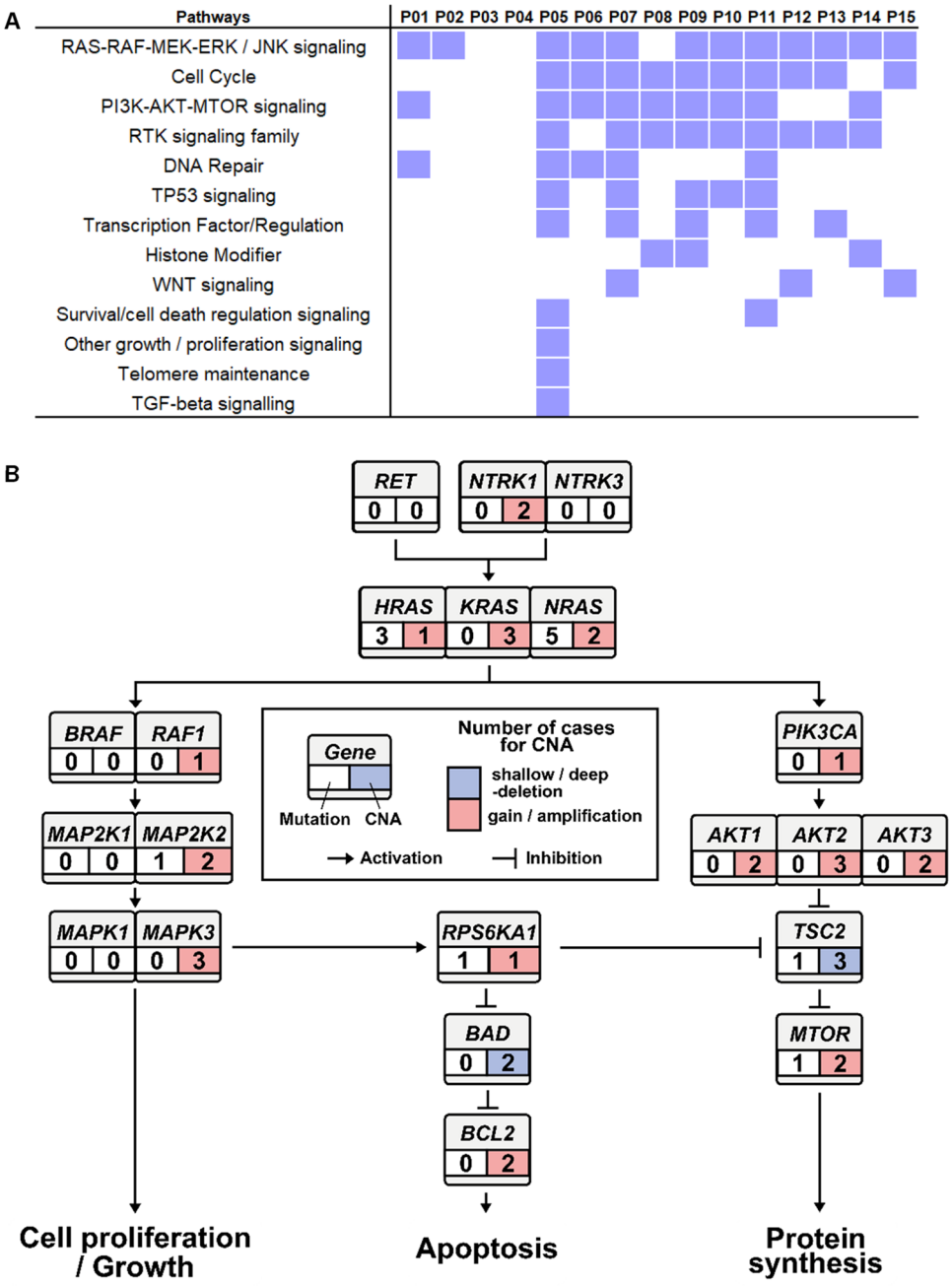


Figure 3. Altered canonical cancer-associated pathways of somatic genomic alterations in patients with PDTC. Somatic genomic alterations (i.e., mutations and CNAs) from the WES data of 15 patients with PDTC were mapped against the major cancer pathways, and significant genomic events were observed. (**A**) Somatic genomic alterations from 13 canonical cancer-associated pathway genes[23] were confirmed from WES data of 15 patients with PDTC at GMC; genomic events frequently occur at the RAS-RAF-MEK-ERK/JNK signaling pathway. (**B**) Somatic genomic alterations from 15 patients with PDTC at GMC were mapped against the commonly known thyroid cancer oncogenic pathways (BRAF and MAPK pathways) and the alteration of the *RAS* gene was found to be located upstream of the RAS-RAF-MEK-ERK/JNK signaling pathway. The numeral indicates the number of cases having the somatic alterations for the gene. For example, copy number gain/amplification for *BRAF* was observed in one patient, and mutations for the gene were not observed in the patients.

explain the conditions for neoantigens in PDTC, we extracted neoantigens with binding affinities of < 500 nM using pVACseq[32] from the WES data of 15 patients with PDTC at GMC. Of the 2,825 mutations (SNV and INDEL) found in the 15 patients, 58 (1.7%) had neoantigens (IC50 < 500 nM), and there were 54 neoantigens from SNV and four neoantigens from INDEL mutations. With respect to the binding affinity with MHC, there

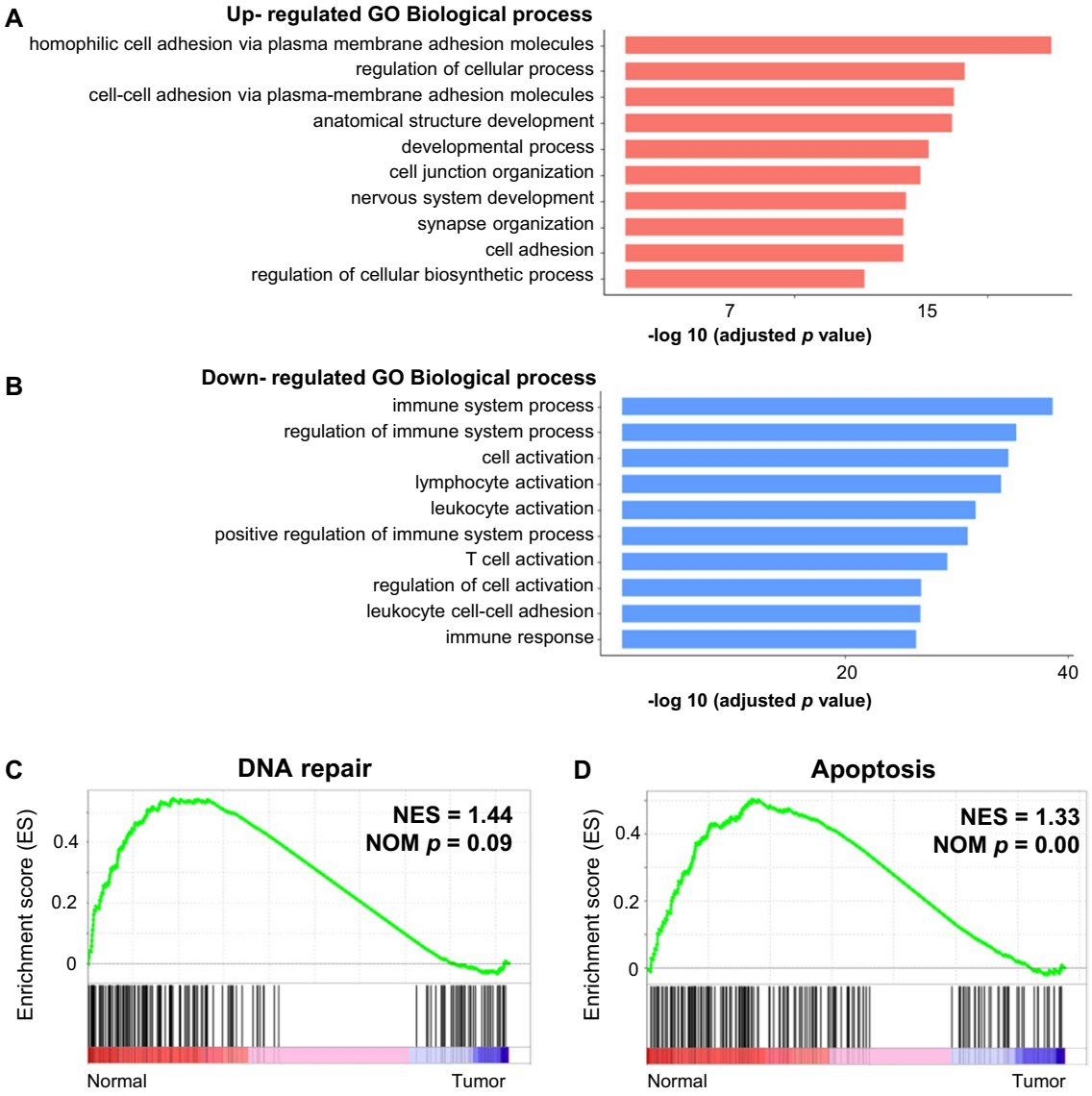
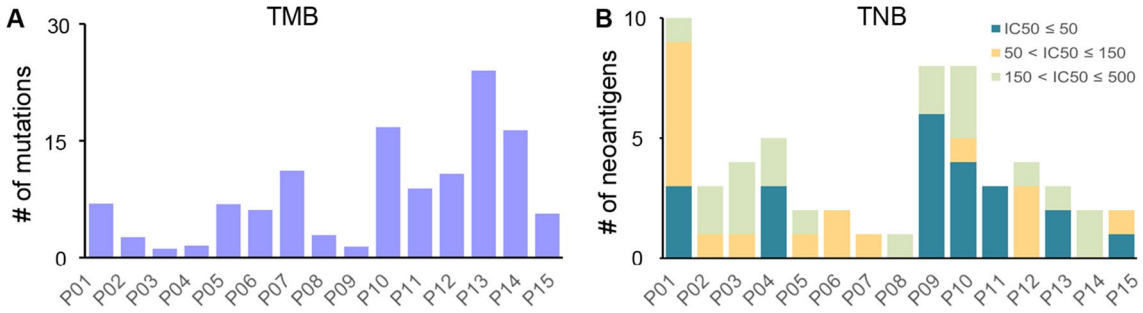


Figure 4. Differentially expressed Genes in PDTC are associated with cell adhesion and immune response. Results from the GO enrichment analysis and GSEA performed using RNA-seq data from three patients with PDTC. (**A**) GOBP enrichment analysis of DEG that are upregulated in PDTC tumor tissue compared to the matched, normal thyroid tissue. They are enriched in terms related to cell adhesion. (**B**) GOBP enrichment analysis of DEG downregulated in PDTC tumor tissue compared to the matched, normal thyroid tissue. They are commonly enriched in terms related to the immune response. The graphic shows the top ten enriched ($p < 0.05$) GOBP terms, starting with the lowest *P*-value. (**C–D**) GSEA results for the hallmark gene sets using RNA-seq data from three patients with PDTC. The graphic shows a significant negative association for DNA repair (**C**) and apoptosis (**D**) for tumor tissues compared to the matched, normal thyroid tissue.

were 22 neoantigens extracted with a high affinity (< 50 nM), 17 neoantigens with a medium affinity (< 150 nM), and 19 neoantigens with a low affinity (< 500 nM; Fig. 5B). Of the neoantigens with a strong affinity, four originated from tumor related genes (*ARID2, DICER1, PAX8*, and *PTEN*; Fig. 5C).

Somatic mutations of PDTC are potentially related to the clinicopathological characteristics

One-way ANOVA and Fisher's exact test were performed to analyze the relationship between clinicopathological somatic mutation genes from 15 patients with PDTC at GMC. Analysis was performed for genes (n = 100) that had somatic mutations in at least three patients. Mutation status and clinicopathological characteristics were analyzed using one-way ANOVA, and six clinicopathological phenotypes were significantly related to 21 genes (Fig. 6A). Fisher's exact test showed that four clinicopathological phenotypes were related to nine genes ($p < 0.05$; Figs. 6B, 6C, and 6D; Supplementary Figure S3). *ABCA12, CLIP1,* and *NRAS* were related to trabecular pattern (Supplementary Figures S3A, B, and C); *ABCA4* and *SLIT3* were related to insular pattern (Supplementary Fig. S3D and E); *ABCA12, ATP13A3*, and *CLIP1* genes were related to the status of vascular invasion (Figs. 6B–D);



C

Patients	Gene	AA_change	HLA Allele	MT Epitope Seq	Variant Type
P01	*FGD1*	D455Y	HLA-A*24:02	RYPRIGYIL	missense
	CCDC170	H127P	HLA-B*40:01	TEITAPAAI	missense
	ARID2	L94LX	HLA-C*03:04	FALKQYYLAL	FS
P04	*PCDHGA4*	R403Q	HLA-A*24:02	TYGNYYQLL	missense
	AGAP3	S829Y	HLA-A*24:02	LYSAMANVVF	missense
	PAX8	S214X	HLA-A*03:01	RPMPPPATPK	FS
P09	*GCN1L1*	S287L	HLA-A*26:02	ETISSLLAL	missense
	CWH43	G188R	HLA-A*33:03	EVATRMASR	missense
	CYHR1	V311X	HLA-A*33:03	WSATSCWLPR	FS
	GPR112	D598V	HLA-A*26:02	ETIWTSRPV	missense
	ACACA	D476Y	HLA-B*15:01	FAYSQFGHCF	missense
	DICER1	E1813A	HLA-B*15:01	FASLAGAIY	missense
P10	*SMARCA2*	H706P	HLA-C*14:02	SYYTVAPAI	missense
	PTEN	D92H	HLA-B*54:01	YPFEHHNPP	missense
	AMOTL1	A367V	HLA-B*54:01	LPLPMALGV	missense
	INPP4A	E725K	HLA-C*14:02	STYGEKLAM	missense
P11	*TBCEL*	G97V	HLA-B*40:01	KEHVLVPSL	missense
	SORL1	N828D	HLA-B*40:01	QEVIIDSGL	missense
	OR2A42	H55P	HLA-A*24:02	RLPTPMYFF	missense
P13	*VPS8*	H172P	HLA-C*03:04	IAVGTSPGL	missense
	NOTCH3	N1588H	HLA-B*40:02	LEIDHRLCL	missense
P15	*DRD2*	T343I	HLA-A*02:01	IQMLAIVLGV	missense

Figure 5. Characteristics of neoantigens predicted from somatic mutations in PDTC. Results of neoantigen prediction analysis using WES data from 15 patients with PDTC at GMC. (**A**) TMB in this patient population is represented. (**B**) The number of neoantigens that have MHC binding affinity < 500 nM in each patient is represented. The shaded area represents MHC binding affinity. **C** List of neoantigens that have a strong affinity (IC_{50} < 50 nM) with MHC predicted from patients with PDTC.

INTS1 and *RYR1* genes were related to the status of capsular invasion (Supplementary Figure S3F and G); and the *SORL1* gene was related to the status of hypertension (Supplementary Figure S1H).

Discussion

Our study aimed to shed light on the genetic landscape of PDTC, an area with limited existing knowledge. To achieve this, we conducted WES and RNA-Seq to uncover the molecular genetic characteristics of PDTC and to explore their associations with the disease. To comprehensively analyze these aspects, we employed various bioinformatics tools and methodologies, including mutation profile analysis, oncogenic pathway analysis, assessment of tumor mutation burden, prediction of neoantigens, DEG analysis, gene ontology enrichment analysis, and investigations into the associations between clinical parameters and somatic mutation profiles. It is worth noting that our study stands out in terms of WES analysis by utilizing data from 15 patients of PDTC originated from FTC, representing a relatively large sample size compared to previous PDTC WES studies[9,14,15]. This substantial dataset allowed us to draw robust conclusions and identify potential clinical implications.

In this study, the *RAS* gene mutation was frequently seen in our patients, and the *BRAF* mutation[12], which is typically known in thyroid cancer, was not seen in our patients. All of the PDTC in our patient population

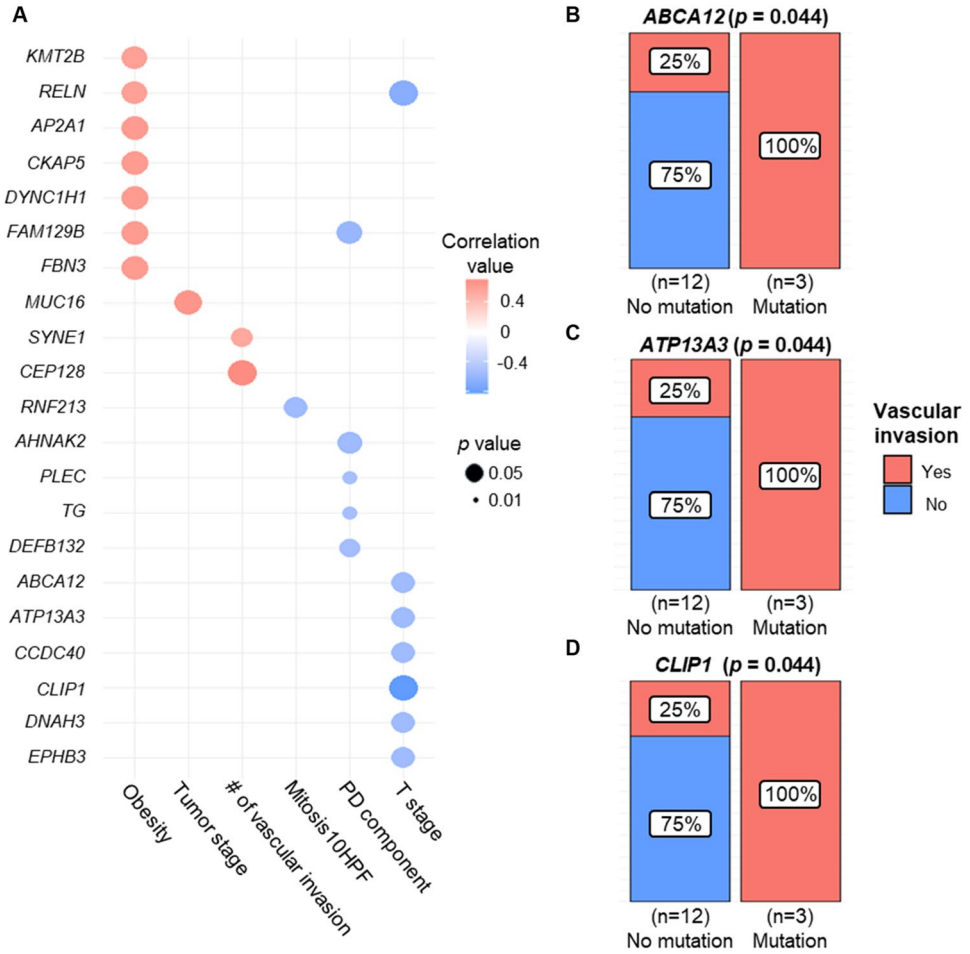


Figure 6. Somatic mutation genes potentially related to PDTC clinicopathological characteristics. We performed association analysis between clinicopathological characteristics of 15 patients with PDTC and the somatic mutation genes. **A** Clinicopathological characteristics and somatic mutation genes with a statistical significance ($p < 0.05$) using one-way ANOVA results. The Y-axis indicates the somatic mutation genes and the x-axis the clinicopathological characteristics. # of vascular invasion, number of foci of vascular invasions; Mitosis10HPF, Mitotic count per 10 high power fields; PD component, the proportion of poorly differentiated components in total tumor tissue. (**B**–**D**) Clinicopathological characteristics and somatic mutation genes that have a statistical significance ($p < 0.05$) using Fisher's exact test. Vascular invasion and *ABCA12* (**B**), *ATP13A3* **(C)**, *CLIP1* (**D**) are significantly correlated.

originated from FTC, and FTC has a characteristic of low *BRAF* gene mutation frequencies[33]. Genomic events were found in the thyroid cancer, MAPK via RAS signaling, and PIK3CA/AKT signaling pathways in 13 patients[25].

In Fig. 1B, the *RAS (H/N/KRAS)* genes are oncogenes, which are often mutated. The genes are related to cell proliferation and survival[33]. *RAS* genes are located at the upstream of the MAPK and PIK3CA/AKT signaling pathways. The MAPK pathway mutation is found in most cancers, and the PIK3CA/AKT pathway is known to be upregulated in thyroid cancer[34]. In Fig. 1B, *MAP2K2* is a gene that encodes mitogen-activated protein kinase kinase 2 (MEK2) and regulates cell growth and proliferation by phosphorylating ERK1/2 in the MAPK pathway, thereby activating it[35]. We found a somatic mutation or copy number gain in *MAP2K2* in three patients (P05, P09, and P10 in Fig. 1B). It is noted that the somatic mutation in *MAP2K2* was not previously reported in targeted gene panel sequencing of PDTC[9,11–13].The *MAP2K2* mutation is related to a high resistance in MEK inhibitors in melanomas with a *BRAF* mutation[35], and *MAP2K2* upregulation is known to induce the progress of clear cell renal carcinoma[36]. However, its oncogenic role in thyroid cancer is unknown.

Cancer immune escape is the most important barrier in improving cancer treatment strategies[37]. Based on the results of GO enrichment analysis using RNA-Seq data from three patients (P02, P03, and P04) with PDTC, genes that were downregulated in PDTC tumor tissues compared to normal thyroid tissues were enriched in the terms related to the immune responses of GO biological processes (GOBP in Fig. 4B). All three patients demonstrated relatively low TMB and high TIDE scores, implying a higher likelihood of tumor immune evasion. This suggests a lower probability of benefiting from treatment with immune checkpoint inhibitors[27,28]. Immune response-related genes that were downregulated in the tumor tissues of patients with PDTC compared to normal

thyroid tissues included *TAP1, AIM2,* and *MICB. TAP1* downregulation is associated with immune evasion and unfavorable prognosis in rectal cancer[38], and *AIM2* is a known tumor suppressor gene in melanoma[39]. *MICB* is a gene that codes the major histocompatibility complex class I related B (MIC-B), and it is related to the innate immune system[40]. *MIC-B* downregulation is the cancer cell immune evasion strategy to avoid NK-mediated cytotoxic killing[41], and *MIC-B* downregulation is associated with poor prognosis of colorectal cancer[42].

TMB is known as a biomarker for immune checkpoint blockade (ICB) response[43]. Well-differentiated thyroid cancers, such as PTC and FTC, exhibit relatively lower TMB compared to other cancer types[44]. Progressive thyroid cancer subtypes like PDTC have higher TMB, with an even higher TMB observed in ATC[44]. Previous studies have proposed an optimal TMB cut-off (> 13 mut/Mb) for ICB response. Among our 15 patients, three (P10, P13, and P14) exceeded this cut-off, each having TMB values of 16.8, 24.0, and 16.4 mut/Mb, respectively, indicating likely hyper-mutated cases. According to previous ATC research, such a 'hypermutator phenotype' may arise in the presence of mutations in DNA mismatch repair (MMR) genes[45], and one (P13) of our three patients indeed had a mutation in a DNA MMR gene, *MSH2*.

In Fig. 5, we detected neoantigens that originated from four cancer related genes (*ARID2, DICER1, PAX8,* and *PTEN*) using neoantigen prediction from WES data and 22 neoantigens had strong affinity. *ARID2* is a tumor suppressor gene, and its mutants occur in various cancer types, including stomach cancer, melanoma, and colorectal cancer[46]. Tumors with *ARID2* mutants can potentially be more sensitive to immune-checkpoint inhibition[47]. *DICER1* is an important tumor suppressor gene in thyroid cancer[48], and its mutants have been reported in thyroid cancer[1,9]. *PAX8* is an important gene in thyroid cell development[49], and PAX8/PPARγ fusion protein from the translocation is a driver mutation that is often seen in FTC[33]. *PTEN* is a tumor suppressor gene that is commonly mutated in human cancers[50]. Loss of *PTEN* function can contribute to the development of cancer by activating PI3K/AKT/mTOR signaling[50].

Recurrence and metastasis occur more commonly in PDTC relative to FTC or PTC[8]. It has been reported that patients with PDTC with vascular invasion have a higher chance of recurrence or metastasis in comparison to patients with PDTC without vascular invasion[51], and we confirmed that six patients with PDTC had vascular invasion (Fig. 6B). Given the significant correlation observed in the association analysis (Fig. 6B) between *ABCA12*, *ATP13A3*, and *CLIP1* somatic mutations and vascular invasion, these somatic mutations may hold promise as potential biomarkers for predicting the recurrence or metastasis of PDTC following surgery. Notably, *ABCA12* knockdown resulted in decreased cell motility, proliferation, and invasion in pancreatic cancer cells, such as SW199[52]. Among the observed *ABCA12* mutations in three patients (P01, P05, and P07), the presence of these mutations suggests a potential gain-of-function effect on cell motility.

ATP13A3 was recently suggested to be a major mammalian polyamine transporter, and it was found to be expressed at a high level in pancreatic cancer patients, thereby reducing survival rates[53]. CLIP1 encodes a protein that controls the dynamics of microtubule cytoskeleton in cells[54] and its mutations are known carcinogenic factors in non-small-cell lung cancer[54].

One limitation of our study is the small sample size for RNA-Seq analysis. We conducted gene expression profiling in only three patients (P02, P03, and P04), who exhibited relatively lower TMB when compared to the remaining 12 patients (Figs. 1A and 3A). Consequently, the gene expression profiling findings in our study might not accurately reflect the characteristics of Korean patients with PDTC. Therefore, it is advisable to exercise caution when interpreting these results. Nevertheless, the DEG ontology analysis performed in the study may expand the understanding of PDTC. Validation of the findings through larger cohort studies and additional studies for the profiling of tumor from single cell RNA-Seq will be needed in the future.

In our patients, genetic alterations within known oncogenic pathways were observed to occur more frequently in copy number alterations (CNAs) than in point mutations. This suggests the potential involvement of CNAs as driver events in tumorigenesis[55]. However, we were unable to simultaneously evaluate both copy number changes and alterations in gene expression levels. Previous studies have indicated a higher likelihood of genes with DNA copy number amplification and elevated gene expression levels being associated with oncogenes[55]. Validation of our findings will be necessary through larger-scale cohort studies and further research involving multi-omics tumor profiling, including single-cell RNA-Seq, in the future.

Conclusion

In conclusion, we identified that somatic mutation genes in patients with PDTC are associated with the clinicopathological characteristics of the tumor, and we confirmed the possibility of immune escape through transcriptome analysis. These findings provide insights into novel treatment strategies for PDTC.

Materials and methods

Study subjects and patients

This study included 15 patients with PDTC originated from FTC at Gachon University Gil Medical Center (GMC), South Korea. Primary tumor and matched, adjacent normal thyroid tissue were collected from the 15 patients for mutation profiling and mRNA expression profiling. Clinicopathological characteristics of the patients are summarized in Table 1. Extrathyroidal extension was found in one patient, and capsule formation was found in 14 patients. Ten patients had capsular invasion, and vascular invasion was confirmed in six patients. None of the 15 patients had lymph node metastasis.

The study was approved by the GMC Ethics Committee (GBIRB2020-098) and conducted in accordance with the Declaration of Helsinki. All participants provided written informed consent.

Patient characteristics	N
Age	
≥ 60	5
< 60	10
Sex	
Female	10
Male	5
Mean tumor size (range), cm	4.8 (1.3–9.5)
Extrathyroidal extension	1
Capsule formation	14
Capsular invasion present	10
Vascular invasion present	6
Lymph node metastasis	0
Solid pattern	13
Insular pattern	5
Trabecular pattern	9
TNM stage	
I/II	13
III/IV	2

Table 1. Clinicopathological characteristics.

DNA extraction and sequencing analysis

For mutation profiling in 15 patients with PDTC, whole-exome sequencing (WES) was performed using tumor tissue and matched, normal thyroid tissue. The SureSelect XT V6 Library Prep Kit (Agilent, Santa Clara, CA, USA) was used as the exome library, and sequencing was performed using the Illumina sequencing platform (Macrogen Inc., Seoul, Korea; Supplementary Table S4).

Sequenced data (FASTQ) was trimmed (q > 30) according to the quality score using Sickle (v1.33) and was aligned for hg19 using BWA-mem (v0.7.15). Duplicate reads were confirmed and edited using Picard MarkDuplicates (v2.5.0). Using the Genome Analysis Toolkit (v3.6)[56], sorting errors from insertion and deletion (INDEL) were corrected and the quality score was edited. For BAM files, which were realigned and edited, Mutect2[56] was used along with the matched, normal thyroid tissue to call for the somatic single nucleotide variants (SNV) and INDEL. To remove low quality variants, we selected data with the 'PASS' label from the results of Mutect2. Somatic SNV and INDELs were annotated using Annovar (v2016-02-02).

To find somatic mutations, variants with ≥ 1% minor allele frequency (MAF) in more than one of the three germline population variant's databases (1000 Genomes Phase 3, ESP6500, ExAC) were removed. Finally, to select for the somatic coding mutation, we selected missense, nonsense, and INDEL mutations from the exon and splice sites. Tumor mutation burden was calculated using the somatic nonsynonymous mutation (SNV, indel) per mega base in the coding region (exon and splice region)[1].

Copy number alterations (CNAs)

According to the workflow recommended by VarScan (v2.4.4)[57], somatic CNAs were recalled from WES BAM files. Using the online software package, GenePattern GISTIC 2.0 tools, genomic level CNAs and recurrent events in the cohort were identified.

RNA sequencing analysis

RNA-Seq was performed using tumor tissue from three GMC patients with PDTC and matched normal thyroid tissue (Supplementary Table S5). An RNA-Seq library was prepared using the SureSelectXT RNA Direct Reagent Kit (Agilent, Santa Clara, CA, USA), and it was sequenced using the Illumina platform (Illumina, San Diego, CA, USA; Supplementary Table S6). The sequences were mapped against the known reference genome using the HISAT2 program[58], and transcript assembly was done using the StringTie. DEG analysis was performed using DESeq2.

Collection of publicly available PDTC and PTC data sets for comparison of mutation profiles

We collected mutation profile data from patients with PDTC and PTC available on cBioPortal[59]. Mutation profiles from the 341 major cancer genes of 84 patients with PDTC provided by the Memorial Sloan Kettering Cancer Center (MSKCC) data set[12] were collected. From The Cancer Genome Atlas (TCGA) data set[1], in which WES was performed in 315 patients with PTC, mutation profiles were collected. Additionally, we collected mutation profiles from previous studies that performed target sequencing in patients with PDTC[9,11–13].

Mutational signature analysis

Using Mutalisk[18], which is a somatic mutation analysis tool, single base substitution (SBS) signatures were confirmed. The SBS signature was dissembled using the SBS signature (v3.3) provided by the Catalogue of Somatic Mutations in Cancer (COSMIC)[60]. The relative contribution of the SBS signature was calculated using linear regression for all 78 SBS signatures. The variant matrix for doublet base substitution (DBS) signatures and small insertion and deletion (ID) signatures was composed using 'SigProfilerMatrixGenerator', and the contributions to DBS Signatures (v3.3) and ID signatures (v3.3) provided by COSMIC were calculated using the 'deconstructSigs (v1.9.0)' package in R software.

Human leukocyte antigen (HLA) typing and neoantigen prediction

Using the Polysolver program[61], the HLA type was determined from the human leukocyte antigen (HLA) class I gene (HLA-A, HLA-B, HLA-C) of 15 patients with PDTC at 4 × resolution. In order to predict the neoantigens, the pVAC-Seq pipeline[32] was used for analysis using the variant call format (VCF) file, which has ensemble variant effect predictor (VEP) annotation. Major histocompatibility complex (MHC) binding affinity for the whole peptide was estimated using the binding prediction algorithms for the pVAC-Seq pipeline, NetMHC and NetMHCpan (< 500 nM binding affinity). The pVec-Seq parameters were set as default, and the peptide length for the measurement of MHC and binding affinity was set as 8–10 amino acid (aa). We selected peptides that were 8–10 aa in length with < 500 nM for MHC and binding affinity, as well as those that appeared only in tumor tissues, as neoantigen candidates. We defined neoantigens with a binding affinity of < 50 nM as 'strong affinity', those with a binding affinity of < 150 nM as 'medium affinity', and those with a binding affinity of < 500 nM as 'weak affinity'.

Gene ontology (GO) enrichment analysis and gene set enrichment analysis (GSEA)

For the biological functional annotation of significantly differentially expressed genes from the tumor tissues of three patients with PDTC against the matched, normal thyroid tissue, we used the 'g:Profiler' tool[26] to perform GO enrichment analysis. Using tumor tissue from three patients with PDTC and the matched, normal thyroid tissue gene expression profiles, we performed gene set enrichment analysis (GSEA, v4.2.2)[62] for functional annotation. For the reference gene set, the hallmark gene set (v7.5.1)[62] was used, which includes the well-defined biological states or processes provided by the Molecular Signatures Database (MSigDB)[62]; all variables were set as default. We determined a False Discovery Rate (FDR) of < 0.25 as statistically significant[62].

Prediction of the immune checkpoint blockade (ICB) response

Predicting the Immune Checkpoint Blockade (ICB) Response the Tumor Immune Dysfunction and Exclusion (TIDE)[28,63] was developed to predict potential immune checkpoint blockade (ICB) responses. Using RNA-Seq data from three patients, we predicted ICB responses following the procedures outlined on the TIDE website in order to evaluate their response to immunotherapy.

Clinical correlation analysis between mutation profiles and clinicopathological data

To find somatic mutations that have significant statistical relevance to clinicopathological characteristics, we analyzed mutations profiles from 15 patients with PDTC and the clinicopathological data using Fisher's exact test and one-way ANOVA. A *P*-value < 0.05 was considered as statistically significant.

Ethics approval and consent to participate

The study was approved by the GMC Ethics Committee (GBIRB2020-098) and conducted in accordance with the Declaration of Helsinki.

Data availability

The WES and RNA-Seq data generated in this study are deposited in the NCBI Sequence Read Archive (SRA) Database (PRJNA883015).

Received: 26 July 2023; Accepted: 1 November 2023
Published online: 06 November 2023

References

1. Cancer Genome Atlas Research, N. Integrated genomic characterization of papillary thyroid carcinoma. *Cell* **159**, 676–690. https://doi.org/10.1016/j.cell.2014.09.050 (2014).
2. Sherman, S. I. Thyroid carcinoma. *Lancet* **361**, 501–511. https://doi.org/10.1016/s0140-6736(03)12488-9 (2003).
3. Tong, J. *et al.* Poorly differentiated thyroid carcinoma: a clinician's perspective. *Eur. Thyroid J.* **11**, https://doi.org/10.1530/ETJ-22-0021 (2022).
4. Asioli, S. *et al.* Poorly differentiated carcinoma of the thyroid: validation of the Turin proposal and analysis of IMP3 expression. *Mod Pathol* **23**, 1269–1278. https://doi.org/10.1038/modpathol.2010.117 (2010).
5. Rossi, E. D., Adeniran, A. J. & Faquin, W. C. Pitfalls in thyroid cytopathology. *Surg Pathol Clin* **12**, 865–881. https://doi.org/10.1016/j.path.2019.08.001 (2019).
6. Kim, N. R., Seok, J. Y., Chung, Y. S., Lee, J. H. & Chung, D. H. Contribution of cytologic examination to diagnosis of poorly differentiated thyroid carcinoma. *J Pathol Transl Med* **54**, 171–178. https://doi.org/10.4132/jptm.2019.12.03 (2020).
7. Baloch, Z. W. *et al.* Overview of the 2022 WHO classification of thyroid neoplasms. *Endocr Pathol* **33**, 27–63. https://doi.org/10.1007/s12022-022-09707-3 (2022).
8. Wen, D. *et al.* Dedifferentiation patterns in DTC: is PDTC an intermediate state between DTC and ATC?. *Int J Clin Exp Pathol* **12**, 267–274 (2019).
9. Chernock, R. D. *et al.* Poorly differentiated thyroid carcinoma of childhood and adolescence: a distinct entity characterized by DICER1 mutations. *Mod Pathol* **33**, 1264–1274. https://doi.org/10.1038/s41379-020-0458-7 (2020).

10. Yoo, S. K., Song, Y. S., Park, Y. J. & Seo, J. S. Recent Improvements in genomic and transcriptomic understanding of anaplastic and poorly differentiated thyroid cancers. *Endocrinol Metab (Seoul)* **35**, 44–54. https://doi.org/10.3803/EnM.2020.35.1.44 (2020).
11. Ibrahimpasic, T. *et al.* Genomic alterations in fatal forms of non-anaplastic thyroid cancer: identification of MED12 and RBM10 as novel thyroid cancer genes associated with tumor virulence. *Clin Cancer Res* **23**, 5970–5980. https://doi.org/10.1158/1078-0432.CCR-17-1183 (2017).
12. Landa, I. *et al.* Genomic and transcriptomic hallmarks of poorly differentiated and anaplastic thyroid cancers. *J Clin Invest* **126**, 1052–1066. https://doi.org/10.1172/JCI85271 (2016).
13. Gerber, T. S. *et al.* Targeted next-generation sequencing of cancer genes in poorly differentiated thyroid cancer. *Endocr Connect* **7**, 47–55. https://doi.org/10.1530/EC-17-0290 (2018).
14. Kim, J. H. *et al.* Genomic profiling of aggressive thyroid cancer in association with its clinicopathological characteristics. *In Vivo* **36**, 111–120. https://doi.org/10.21873/invivo.12682 (2022).
15. Paulsson, J. O. *et al.* Whole-genome sequencing of synchronous thyroid carcinomas identifies aberrant DNA repair in thyroid cancer dedifferentiation. *J Pathol* **250**, 183–194. https://doi.org/10.1002/path.5359 (2020).
16. Yoo, S. K. *et al.* Integrative analysis of genomic and transcriptomic characteristics associated with progression of aggressive thyroid cancer. *Nat Commun* **10**, 2764. https://doi.org/10.1038/s41467-019-10680-5 (2019).
17. Yoo, S. K. *et al.* Comprehensive analysis of the transcriptional and mutational landscape of follicular and papillary thyroid cancers. *PLoS Genet* **12**, e1006239. https://doi.org/10.1371/journal.pgen.1006239 (2016).
18. Lee, J. *et al.* Mutalisk: a web-based somatic MUTation AnaLyIS toolKit for genomic, transcriptional and epigenomic signatures. *Nucleic Acids Res* **46**, W102–W108. https://doi.org/10.1093/nar/gky406 (2018).
19. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993. https://doi.org/10.1016/j.cell.2012.04.024 (2012).
20. Zamborszky, J. *et al.* Loss of BRCA1 or BRCA2 markedly increases the rate of base substitution mutagenesis and has distinct effects on genomic deletions. *Oncogene* **36**, 746–755. https://doi.org/10.1038/onc.2016.243 (2017).
21. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421. https://doi.org/10.1038/nature12477 (2013).
22. Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101. https://doi.org/10.1038/s41586-020-1943-3 (2020).
23. Iorio, F. *et al.* A landscape of pharmacogenomic interactions in cancer. *Cell* **166**, 740–754. https://doi.org/10.1016/j.cell.2016.06.017 (2016).
24. Prete, A. *et al.* Update on Fundamental mechanisms of thyroid cancer. *Front Endocrinol (Lausanne)* **11**, 102. https://doi.org/10.3389/fendo.2020.00102 (2020).
25. Howell, G. M., Hodak, S. P. & Yip, L. RAS mutations in thyroid cancer. *Oncologist* **18**, 926–932. https://doi.org/10.1634/theoncologist.2013-0072 (2013).
26. Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* **47**, W191–W198. https://doi.org/10.1093/nar/gkz369 (2019).
27. Klempner, S. J. *et al.* Tumor mutational burden as a predictive biomarker for response to immune checkpoint inhibitors: a review of current evidence. *Oncologist* **25**, e147–e159. https://doi.org/10.1634/theoncologist.2019-0244 (2020).
28. Jiang, P. *et al.* Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med* **24**, 1550–1558. https://doi.org/10.1038/s41591-018-0136-1 (2018).
29. Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* **348**, 69–74. https://doi.org/10.1126/science.aaa4971 (2015).
30. Wang, P., Chen, Y. & Wang, C. Beyond tumor mutation burden: tumor neoantigen burden as a biomarker for immunotherapy and other types of therapy. *Front Oncol* **11**, 672677. https://doi.org/10.3389/fonc.2021.672677 (2021).
31. Guo, M. *et al.* tumor mutation burden predicts relapse in papillary thyroid carcinoma with changes in genes and immune microenvironment. *Front. Endocrinol. (Lausanne)* **12**, 674616 https://doi.org/10.3389/fendo.2021.674616 (2021).
32. Hundal, J. *et al.* pVAC-Seq: A genome-guided in silico approach to identifying tumor neoantigens. *Genome Med* **8**, 11. https://doi.org/10.1186/s13073-016-0264-5 (2016).
33. Zaballos, M. A. & Santisteban, P. Key signaling pathways in thyroid cancer. *J Endocrinol* **235**, R43–R61. https://doi.org/10.1530/JOE-17-0266 (2017).
34. Nozhat, Z. & Hedayati, M. PI3K/AKT pathway and its mediators in thyroid carcinomas. *Mol Diagn Ther* **20**, 13–26. https://doi.org/10.1007/s40291-015-0175-y (2016).
35. Nikolaev, S. I. *et al.* Exome sequencing identifies recurrent somatic MAP2K1 and MAP2K2 mutations in melanoma. *Nat Genet* **44**, 133–139. https://doi.org/10.1038/ng.1026 (2011).
36. Ren, L. X. *et al.* Myc-associated zinc-finger protein promotes clear cell renal cell carcinoma progression through transcriptional activation of the MAP2K2-dependent ERK pathway. *Cancer Cell Int* **21**, 323. https://doi.org/10.1186/s12935-021-02020-9 (2021).
37. Aktar, N. *et al.* Understanding of immune escape mechanisms and advances in cancer immunotherapy. *J Oncol* **2022**, 8901326. https://doi.org/10.1155/2022/8901326 (2022).
38. Ling, A. *et al.* TAP1 down-regulation elicits immune escape and poor prognosis in colorectal cancer. *Oncoimmunology* **6**, e1356143. https://doi.org/10.1080/2162402X.2017.1356143 (2017).
39. DeYoung, K. L. *et al.* Cloning a novel member of the human interferon-inducible gene family associated with control of tumorigenicity in a model of human melanoma. *Oncogene* **15**, 453–457. https://doi.org/10.1038/sj.onc.1201206 (1997).
40. Kopp, R., Glas, J., Lau-Werner, U., Albert, E. D. & Weiss, E. H. Association of MICA-TM and MICB C1_2_A microsatellite polymorphisms with tumor progression in patients with colorectal cancer. *J Clin Immunol* **29**, 545–554. https://doi.org/10.1007/s10875-009-9288-6 (2009).
41. Raffaghello, L. *et al.* Downregulation and/or release of NKG2D ligands as immune evasion strategy of human neuroblastoma. *Neoplasia* **6**, 558–568. https://doi.org/10.1593/neo.04316 (2004).
42. Feng, Q. *et al.* High MICB expression as a biomarker for good prognosis of colorectal cancer. *J Cancer Res Clin Oncol* **146**, 1405–1413. https://doi.org/10.1007/s00432-020-03159-0 (2020).
43. Zheng, M. Tumor mutation burden for predicting immune checkpoint blockade response: the more, the better. *J Immunother Cancer* **10**. https://doi.org/10.1136/jitc-2021-003087 (2022).
44. Garcia-Alvarez, A., Hernando, J., Carmona-Alonso, A. & Capdevila, J. What is the status of immunotherapy in thyroid neoplasms? *Front. Endocrinol. (Lausanne)* **13**, 929091. https://doi.org/10.3389/fendo.2022.929091 (2022).
45. Kunstman, J. W. *et al.* Characterization of the mutational landscape of anaplastic thyroid cancer via whole-exome sequencing. *Hum Mol Genet* **24**, 2318–2329. https://doi.org/10.1093/hmg/ddu749 (2015).
46. Mittal, P. & Roberts, C. W. M. The SWI/SNF complex in cancer - biology, biomarkers and therapy. *Nat Rev Clin Oncol* **17**, 435–448. https://doi.org/10.1038/s41571-020-0357-3 (2020).
47. Pan, D. *et al.* A major chromatin regulator determines resistance of tumor cells to T cell-mediated killing. *Science* **359**, 770–775. https://doi.org/10.1126/science.aao1710 (2018).
48. Ghossein, C. A., Dogan, S., Farhat, N., Landa, I. & Xu, B. Expanding the spectrum of thyroid carcinoma with somatic DICER1 mutation: a survey of 829 thyroid carcinomas using MSK-IMPACT next-generation sequencing platform. *Virchows Arch* **480**, 293–302. https://doi.org/10.1007/s00428-021-03212-4 (2022).

49. Macchia, P. E. *et al.* PAX8 mutations associated with congenital hypothyroidism caused by thyroid dysgenesis. *Nat Genet* **19**, 83–86. https://doi.org/10.1038/ng0598-83 (1998).
50. Sansal, I. & Sellers, W. R. The biology and clinical relevance of the PTEN tumor suppressor pathway. *J Clin Oncol* **22**, 2954–2963. https://doi.org/10.1200/JCO.2004.02.141 (2004).
51. Collini, P., Sampietro, G. & Pilotti, S. Extensive vascular invasion is a marker of risk of relapse in encapsulated non-Hurthle cell follicular carcinoma of the thyroid gland: a clinicopathological study of 18 consecutive cases from a single institution with a 11-year median follow-up. *Histopathology* **44**, 35–39. https://doi.org/10.1111/j.1365-2559.2004.01729.x (2004).
52. Zheng, S. *et al.* ABCA12 promotes proliferation and migration and inhibits apoptosis of pancreatic cancer cells through the AKT signaling pathway. *Front Genet* **13**, 906326. https://doi.org/10.3389/fgene.2022.906326 (2022).
53. Sekhar, V., Andl, T. & Phanstiel, O. T. ATP13A3 facilitates polyamine transport in human pancreatic cancer cells. *Sci. Rep.* **12**, 4045. https://doi.org/10.1038/s41598-022-07712-4 (2022).
54. Izumi, H. *et al.* The CLIP1-LTK fusion is an oncogenic driver in non-small-cell lung cancer. *Nature* **600**, 319–323. https://doi.org/10.1038/s41586-021-04135-5 (2021).
55. Zaman, N. *et al.* Signaling network assessment of mutations and copy number variations predict breast cancer subtype-specific drug targets. *Cell Rep* **5**, 216–223. https://doi.org/10.1016/j.celrep.2013.08.028 (2013).
56. Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr Protoc Bioinform.* **43**, 11–10. https://doi.org/10.1002/0471250953.bi1110s43 (2013).
57. Koboldt, D. C. *et al.* VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* **22**, 568–576. https://doi.org/10.1101/gr.129684.111 (2012).
58. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907–915. https://doi.org/10.1038/s41587-019-0201-4 (2019).
59. Cerami, E. *et al.* The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401–404. https://doi.org/10.1158/2159-8290.CD-12-0095 (2012).
60. Tate, J. G. *et al.* COSMIC: The catalogue of somatic mutations in cancer. *Nucleic Acids Res* **47**, D941–D947. https://doi.org/10.1093/nar/gky1015 (2019).
61. Shukla, S. A. *et al.* Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat Biotechnol* **33**, 1152–1158. https://doi.org/10.1038/nbt.3344 (2015).
62. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102**, 15545–15550. https://doi.org/10.1073/pnas.0506580102 (2005).
63. Fu, J. *et al.* Large-scale public data reuse to model immunotherapy response and resistance. *Genome Med* **12**, 21. https://doi.org/10.1186/s13073-020-0721-z (2020).

Author contributions

Conceptualization: SN, YSC; Methodology: SN; Sample collection: J-HL, YSC; Data curation: YL, SM; Formal Analysis: YL, SM, JYS; Funding acquisition: SN, YSC; Investigation: YL, SM; Supervision: SN, YSC; Validation: YL, SM, J-HL; Visualization: YL, SM; Writing-original draft: YL, SM; Writing – review & editing: SN, YSC. All authors have read and agreed to the published version of the manuscript. All participants provided written informed consent.

Funding

This work was supported by the Industrial Technology Innovation program (NO. 20016417, AI prediction platform development for lung and gastric cancer with Korean genetic data and its servitization) funded by the Ministry of Trade, Industry & Energy (MOTIE (KATS)/KEIT, Korea); the Gachon University research fund of 2020 (GCU-202008430003 to SN); and the Gachon University Gil Medical Center (Grant number: FRD2020-09 to YSC).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-46466-5.

Correspondence and requests for materials should be addressed to S.N. or Y.S.C.

Reprints and permissions information is available at www.nature.com/reprints.

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

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

© The Author(s) 2023