

Integrated analysis of fine-needle-aspiration cystic fluid proteome, cancer cell secretome, and public transcriptome datasets for papillary thyroid cancer biomarker discovery

Chia-Chun Wu^{1,*}, Jen-Der Lin^{4,*}, Jeng-Ting Chen⁷, Chih-Min Chang¹, Hsiao-Fen Weng⁴, Chuen Hsueh⁵, Hui-Ping Chien⁵ and Jau-Song Yu^{1,2,3,6}

¹Graduate Institute of Biomedical Sciences, Chang Gung University, Taoyuan, Taiwan

²Department of Cell and Molecular Biology, College of Medicine, Chang Gung University, Taoyuan, Taiwan

³Molecular Medicine Research Center, Chang Gung University, Taoyuan, Taiwan

⁴Division of Endocrinology and Metabolism, Department of Internal Medicine, Chang Gung Memorial Hospital at Linkou, Taoyuan, Taiwan

⁵Department of Pathology, Chang Gung Memorial Hospital at Linkou, Taoyuan, Taiwan

⁶Liver Research Center, Chang Gung Memorial Hospital at Linkou, Taoyuan, Taiwan

⁷Department of Surgery, Department of Medical Research and Development Linkou Branch, Chang Gung Memorial Hospital at Linkou, Taoyuan, Taiwan

*These authors contributed equally to this work

Correspondence to: Jau-Song Yu, **email:** yusong@mail.cgu.edu.tw

Keywords: papillary thyroid carcinoma; fine needle aspiration cystic fluid; proteome profiles; secretome; biomarker

Received: April 20, 2017

Accepted: November 15, 2017

Published: January 04, 2018

Copyright: Wu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Thyroid ultrasound and ultrasound-guided fine-needle aspiration (USG/FNA) biopsies are currently used for diagnosing papillary thyroid carcinoma (PTC), but their detection limit could be improved by combining other biomarkers. To discover novel PTC biomarkers, we herein applied a GeLC-MS/MS strategy to analyze the proteome profiles of serum-abundant-protein-depleted FNA cystic fluid from benign and PTC patients, as well as two PTC cell line secretomes. From them, we identified 346, 488, and 2105 proteins, respectively. Comparative analysis revealed that 191 proteins were detected in the PTC but not the benign cystic fluid samples, and thus may represent potential PTC biomarkers. Among these proteins, 101 were detected in the PTC cell line secretomes, and seven of them (NPC2, CTSC, AGRN, GPNMB, DPP4, ERAP2, and SH3BGRL3) were reported in public PTC transcriptome datasets as having 4681 elevated mRNA expression in PTC. Immunoblot analysis confirmed the elevated expression levels of five proteins (NPC2, CTSC, GPNMB, DPP4, and ERAP2) in PTC versus benign cystic fluids. Immunohistochemical studies from near 100 pairs of PTC tissue and their adjacent non-tumor counterparts further showed that AGRN ($n = 98$), CTSC ($n = 99$), ERAP2 ($n = 98$) and GPNMB ($n = 100$) were significantly ($p < 0.05$) overexpressed in PTC and higher expression levels of AGRN and CTSC were also significantly associated with metastasis and poor prognosis of PTC patients. Collectively, our results indicate that an integrated analysis of FNA cystic fluid proteome, cancer cell secretome and tissue transcriptome datasets represents a useful strategy for efficiently discovering novel PTC biomarker candidates.

INTRODUCTION

Thyroid cancer is the most prevalent malignant endocrine carcinoma in the world. In America, the

incidence rate of thyroid cancer increased an average of 4.6% per year between 2004 and 2013, and the disease is now ranked as the third most common cancer in women [1]. Among all of the histological types of thyroid cancer,

papillary thyroid carcinoma (PTC) accounts for the majority (80–85%) of cases. Most PTC patients have a good prognosis with a 10-year survival rate > 90%; however, a small portion of PTCs are aggressive and may develop distant metastases that are associated with higher mortality [2]. Currently, preoperative ultrasound-guided fine-needle aspiration (USG/FNA) followed by cytopathologic diagnosis is the standard procedure for examining thyroid nodules and determining therapeutic modalities [3]. Among all cases, however, 10–20% yield indeterminate biopsy results on preoperative USG/FNA diagnosis. These patients normally undergo thyroidectomy, but this is often unnecessary, as post-surgery immunohistological analysis has shown that 80% of the suspicious cases are benign [4, 5]. Thus, if we hope to reduce unnecessary thyroidectomies and prevent deterioration in PTC, we urgently need reliable biomarkers that can distinguish between benign and malignant nodules in patients with indeterminate lesions.

Cystic fluid and tissue/cell specimens obtained from thyroid nodules during the FNA procedure represent ideal resources for discovering and verifying PTC biomarkers. Various types of potential PTC biomarkers have been explored in FNA cystic fluid and tissue/cell specimens, including DNA mutations/rearrangements and proteins. Several gene mutations have been identified in thyroid cancer. For example, somatic *RET/PTC* gene rearrangements and the *BRAF*^{V600E} point mutation are frequently found in PTC, and clinicians have used them to decide whether or not to undertake thyroidectomy [2, 6]. However, FNA cystic fluid samples typically contain very few cancer cells, limiting their usefulness for cytologic evaluation or genetic marker screening, and creating ambiguity in the diagnosis of thyroid cancer [7].

To address this issue, several groups have applied proteomic approaches to discover potential PTC biomarkers from FNA cystic fluid samples, PTC cell secretomes, or tissue specimens. For example, two studies identified secreted proteins (secretome analysis; 154 and 83 proteins, respectively) from PTC cell lines using LC-MS/MS analysis [8, 9]. Two other studies investigated the differential protein profiles between benign and malignant cystic fluids using two-dimensional gel electrophoresis/MALDI-TOF-MS or iTRAQ-based LC-MS/MS. These studies then further verified two and three proteins, respectively, by Western blotting, ELISA and/or immunohistochemical analysis of PTC cystic fluids and tissue samples [10, 11]. More recently, Martínez-Aguilar et al. used MS-based quantitative expression analysis of over 1600 proteins in normal thyroid tissues versus PTC to identify ~180 proteins that are deregulated in PTC tumors [12]. Although these proteomic studies have discovered numerous proteins as potential PTC biomarkers, however, challenges remain in determining how researchers should efficiently prioritize and select targets for further validation in a large PTC sample cohort.

In this study, we used an integrated omics approach to identify potential PTC biomarkers. We applied a GeLC-MS/MS strategy to comprehensively analyze the proteome profiles of abundant-protein-depleted FNA cystic fluids (i.e., those depleted of the top 14 most abundant proteins) from benign and PTC patients, as well as secretomes from the PTC cell lines, BHP 7-13 and CGTH W3. Using a criterion of at least two peptide hits for confident protein identification, we identified 346 and 488 proteins expressed in the FNA cystic fluids of benign and PTC patients, and 2105 proteins in the secretomes of the two PTC cell lines. Integrated analysis of these three datasets revealed that 101 proteins found in the PTC cystic fluid but not the corresponding benign samples were also detected in the PTC cell line secretome. We then combined two publicly available mRNA microarray datasets representing PTC tissues, and used them to identify the seven strongest candidates: *NPC2*, *CTSC*, *AGRN*, *GPNMB*, *DPP4*, *ERAP2*, and *SH3BGRL3*. Finally, we used Western blot analysis and/or immunohistochemistry to confirm the up-regulations of selected candidates in PTC specimens, and evaluated their associations with the clinicopathological characteristics of PTC patients.

RESULTS

Generating a thyroid cystic fluid proteome dataset using GeLC-MS/MS

The strategy we used to improve the identification of novel PTC biomarkers is delineated in Figure 1. We used USG/FNA to collect thyroid cystic fluid samples from 12 patients: five cases of PTC and seven cases of benign disease (Supplementary Table 1). The individual cystic fluid samples from seven benign and three PTC specimens were pooled into two groups containing equal amounts of protein. These samples were subjected to top-14-high-abundance-protein depletion followed by GeLC-MS/MS-based proteomic profiling. The protein staining patterns of the abundant-protein-depleted fractions are shown in Figure 2A, and were quite different from those of the un-depleted (crude) and abundant protein fractions (Supplementary Figure 1). Our results demonstrated that the highly abundant proteins accounted for >90% of the total protein mass in the thyroid cystic fluid, as previously seen in human serum [13, 14]. The gel lane of each sample was sliced into 60 fractions, and the proteins were subjected to in-gel tryptic digestion and identified by LC-MS/MS analysis using an LTQ-Orbitrap. Supplementary Table 2A lists the MS-identified proteins in the abundant protein fractions of benign cystic fluid, as captured by a Human 14 MARS column. To examine if the combined use of these technology platforms (i.e., abundant protein depletion and GeLC-MS/MS) can significantly improve the detection rate of low-abundance proteins in the thyroid cystic fluid samples, we used the same GeLC-

MS/MS approach to analyze respectively the undepleted and abundant protein-depleted cystic fluid sample pooled from three PTC specimens. This analysis identified 241 proteins in the undepleted sample (Supplementary Table 2B), which is much less than the number of proteins (488) identified in the depleted sample (Supplementary Table 3B). Regarding the total identified spectra, there was a high proportion (45%, 24689/54411) of abundant plasma proteins in the undepleted sample as compared to a significantly lower proportion (19%, 22474/115905) of these abundant plasma proteins in the depleted sample (Supplementary Table 2C). After removing the abundant plasma protein identities from these two datasets, 194 proteins and 456 proteins remained respectively in the undepleted and depleted sample; importantly, we found that 284 proteins could only be identified in the depleted sample. Collectively, these findings confirmed that abundant plasma protein depletion is very useful to significantly increase the detection rate of low-abundance proteins in thyroid cystic fluid.

From the benign and PTC abundant-protein-depleted samples (60 µg each), the GeLC-MS/MS approach identified 346 and 488 proteins, respectively, with multiple (≥ 2) peptide hits and a false discovery rate (FDR) of 0.58–1.23% (Table 1 and Supplementary Table 3A–3B). Recent studies have reported galectin-3 (LGALS3) as a reliable immunohistochemical and serum biomarker for PTC detection [15, 16]. Thus, it is notable that we identified galectin-3 in our FNA cystic fluid dataset, and our label-free spectral counting approach found that its expression level was much higher (~7.8 fold) in our PTC samples compared to our benign cystic fluid samples (Supplementary Table 4).

GeLC-MS/MS-based secretomic analysis of two PTC cell lines

Serum-free conditioned media of two PTC cell lines (CGTH W3 and BHP 7-13) were concentrated and desalted, and 50 µg of proteins from each sample were resolved by 8–14% gradient SDS-PAGE (Figure 1). The protein staining patterns of the conditioned media from CGTH W3 and BHP 7-13 cells are shown in Figure 2B. As a quality control, we used Western blotting analysis to examine the distribution of α -tubulin, a cytoplasmic abundant protein, between total cell extracts and the conditioned media. We clearly detected α -tubulin in the total cell extracts but obtained little or no such signal from the conditioned media (Figure 2C), confirming that the secreted/shed proteins had been specifically collected from the cultured PTC cell lines. The gel lane of each sample was sliced into 40 fractions, and the proteins were identified by the above-described GeLC-MS/MS approach. Starting from 50 µg of proteins from the conditioned media, we identified 1921 and 830 proteins in the CGTH W3 and BHP 7-13 cell lines, respectively, with

a FDR of 0.12–0.21% (Table 1 and Supplementary Table 3C–D). Our analysis thus yielded a total of 2105 proteins that served as our dataset for PTC biomarker selection (Supplementary Table 5A).

Integrated analysis and secretion-pathway prediction of the MS-identified proteins in the thyroid cystic fluid and PTC cell secretome

A total of 537 unique proteins were detected in the benign and PTC cystic fluid proteomes; of them, 297 were shared between the datasets (Figure 3A and Supplementary Table 4). Using a label-free quantification strategy and spectral counting, we determined that 86 of the shared proteins (29%) were up-regulated and 36 (12%, 36 of 297) were down-regulated more than two fold in the PTC versus benign samples (Supplementary Table 4). The up-regulated proteins included some potential biomarkers previously reported as having elevated levels in PTC [10, 15–28], such as FTL, FTH1, LGALS3, JUP, S100A11, SERPINA1, TF, ECM1, CD163, BCHE, SOD2, CP, FN1, S100A4, and C5 (Supplementary Table 4). This confirms that our label-free quantification strategy can be used to discover PTC biomarkers from proteomic datasets generated by abundant protein depletion of cystic fluid samples coupled with GeLC-MS/MS.

The identified proteins were further analyzed using bioinformatics programs designed to predict protein secretion pathways. Among the 537 non-redundant proteins identified in our thyroid cystic fluid samples, 246 proteins were predicted to be classical secreted proteins, as assessed by the SignalP 4.0 program (SignalP-TM probability ≥ 0.50 , SignalP-noTM ≥ 0.45) based on the presence of a signal peptide in target proteins with or without transmembrane (TM) sequences [29]. The SecretomeP 2.0 program predicted that 131 proteins would be released through the non-classical pathway (SignalP score < 0.5 or 0.45 and Secretome score ≥ 0.50) [30]. Among them, only five proteins were determined to be integral membrane proteins, as assessed by TMHMM [31]. Among the 2105 proteins of the PTC cell secretome dataset, 426 were predicted to be classical secreted proteins, 723 were predicted to be non-classical secreted proteins, and 81 were predicted to be integral membrane proteins (Table 2). Taken together, our results indicated that 71.1% (382 of 537) of the cystic fluid proteins and 58.4% (1230 of 2105) of the conditioned-media proteins of cultured PTC cells could be secreted/released via different mechanisms. Notably, the cystic fluid proteins of the up-regulated group described above included a higher percentage (84.9%, 73 of 86) of secretory proteins than those of the down-regulated group (63.9%, 23 of 36) (Supplementary Tables 4 and 6). Moreover, among the 537 cystic fluid proteins, the 191 PTC-unique proteins included a relatively high percentage (74.3%, 142 of 191) of secretory proteins (Table 2, Supplementary Table 4

Table 1: Number of proteins identified in thyroid cystic fluids and PTC cell secretomes

Thyroid cystic fluid or PTC cell lines	No. of identified proteins ^a	FDR ^b
Benign cystic fluid	346	0.58%
PTC cystic fluid	488	1.23%
BHP 7-13	830	0.12%
CGTH W-3	1921	0.21%

^aThe numbers represent identified proteins for which at least two unique peptides of 95% peptide possibility and 95% protein possibility were identified using the Scaffold software.

^bThe false discovery rate (FDR) was calculated as the ratio of the spectra assigned to a random database over those assigned to a normal database.

and Supplementary Table 6) compared to those uniquely identified in the benign cystic fluid sample (65.3%, 32 of 49) (Supplementary Tables 4 and 6). These results indicate that most PTC-related proteins are predicted to be secretory, and thus PTC biomarkers could be exploited in cystic fluid. In addition, we used the Gene Ontology (GO) annotation tool in the DAVID Bioinformatics Resources (v6.8) to perform functional annotation of those 277 proteins with two-fold up-regulation or solely detected in PTC cystic fluids, including biological process, molecular function and proteins location of cellular components. Regarding the biological process, the top three categories were platelet degranulation, complement activation (classical pathway) and complement activation (Supplementary Figure 2A). The major molecular functions were classified as serine-type endopeptidase activity, structural molecule activity and heparin binding (Supplementary Figure 2B). The protein localization was mainly classified into extracellular such as extracellular exosome, extracellular space and extracellular region (Supplementary Figure 2C).

Selecting novel PTC cystic fluid biomarker candidates through a combined analysis of the PTC cystic fluid proteome, cell secretome, and tissue transcriptome

To search for PTC-specific biomarkers, the 191 proteins (Figure 3A) uniquely detected in PTC cystic fluid samples were compared with the 2105 proteins (Figure 3B) identified in the secretomes of the BHP 7-13 and CGTH W3 cell lines. We also mined two public cDNA microarray datasets (the E-GEOD-3678 dataset from EBI-ArrayExpress and the GDS1665 dataset from Gene Expression Omnibus) for a total of 4681 genes whose tissue mRNA expression levels were reported to be significantly elevated in PTC versus adjacent normal tissues (Figure 3C and Supplementary Table 7) [32]. Our combined analysis of the three datasets (Figure 3D) identified seven highly relevant potential PTC candidate biomarkers: epididymal secretory protein E1 (NPC2), dipeptidyl peptidase 1 (CTSC), agrin

(AGRN), transmembrane glycoprotein NMB (GPNMB), dipeptidyl peptidase 4 (DPP4), endoplasmic reticulum aminopeptidase 2 (ERAP2), and SH3 domain-binding glutamic acid-rich-like protein 3 (SH3BGRL3) (Table 3). All seven are predicted to be secreted via the classical or nonclassical secretion pathways, and four of them (CTSC, AGRN, DPP4, and SH3BGRL3) are known as exosomal proteins [33]. Of them, DPP4 was previously detected in the human serum proteome dataset [34] (Table 3), and was verified as a potential PTC biomarker using tissue specimens [35]. Otherwise, ERAP2 was also elucidated highly expressed in PTC tissue with cervical lymph node metastasis [36], but other five candidates had not previously been reported.

Verifying potential PTC biomarkers in the conditioned media of PTC cell lines and thyroid cystic fluids

We used Western blot analysis to verify the seven selected biomarkers in conditioned media of the CGTH W3 and BHP 7-13 cell lines, as well as in cell extracts. Galectin-3, which was previously reported as a promising PTC biomarker [37], was used as a positive control. As shown in Figure 4, all of the tested proteins except SH3BGRL3 could be detected in the conditioned media of CGTH W3 and/or BHP 7-13 cells, and their relative expression levels fitted well with the total spectral numbers detected in the PTC cell secretome (Supplementary Table 5B). Similar experiments were performed to evaluate the levels of these potential biomarkers in twelve abundant-protein-depleted cystic fluid samples including ten samples used for initial discovery experiment and two subsequently collected PTC cystic fluid samples. As shown in Figure 5A, five candidates (GPNMB, DPP4, ERAP2, CTSC, and NPC2) were successfully confirmed. Their levels were generally higher in PTC cystic fluid samples compared to benign cystic fluid samples when equal amounts of total protein were analyzed (Figure 5B). We examined SFTP as an additional candidate, as it was detected solely in our PTC cystic fluid data set with a very

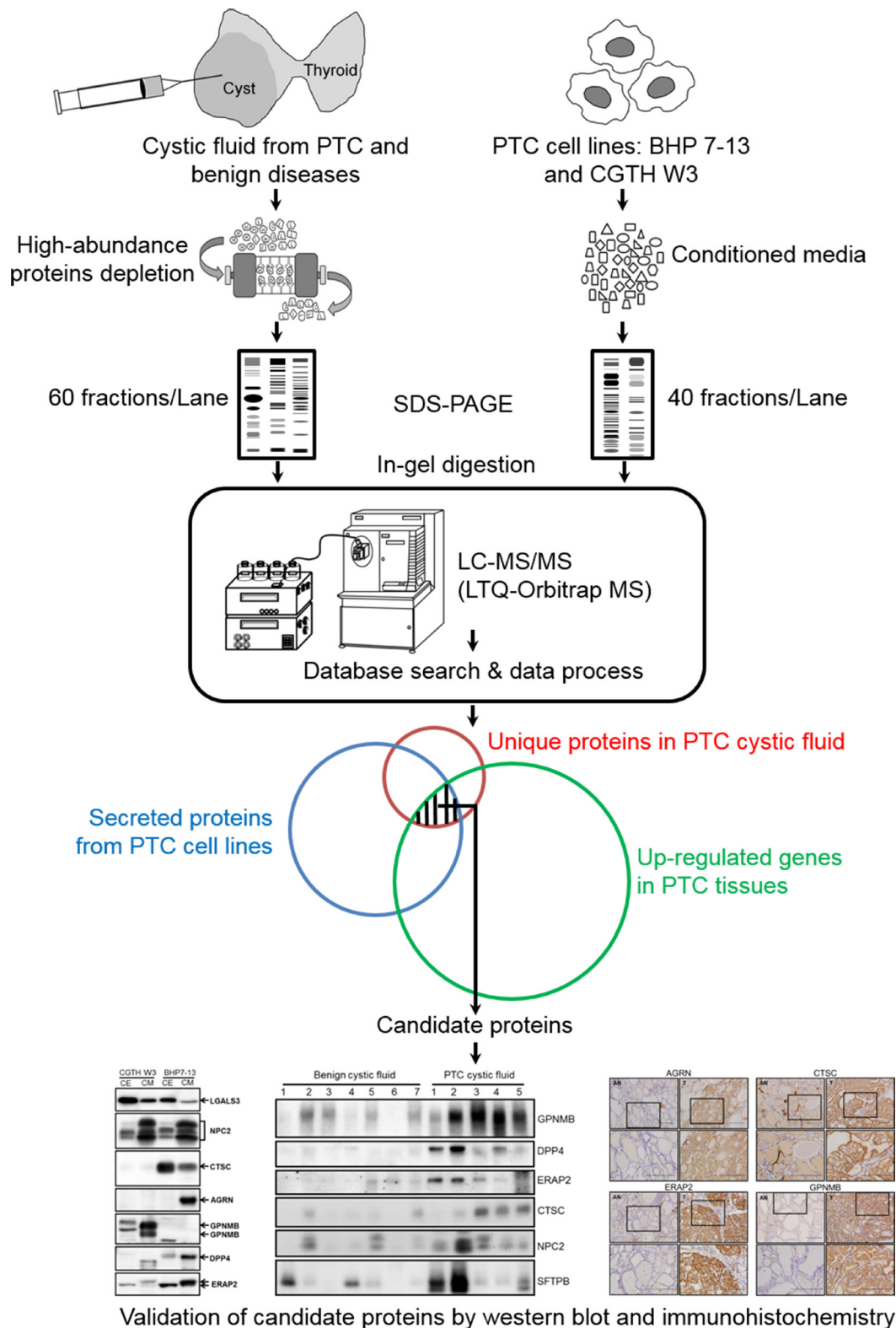


Table 2: Predicted secretion pathways of proteins identified in thyroid cystic fluids and the two PTC cell secretomes

Thyroid cystic fluid and PTC cell lines	No. of identified proteins				Percentage of predicted secreted proteins (%)
	Classical secretion ^b	Nonclassical secretion ^c	Membrane protein ^d	Others ^e	
Benign cystic fluid	163	71	5	107	69.1
PTC cystic fluid	232	115	3	138	71.7
Benign U PTC cystic fluid	246	131	5	155	71.1
PTC cystic fluid alone ^a	83	59	0	49	74.3
CGTH W3	321	683	73	844	56.1
BHP 7-13	290	209	22	309	62.8
CGTH W3 U BHP 7-13	426	723	81	875	58.4

^aA total of 191 proteins were identified as being unique to PTC from the thyroid cystic fluid dataset.

^bProteins predicted by the SignalP program to be secreted via the classical secretion pathway (SignalP probability ≥ 0.45 without TM; ≥ 0.50 with TM).

^cProteins predicted to be secreted by the nonclassical secretion pathway, as assessed using SignalP and SecretomeP (SignalP probability lower than the above criteria and SecretomeP score ≥ 0.50).

^dProteins predicted by the TMHMM to form integral membrane proteins that were not predicted to be secreted via the classical or nonclassical secretion pathways.

^eProteins that could not be classified as classical secreted, nonclassical secreted, or integral membrane proteins.

high spectral count (Supplementary Table 4), and was previously reported to be upregulated in PTC tissues [38]. Consistent with these observations, we found that SFTPB was drastically increased in two of the PTC cystic fluid samples (Figure 5A). Taken together, these results indicate that the proteins selected based on our integrated analysis of the FNA cystic fluid proteome, PTC cell secretome, and tissue transcriptome represent strong potential biomarkers for detecting PTC from FNA cystic fluid.

Overexpression of AGRN, CTSC, ERAP2, and GPNMB in PTC tissues

To further examine the expression levels of the seven prioritized targets in PTC tissues, we surveyed antibodies suitable for their immunohistochemical (IHC) analysis. We obtained suitable antibodies against AGRN, CTSC, ERAP2, and GPNMB and used them to stain tissue specimens from 114, 117, 115, and 115 patients, respectively (Supplementary Table 8). Among them, 98, 99, 98, and 100 specimens, respectively, contained PTC tumor and adjacent non-tumor tissues. Representative paired-tissue staining patterns for the four target proteins are shown in Figure 6A. Analysis of the staining scores revealed that strong staining (2+ and 3+) for AGRN, CTSC, ERAP2, and GPNMB was respectively detected in 20.4% (20/98), 32.3% (32/99), 83.7% (82/98), and 56% (56/100) of regions containing tumor cells, while weak staining (1+) was respectively detected in 91.8% (90/98), 94.9% (94/99), 40.8% (40/98), and 99% (99/100) of

regions harboring paired adjacent non-tumor cells (Figure 6B). Statistical analysis showed that the expression levels (IHC staining scores) of all four proteins were significantly higher in the tumor parts than in the adjacent non-tumor parts (89.13 ± 49.8 vs. 52.35 ± 30.96 , $p < 0.001$ for AGRN; 75.71 ± 54.11 vs. 35.45 ± 32.18 , $p < 0.001$ for CTSC; 203.6 ± 88.27 vs. 70.05 ± 63.89 , $p < 0.001$ for ERAP2 and 56.78 ± 55.4 vs. 3.3 ± 5.7 , $p < 0.001$ for GPNMB) (Figure 6C). It is noteworthy that little or no expression of GPNMB was detected in adjacent non-tumor tissues.

Associations of AGRN, CTSC, ERAP2, and GPNMB expression with clinicopathological characteristics

We next used the median IHC staining score value as the cutoff for each protein, and explored the relevance of the observed protein expression levels to different clinicopathological characteristics using these IHC stained specimens (114 for AGRN, 117 for CTSC, 115 for ERAP2 and 115 for GPNMB) (Table 4). Higher expression levels of AGRN and CTSC were found to be significantly correlated with lymph node metastasis, distant metastasis at diagnosis, tumor multicentricity, TNM stage, and disease-specific mortality. Notably, high CTSC expression was also significantly associated with an age greater than 45 years and with locoregional recurrence. In contrast, the expression levels of ERAP2 and GPNMB did not show any significant correlation with the tested manifestations.

Association of AGRN and CTSC expression with disease-specific survival (DSS)

To assess the correlation the expression of select proteins and patient survival, we used Kaplan-Meier plots to estimate the DSS rates of PTC patients. We found that the 5-year DSS for patients stratified by low or high protein expression were respectively 95.5% vs. 80.9% for AGRN ($p = 0.015$) and 96.4% versus 78.3% for CTSC ($p = 0.0004$), whereas there was no significant expression-related difference in DSS for ERAP2 or GPNMB (Figure 7). These results indicate that the tissue expression levels of AGRN and CTSC appear to be correlated with the DSS of PTC patients.

DISCUSSION

The integrated analysis of multiple omic datasets has shown promise for the identification of potential biomarkers and/or therapeutic targets for different cancer

types [39, 40]. Here, we set out to discover novel PTC biomarkers that are overexpressed in PTC tumors and can be secreted/released by PTC cells, and thus may be detected at elevated protein levels in PTC cystic fluid samples. Through an integrated analysis of in-house-generated FNA cystic fluid proteome and cell line secretome datasets, as well as public-domain tissue transcriptome datasets, we identified seven proteins that fit the above criteria for an ideal PTC biomarker (Figure 3). Further immunoassays confirmed that the levels of four or five of the seven target proteins appeared to be elevated in PTC cystic fluids or tumor tissues (Figures 5 and 6). Higher tumor tissue expressions of two proteins, AGRN and CTSC, were found to be significantly associated with lymph node metastasis, distant metastasis at diagnosis, and poor prognosis of PTC patients (Table 4). Our findings demonstrate that an integrated analysis of multiple omic datasets can be used to identify PTC biomarker candidates that may be of high clinical utility.

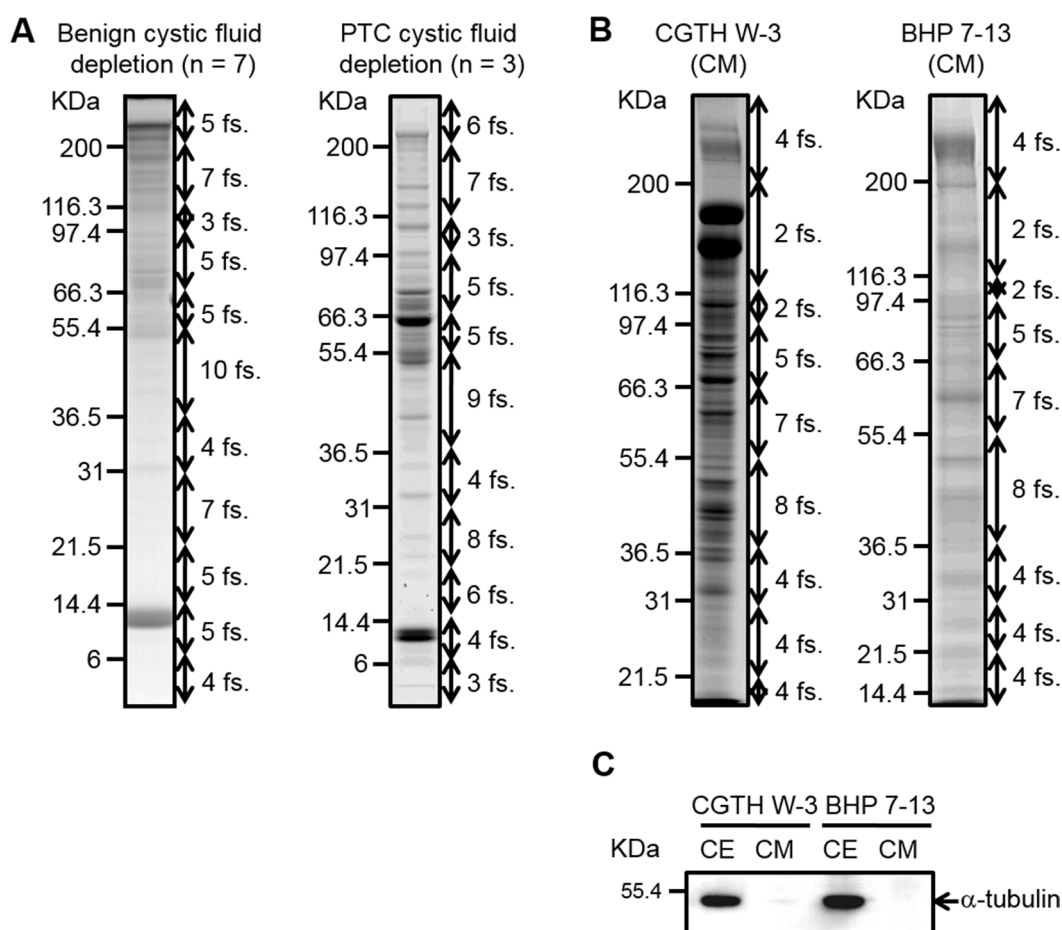


Figure 2: SDS-PAGE analysis of thyroid cystic fluid samples and conditioned media harvested from PTC cells. (A) Thyroid cystic fluid samples of seven benign cases and three PTC cases were independently pooled and depleted using Hu-14 columns. The depleted proteins (60 μ g) were resolved on 8–14%-gradient SDS gels and stained with Coomassie Blue. The gel lanes were sliced into 60 fractions (fs.) for further analysis. (B) The conditioned media (CM) of CGTH W3 and BHP 7-13 cells were collected and processed as described in the “Experimental Procedures.” Proteins (50 μ g) from the concentrated CM were resolved on 8–14%-gradient SDS gels and stained with Coomassie Blue. The gel lanes were sliced into 40 pieces for further analysis. (C) Proteins (40 μ g) from the CM and extracts of PTC cell lines (CE) were resolved by SDS-PAGE and subjected to Western blot analysis using an antibody against α -tubulin.

Table 3: List of seven high-potential candidate proteins for PTC biomarkers

Swiss-Prot		Gene name	Protein Name	Classical secretion	Nonclassical secretion	Membrane protein	HPPP ^a	ExoCarta ^b
ID_HUMAN	Code							
NPC2 ^c	P61916	<i>NPC2</i>	Epididymal secretory protein E1	V				
CATC	P53634	<i>CTSC</i>	Dipeptidyl peptidase 1	V				V
AGRIN ^c	O00468	<i>AGRN</i>	Aggrin	V				V
GPNMB	Q14956	<i>GPNMB</i>	Transmembrane glycoprotein NMB	V		V		
DPP4 ^c	P27487	<i>DPP4</i>	Dipeptidyl peptidase 4	V		V	V	V
ERAP2	Q6P179	<i>ERAP2</i>	Endoplasmic reticulum aminopeptidase 2		V			
SH3L3	Q9H299	<i>SH3BGRL3</i>	SH3 domain-binding glutamic acid-rich-like protein 3		V			V

^aProteins were detected in the Human Plasma Proteome Project (HPPP) database, which contains 3020 proteins.

^bProteins were detected in the human exosome protein database of ExoCarta.

^cThe proteins were previously reported as potential PTC biomarkers.

Several groups have previously applied proteomic approaches to discover potential PTC biomarkers [8, 10]. The study performed by Martínez-Aguilar et al. is of particular interest. The authors used SWATH-MS (sequential windowed acquisition of all theoretical fragment ion mass spectra) and MRM-HR (high-resolution multiple reaction monitoring) to analyze the proteomes in frozen thyroid tissues that included normal, follicular adenoma, follicular thyroid carcinoma, and PTC samples. They identified 1512 proteins in PTC tissues; of them, ~180 proteins were deregulated in PTC tumors compared to normal tissues [12]. When we compared the PTC tissue proteome dataset (1512 proteins) with our present cystic fluid proteome dataset (537 proteins), we found 303 proteins commonly detected in both datasets (Supplementary Table 9). Among these 303 proteins, 101 proteins (33.3%, 101/303) showed similar up- or down-regulation trend in both datasets; 73 were significantly up-regulated (≥ 2 fold) or solely detected in PTC, and 28 proteins were significantly down-regulated (≥ 2 fold) in PTC or solely detected in benign lesions. Notably, five out of the seven candidates identified in our present integrated analysis of multiple omic datasets (NPC2, CTSC, GPNMB, ERAP2, and SH3BGRL3) were among the proteins that Martínez-Aguilar et al. identified as being up-regulated (>1.5 to 4 fold increase) in PTC versus normal tissues; however, the levels of AGRN were down-regulated and DPP4 was not detected in PTC tissue specimens (Supplementary Table 9). Different approaches and study materials used in our current study and the work from Martínez-Aguilar et al. may account for the discrepancy of proteomic findings between the two datasets. Consistent with the previously reported SWATH-MS data made by Martínez-Aguilar et al., our IHC analysis of ~100 tissue specimens also revealed

that CTSC, GPNMB, and ERAP2 are overexpressed in PTC (Figure 6). Indeed, these three proteins have been consistently found to be overexpressed in PTC samples analyzed by different technologies in distinct areas of the world, including transcriptome analysis of specimens from Finland [32], SWATH-MS analysis of specimens from Australia [12], and our IHC analysis of specimens from Taiwan (this study). Our integrated analysis added further evidence that these proteins could be good PTC biomarker candidates by showing that they could be secreted/released by PTC cell lines and detected at elevated levels in FNA cystic fluids (Figures 4 and 5).

Moreover, we also included a quantitative cystic fluid proteome dataset analyzed by Dinets et al. [11] for further comparison. The authors used iTRAQ labeling coupled with LC-MS/MS approach to analyze the proteome in high abundant proteins-depleted cystic fluid samples from 7 benign and 6 PTC patients. They identified 1581 and quantified 841 proteins in cystic fluid samples. This integrated analysis (191 PTC-specific proteins from our cystic fluid proteome, 2105 proteins from our cell secretome and 1834 proteins from both tissue proteome of Martínez-Aguilar's study and cystic fluid proteome of Dinets's study) identified 82 proteins commonly detected in our and Martínez-Aguilar's (or Dinets's) datasets (Supplementary Figure 3 and Supplementary Table 10). Among them, 41 proteins with more than two-fold up-regulation (PTC/N) and 23 proteins with down-regulation in PTC lesions were observed in the tissue dataset of Martínez-Aguilar et al. study. Regarding the Dinets's dataset, 19 up-regulated and 39 down-regulated proteins could be identified, but most of them had fold-change (PTC/Benign) ≥ 2 (Supplementary Table 10). Four out of seven candidates we identified (NPC2, DPP4, SH3BGRL3 and GPNMB) were also detected in Dinets's dataset, with up-regulation of NPC2

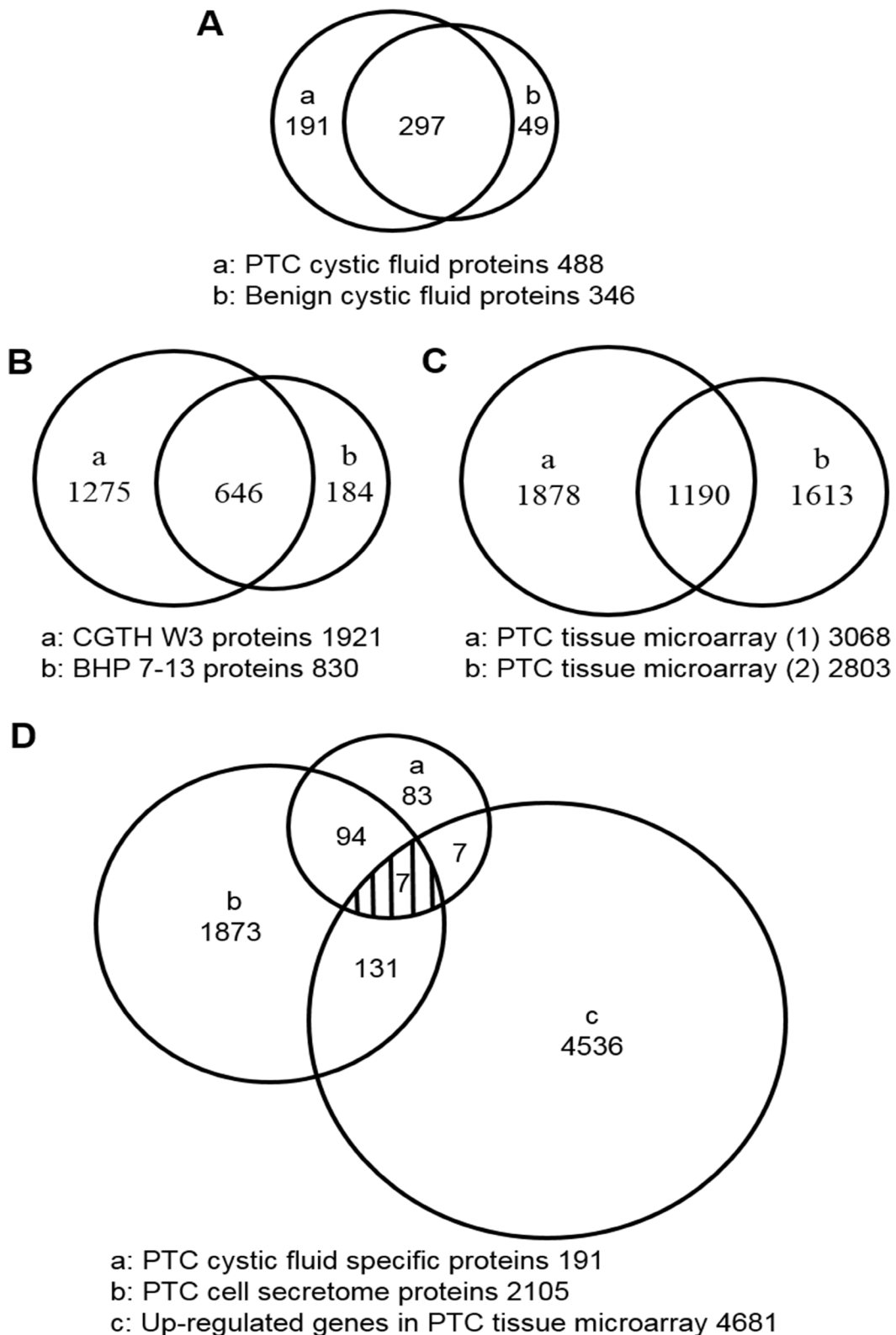


Figure 3: Venn diagrams showing the overlap of proteins in the thyroid cystic fluid, PTC cell secretome, and PTC tissue microarray datasets. (A) The number of proteins identified in thyroid cystic fluid samples from PTC patients and patients with benign lesions. (B) The numbers of proteins identified in the secretomes of the PTC cell lines, CGTH W3 and BHP 7-13. (C) The numbers of up-regulated genes in the two PTC tissue transcriptome datasets. (D) Integrated analysis of the PTC cystic fluid proteome (191 proteins), the PTC cell secretome dataset (2105 proteins), and the up-regulated genes in the two PTC tissue transcriptome datasets (4681 genes). This analysis identified seven proteins as strong PTC biomarker candidates.

Table 4: Correlations between AGRN, CTSC, ERAP2, and GPNMB expression levels and clinicopathological characteristics of PTC patients

Cut off value: Median	AGRN			CTSC			ERAP2			GPNMB		
	Low (n = 45)	High (n = 69)	p-value	Low (n = 57)	High (n = 60)	p-value	Low (n = 57)	High (n = 58)	p-value	Low (n = 50)	High (n = 65)	p-value
Gender ^a , n (%)												
Female	33 (73.3)	46 (66.7)	0.535	43 (75.4)	40 (66.7)	0.316	39 (68.4)	42 (72.4)	0.686	36 (72)	46 (70.8)	1
Male	12 (26.7)	23 (33.3)		14 (24.6)	20 (33.3)		18 (31.6)	16 (27.6)		14 (28)	19 (29.2)	
Age (years) ^a , n (%)												
<45	24 (53.3)	25 (36.2)	0.084	35 (61.4)	17 (28.3)	0.0004 ^c	30 (52.6)	21 (36.2)	0.092	25 (50)	25 (38.5)	0.257
≥45	21 (46.7)	44 (63.8)		22 (38.6)	43 (71.7)		27 (47.4)	37 (63.8)		25 (50)	40 (61.5)	
Tumor size (cm) ^a , n (%)												
<2	9 (20)	23 (33.3)	0.259	16 (28.1)	18 (30)	0.374	16 (28.1)	17 (29.3)	0.468	14 (28)	19 (29.2)	0.369
≥2 and ≤4	27 (60)	32 (46.4)		33 (57.9)	28 (46.7)		27 (47.4)	32 (55.2)		24 (48)	37 (56.9)	
>4	9 (20)	14 (20.3)		8 (14)	14 (23.3)		14 (24.6)	9 (15.5)		12 (24)	9 (13.8)	
Lymph node metastasis ^a , n (%)												
No	28 (62.2)	56 (81.2)	0.031 ^c	37 (64.9)	50 (83.3)	0.033 ^c	41 (71.9)	44 (75.9)	0.675	36 (72.0)	49 (75.4)	0.831
Yes	17 (37.8)	13 (18.8)		20 (35.1)	10 (16.7)		16 (28.1)	14 (24.1)		14 (28.0)	16 (24.6)	
Extrathyroid invasion ^a , n (%)												
No	32 (71.1)	50 (72.5)	1	42 (73.7)	41 (68.3)	0.548	43 (75.4)	38 (65.5)	0.308	41 (82)	42 (64.6)	0.058
Yes	13 (28.9)	19 (27.5)		15 (26.3)	19 (31.7)		14 (24.6)	20 (34.5)		9 (18)	23 (35.4)	
Distant metastasis at diagnosis ^a , n (%)												
No	43 (95.6)	46 (66.7)	0.0002 ^c	54 (94.7)	39 (65)	0.0001 ^c	45 (78.9)	47 (81)	0.819	35 (70)	55 (84.6)	0.071
Yes	2 (4.4)	23 (33.3)		3 (5.3)	21 (35)		12 (21.1)	11 (19)		15 (30)	10 (15.4)	
Tumor multicentricity ^a , n (%)												
No	36 (80)	41 (59.4)	0.025 ^c	44 (77.2)	35 (58.3)	0.032 ^c	42 (73.7)	37 (63.8)	0.316	37 (74)	41 (63.1)	0.234
Yes	9 (20)	28 (40.6)		13 (22.8)	25 (41.7)		15 (26.3)	21 (36.2)		13 (26)	24 (36.9)	
TNM stage ^a , n (%)												
Stage I-II	29 (64.4)	29 (42)	0.022 ^c	40 (70.2)	21 (35)	0.0002 ^c	34 (59.6)	26 (44.8)	0.137	28 (56)	31 (47.7)	0.453
Stage III-IV	16 (35.6)	40 (58)		17 (29.8)	39 (65)		23 (40.4)	32 (55.2)		22 (44)	34 (52.3)	
Postoperative 131I cumulative dose (mCi) ^b	232.7 ± 246.5 (1023, 30)	368.4 ± 480.7 (2000, 0)	0.5853	238.9 ± 228.2 (930.6, 30)	389.1 ± 505 (2000, 0)	0.9152	285.5 ± 413.7 (2000, 0)	340.2 ± 381.4 (1370, 0)	0.3215	328.1 ± 483.8 (2000, 0)	304.7 ± 341.4 (1640, 30)	0.2768
Locoregional recurrence ^a , n (%)												
No	31 (68.9)	34 (49.3)	0.053	40 (70.2)	27 (45)	0.009 ^c	31 (54.4)	35 (60.3)	0.574	25 (50)	41 (63.1)	0.186
Yes	14 (31.1)	35 (50.7)		17 (29.8)	33 (55)		26 (45.6)	23 (39.7)		25 (50)	24 (36.9)	
Disease-specific mortality (%) ^a												
	4 (9.1)	18 (26.5)	0.029 ^c	4 (7.3)	19 (31.7)	0.001 ^c	11 (19.6)	11 (19.3)	1	11 (22.4)	11 (17.2)	0.632

Low, IHC score below to the median. High, IHC score equal to or above the median. The p-values were generated using ^aChi-square test, number (%) or ^bMann-Whitney test, mean ± SD (maximum, minimum). ^cStatistically significant, p-value ≤ 0.05.

(1.7 fold) and DPP4 (1.5 fold), down-regulation of GPNMB (0.35 fold) and no obvious change of SH3BGL3 (1.07 fold) in PTC cystic fluid. As expected, both similarity and differences in the proteomic findings could be found in our

and other datasets. Nevertheless, the strategy to integrate different PTC-related tissue and cystic fluid proteome datasets represents a useful mean to prioritize more potential PTC markers for follow-up.

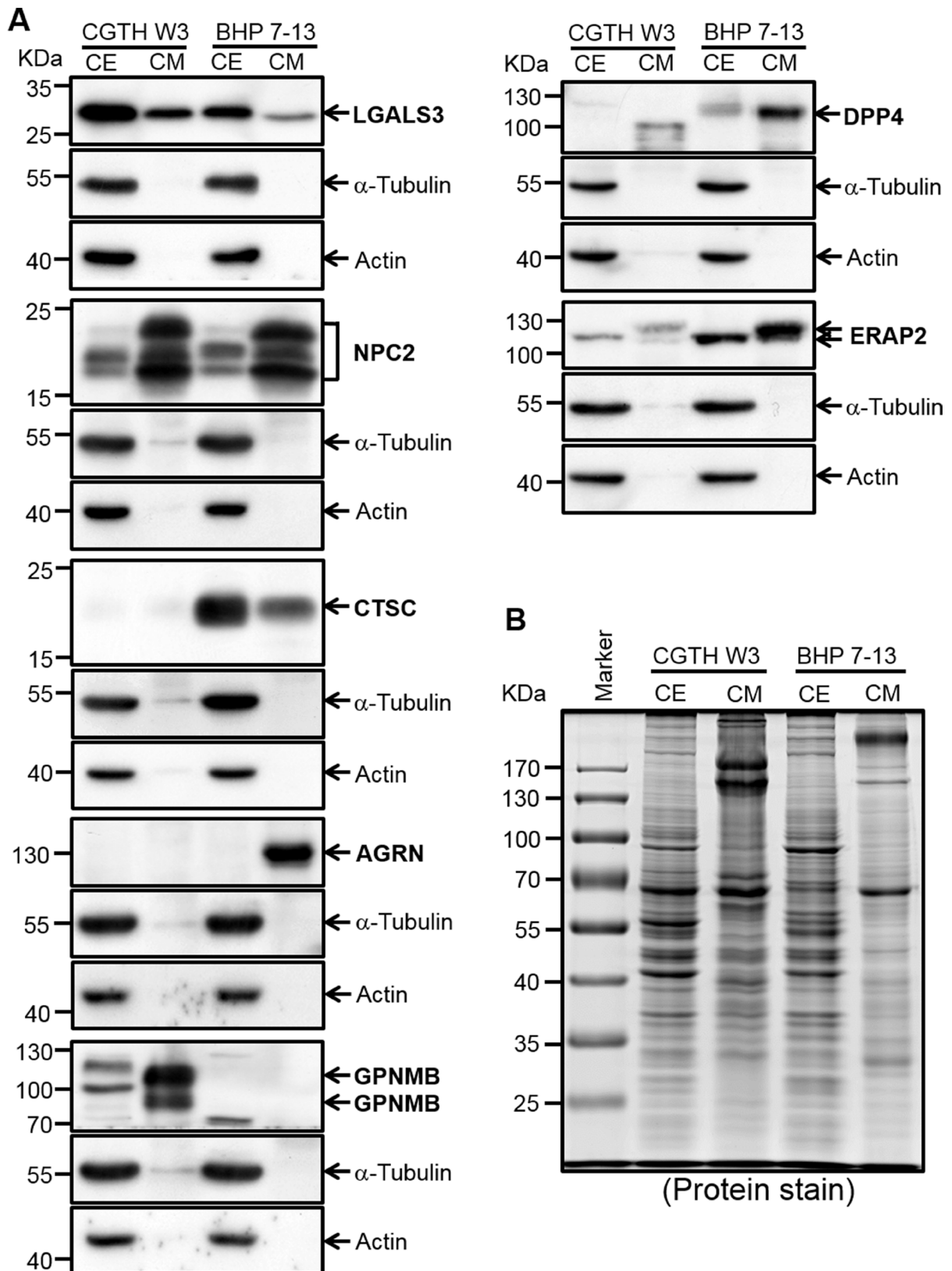


Figure 4: Verification of potential PTC biomarkers in PTC cell extracts and conditioned media by Western blot analysis. (A) Proteins (50 μ g) from conditioned media (CM) and cell extracts (CE) were subjected to Western blot analysis. The utilized PTC cell lines are denoted at the top of the blot. LGALS3, which was previously reported as a potential PTC biomarker, was also included in this analysis. Two cytosolic proteins, α -tubulin and actin, were detected as controls. (B) The Coomassie Blue-stained protein profile was used as a loading control.

Although several potential PTC biomarkers have been identified in the present study, there are several limitations of our approach. First, only a single proteomics analysis of a single pooled cystic fluid of each clinical group and secretome data of only two cancer cell lines were used for the discovery proteomics experiments. This may not be able to capture tumor heterogeneity and thus other potential biomarkers reflecting the tumor heterogeneity. Second, the poor correlation between expression levels of mRNA and protein may complicate the selection of secreted protein candidates deserved for further study. Third, the number of cystic fluid samples used for verification was small. Further study using larger numbers of samples is needed to prove the clinical utility of these candidate biomarkers.

Aggrin (AGRN) is a multifunctional heparan sulfate proteoglycan of the extracellular matrix. It is localized in the basement membrane of the vessels and ducts, and may critically regulate blood-brain barrier conformation and/or synaptogenesis at neuromuscular junctions [41]. Significant overexpression of AGRN was observed during neoangiogenesis in liver cirrhosis and hepatocellular carcinoma, supporting the notion that AGRN stimulates tumor vascularization [42]. AGRN can be detected in ascitic fluid from patients with ovarian cancer, and is secreted by small cell lung cancer cells [43, 44]. An immunoscreening of the extracellular proteome of colorectal cancer cells identified AGRN as an antigen that may be recognized by autoantibodies that exist in sera from colorectal cancer patients [45]. These observations together with our data suggest that AGRN, a proteoglycan that can be secreted by cancer cells via the exosomal pathway, may be a promising PTC marker candidate in cystic fluid.

Cathepsin C (dipeptidyl peptidase I, CTSC) belongs to the papain family of proteinases and participates in the catalytic activation of lysosomal cysteine hydrolase and leukocyte-derived serine proteases [46, 47]. CTSC appears to regulate the degradation of extracellular matrix components that is associated with the metastasis of oral and ovarian cancer cells [48–50]. These observations are consistent with our findings that higher expression of CTSC is correlated with a higher percentage of distant metastasis at diagnosis, locoregional recurrence, and poor prognosis of PTC patients (Table 4 and Figure 7).

Glycoprotein nonmetastatic melanoma protein B (GPNMB) is a type I transmembrane glycoprotein that contains three binding motifs of heparin sulfate proteoglycan, lysosome and integrin; it is highly expressed in bone, where it modulates osteoblast maturation and matrix mineralization [51, 52]. Overexpression of GPNMB has been correlated with tumor formation and metastasis in melanomas, gliomas, hepatocellular carcinoma, and breast cancer [53–56]. The antibody-drug conjugate, glembatumumab vedotin, in which a fully human monoclonal antibody against GPNMB is linked to

the potent cytotoxin, monomethyl auristatin E, has been approved by the Food and Drug Administration for phase II clinical trials in stage III or IV melanoma and GPNMB-expressing metastatic triple negative breast cancer [57, 58]. Although our data do not suggest that GPNMB is a prognostic indicator for PTC, its levels were found to be dramatically elevated in PTC cells and cystic fluid samples (Figures 4 and 5), suggesting that GPNMB is a strong potential candidate for the targeted therapy and/or diagnosis of PTC.

The endoplasmic reticulum aminopeptidases (ERAPs), which include EARP1 and ERAP2 (also called LRAP), play central roles in trimming longer precursors in the endoplasmic reticulum to generate antigenic peptides that are presented on major histocompatibility complex class I (MHC I) molecules [59]. Animal model studies have shown that altered levels of ERAP1 and ERAP2 can facilitate tumor immune evasion [60, 61]. A recent study investigated the expression of both aminopeptidases in a variety of solid tumors and their normal counterpart tissues, and found that the tumor tissues retained, lost, or acquired expression of either or both aminopeptidases, compared to their normal counterparts, depending on the tumor histotype [62]. Of the thyroid specimens examined in this study, only four exhibited high-level expression of both aminopeptidases in tumor cells but none in their normal counterpart tissues. However, our IHC analysis of ~100 PTC tissue specimens demonstrated that ERAP2 is overexpressed in PTC (Figure 6). Future studies are warranted to clarify the role of ERAP2 in PTC carcinogenesis and its potential as a PTC biomarker.

In conclusion, our data showed the significant overexpression of AGRN, CTSC, ERAP2 and NPC2 in PTC tissues, and the tissue expression levels of AGRN and CTSC were significantly associated with metastasis and poor prognosis of PTC patients. Therefore, we considered that integration of multiple omics profiling datasets, from FNA cystic fluid proteome, cancer cell secretome to tissue transcriptome, can be a useful approach to discover novel PTC biomarker candidates while there were not enough clinical samples for large scale and comprehensive analysis. Further studies such as using workable ELISAs to verify these candidates in cystic fluids from a large cohort of patients are warranted to evaluate their utilities in clinical settings.

MATERIALS AND METHODS

Patient characteristics and clinical specimens

The collection and preparation of thyroid cyst fluids were performed as previously described [63]. Briefly, a real-time ultrasonographic machine with a 10-MHz transducer (ALOKA, Tokyo, Japan) was used to detect thyroid nodule and guide fine-needle aspiration with 22- or 25-gauge needles (Becton Dickinson, Singapore).

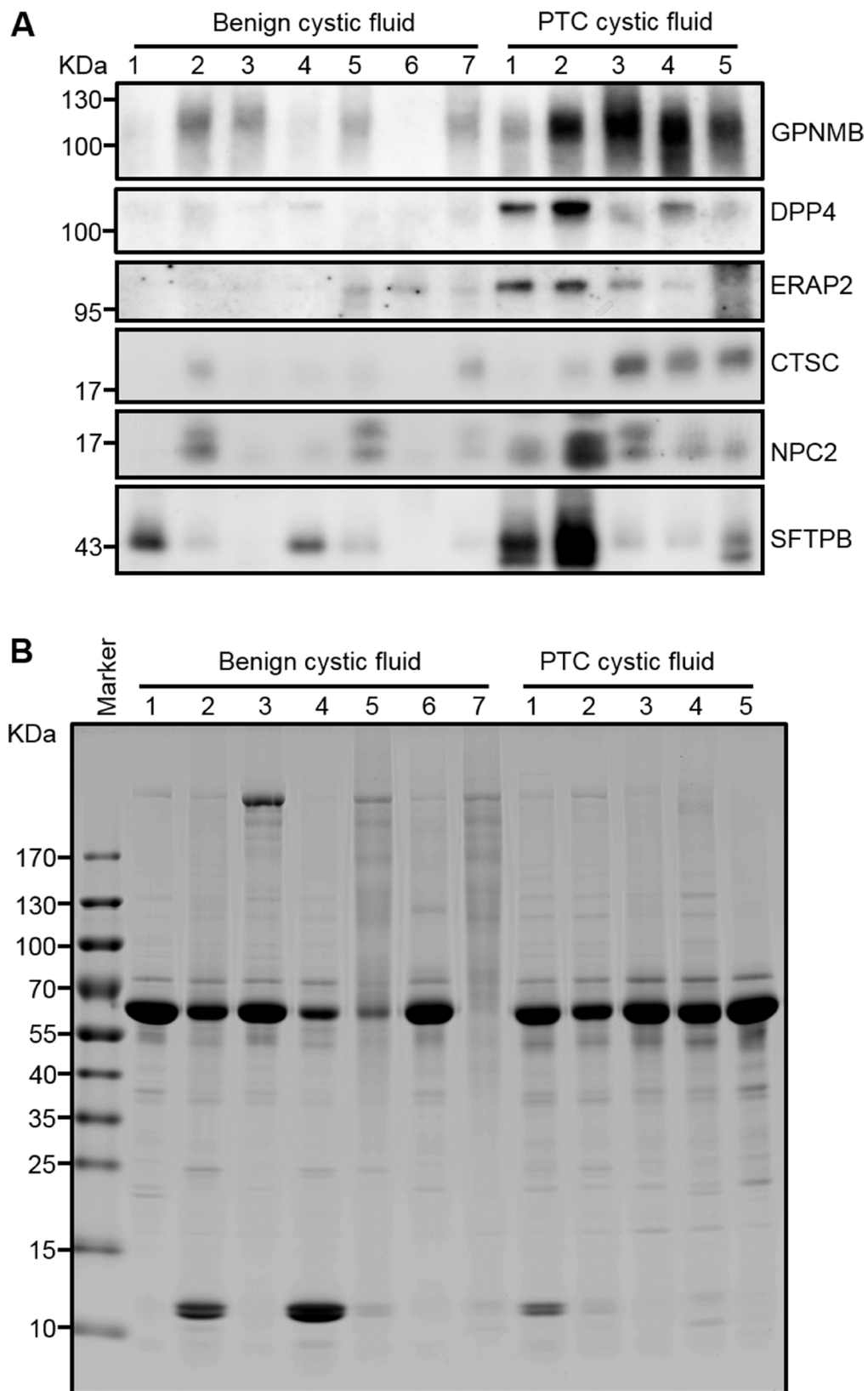


Figure 5: Verification of GPNMB, ERAP2, CTSC, and NPC2 in individual FNA cystic fluid samples by Western blot analysis. (A) Equal amounts of proteins (20 µg) prepared from individual FNA cystic fluid samples were separated by 6-18%-gradient SDS-PAGE, transferred to PVDF membranes, and probed with specific antibodies against the indicated target proteins. SFTPb, which was previously reported as a potential PTC biomarker, was also included in this analysis. (B) The Coomassie-Blue-stained protein profile was used as a loading control.

After the fluid was smeared onto slide, air dried and stained by the Romanowsky-based Liu method [64], fine-needle aspiration cytology (FNAC) was performed and the cytological result was interpreted by a pathologist. Before sample collection, the informed consent forms had been acquired from all patients. Based on cytological and pathological examination, we collected 12 FNA cystic fluid samples from Chang Gung Memorial Hospital (Linkou, Taiwan): seven from patients with benign lesions (2 males and 5 females, mean (\pm SD) age 41.14 ± 13.16 years, age range 21–55) and five from PTC patients (2 males and 3 females, mean age 47 ± 16.46 years, age range 24–70). The cytology tests for the PTC patients were positive for cancer. The features of the FNA cystic fluid samples were assessed by histological analysis. Formalin-fixed and paraffin-embedded samples of thyroid tissues were stained with hematoxylin and eosin (H&E). The guidelines found in “Pathology and Genetics of Tumours of Endocrine Organs” (edited by the World Health Organization, 2004) were used to classify the histopathologic features of the tumor specimens, and clinical staging was based on the definitions of the American Joint Committee on Cancer, 2002 [65, 66]. The characteristics of all study subjects are summarized in Supplemental Table 1. This study was approved by the Institutional Review Board of Chang Gung Memorial Hospital (IRB number 99-3565B).

Depletion of high-abundance proteins from FNA cystic fluid samples

The typical volumes of cyst fluids used for the top 14 abundant protein depletion were 4.28 μ L (protein concentration: 233.64 μ g/ml) and 8.82 μ L (protein concentration: 113.38 mg/ml) which were respectively pooled from 3 malignant and 7 benign cystic fluid samples and then were subjected to depletion of 14 highly abundant proteins using Agilent Human 14 Multiple Affinity Removal System (MARS) columns (4.6 X 100 mm; Agilent, Palo Alto, CA, USA), which harbor antibodies raised against human albumin, IgG, antitrypsin, IgA, transferrin, haptoglobin, fibrinogen, alpha2-macroglobulin, alpha1-acid glycoprotein, IgM, apolipoprotein A1, apolipoprotein A11, complement C3, and transthyretin. Briefly, the FNA cystic fluid sample was prepared at 1 mg/40 μ L in ddH₂O and diluted four-fold (to 1 mg/160 μ L) with 120 μ L buffer A of the MARS column system. The diluted sample was processed using the suggested column run cycle which was coupled with AKTApurifier 10 fast protein liquid chromatography (FPLC) (GE Healthcare Life Sciences, Piscataway, NJ, USA) including sample loading, flow-through collection (depleted fraction), washing, eluting the bound proteins with buffer B of the MARS column system and re-equilibrating the column for the next run. The depleted and bound fractions were desalted and concentrated with Amicon Ultra-15 Centrifugal Filter Devices (MW cutoff,

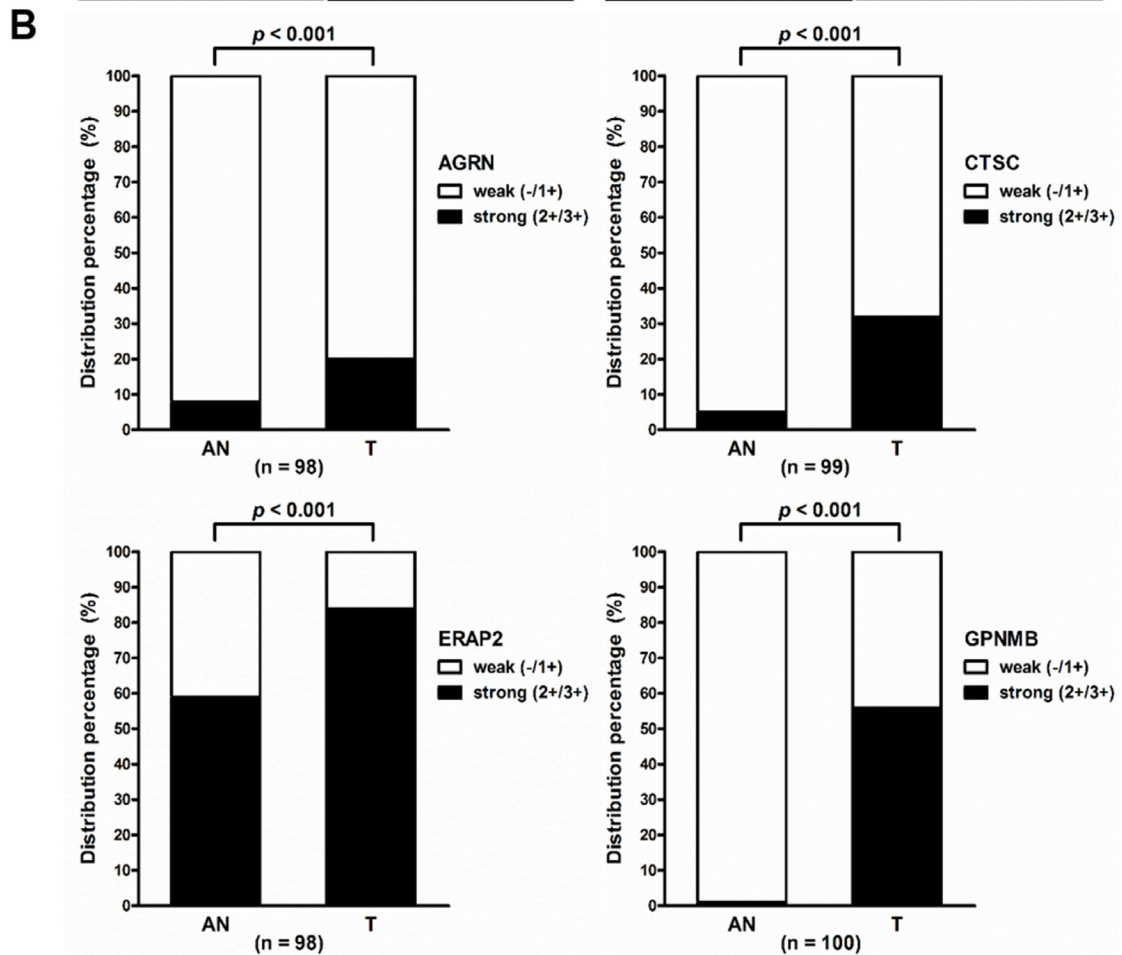
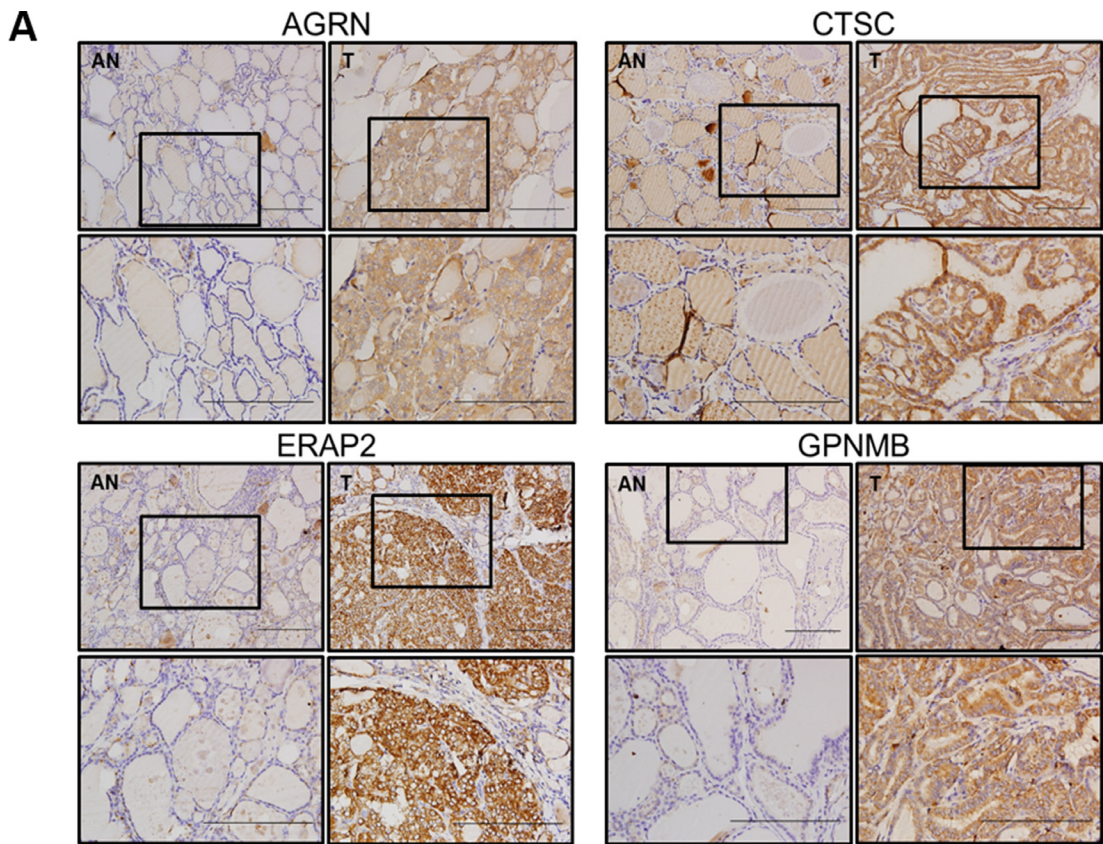
3000 Da; Millipore, Billerica, MA, USA). The proteins were then suspended in ddH₂O, quantified using a Pierce BCA Protein Assay Kit (Thermo Scientific, Hudson, NH, USA), and stored at -80°C for further study.

Cell culture and collection of conditioned media and cells extracts

The PTC cell lines, CGTH W3 and BHP 7-13 were provided from Dr. Jen-Der Lin (Chang Gung Memorial Hospital at Taoyuan, Taiwan) and cultured in RPMI 1640 supplemented with 10% fetal bovine serum and 100 units/mL of penicillin/streptomycin (Invitrogen, Carlsbad, CA, USA) in a humidified 5% CO₂ atmosphere at 37°C. Cells were grown to approximately 80% confluence in 15-mm culture dishes, washed twice with phosphate-buffered saline (PBS) and once with serum-free medium, and then incubated in serum-free medium at 37°C for 24 h. The conditioned media were harvested and centrifuged at 1000 rpm for 10 minutes to eliminate the suspended cells. A proteinase inhibitor cocktail was added to the supernatants (final concentrations: 1 mM phenylmethylsulfonyl fluoride [PMSF], 1 mM benzamidine, 0.5 μ g/ml leupeptin), which were then concentrated and desalted in Amicon Ultra-15 tubes. The cells that had adhered to the dishes were washed twice with PBS and lysed in homogenization buffer (10 mM Tri-HCl, 1 mM EDTA, 1 mM EGTA, 50 mM NaCl, 50 mM NaF, 20 mM Na₄P₂O₇, 1 mM Na₃VO₄, 1 mM PMSF, 1 mM benzamidine, 0.5 μ g/ml leupeptin, and 1% Triton-X100, pH 7.4) on ice for 15 minutes. The cell lysate was collected, sonicated on ice, and centrifuged at 11,000 rpm for 20 minutes at 4°C. The supernatant was harvested as the cell extract. The BCA protein assay reagent (Thermo Scientific Pierce, Rockford, IL, USA) was used to determine the protein concentrations of the cell extracts and conditioned media, which were then stored at -80°C for further use.

One-dimensional SDS-PAGE and in-gel tryptic digestion

The abundant-protein-depleted thyroid FNA samples (60 μ g each) or the conditioned media of the PTC cell lines (50 μ g each) were separated by 8–14% large-gradient SDS-PAGE (gel dimensions: 0.15 \times 12.5 \times 14 cm) and stained with Coomassie Brilliant Blue. The gel-separated proteins were processed for MS analysis using in-gel tryptic digestion, as previously described [39]. Briefly, each gel lane was cut into 60 (for FNA cystic fluid samples) or 40 (for conditioned media of PTC cells) pieces, destained three times (15 min each time) with 30 mM ammonium bicarbonate containing 40% acetonitrile, dehydrated in 100% acetonitrile, and dried in a laminar flow hood. The proteins were then reduced with 25 mM ammonium bicarbonate containing 10 mM dithiothreitol (Merck KGaA, Darmstadt, Germany) at 60°C for 1 h and alkylated with 55 mM iodoacetamide (GE



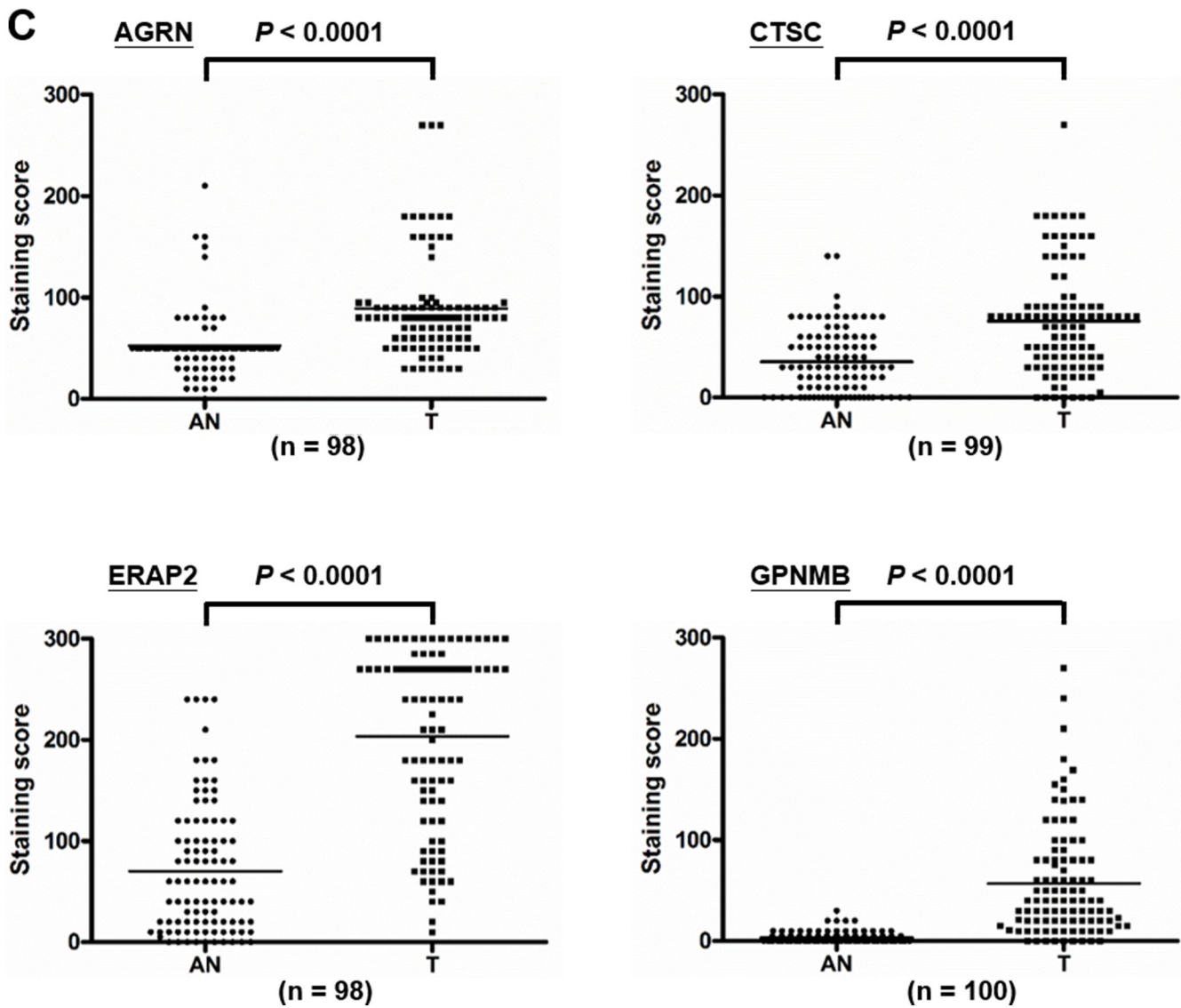


Figure 6: Elevated expression of AGRN, CTSC, ERAP2, and GPNMB in PTC tissues. (A) Immunohistochemical staining for AGRN, CTSC, ERAP2, and GPNMB in paired tumor (*T*) and adjacent non-tumor (*AN*) tissues from two representative cases (*scale bar* = 200 μ m). The boxed areas indicated in the upper panels are enlarged and shown in the lower panels. (B) Statistical analysis of the expression levels of AGRN, CTSC, ERAP2, and GPNMB in PTC specimens harboring both tumor and adjacent non-tumor cells. High-level expression is indicated by strong staining (2+/3+), whereas low-level expression is indicated by weakly positive (1+) and absent (-) staining. (C) Statistical analysis of the staining scores from specimens harboring both *T* and *AN* tissues. The dots indicate each case in the scatter dot plot; the middle line indicates the median.

Healthcare Life Sciences) at room temperature in the dark for 30 min. Sequencing-grade modified porcine trypsin (20 μ g/ml in 25 mM ammonium bicarbonate; Promega, Fitchburg, WI, USA) was used for protein digestion at 37°C for 18 h, and the tryptic peptides were extracted with 100% acetonitrile containing 1% trifluoroacetic acid. Finally, the peptides were concentrated and dried by lyophilization.

Reverse phase liquid chromatography-tandem mass spectrometry

For LC-MS/MS analysis, peptide samples were reconstituted in HPLC buffer A (0.1% formic acid), loaded

across a reversed-phase trapping column (Zorbax 300SB-C18, 0.3 x 5 mm; Agilent Technologies, Wilmington, DE, USA) at a flow rate of 0.2 μ l/min in buffer A, and separated on a 10-cm analytical C₁₈ column (inner diameter, 75 μ m; New Objective, Woburn, MA, USA) using a 15- μ m tip (New Objective). The peptides were eluted using a linear gradient of 0–10% HPLC buffer B (99.9% acetonitrile containing 0.1% formic acid) for 3 min, 10–30% buffer B for 35 min, 30–35% buffer B for 4 min, 35–50% buffer B for 1 min, 50–95% buffer B for 1 min, and 95% buffer B for 8 min, all at a flow rate of 0.25 μ l/min. The LC device was on-line coupled with a two-dimensional linear ion-trap mass spectrometer LTQ-

Orbitrap (Thermo Fisher, San Jose, CA, USA) operated with the Xcalibur 2.0.7 software (Thermo Fisher). The MS full-scan was set to the 350-2000 Da range and the intact peptides were detected at a resolution of 30,000. The ion signal of $(\text{Si}(\text{CH}_3)_2\text{O})_6\text{H}^+$ at m/z 445.120025 was used as a lock mass for internal calibration. The utilized data-dependent analytical mode alternated between one full-scan MS and six MS/MS scans for the six most abundant precursor ions in the MS survey scan. The m/z values selected for MS/MS were dynamically excluded for 40 s. The voltage of the electrospray ionization was 1.8 kV. The MS and MS/MS spectra were both obtained using one microscan with maximum full times of 1000 and 100 ms for MS and MS/MS, respectively. Automatic gain control was performed to prevent the ion trap from becoming overfilled; 5×10^3 ions were collected in the ion trap for the generation of MS/MS spectra.

MS data analysis and label-free spectral quantification

The RAW files of the spectra obtained from the LTQ-Orbitrap were searched against 20,367 Homo sapiens entries in the SwissProt-human_56.0 database, with trypsin assumed as the digestive enzyme. The MASCOT Daemon algorithm (version 2.2.03; Matrix Science, London, UK) was used for data processing, and one missed cleavage was allowed. The MS tolerance for the monoisotopic peptide window was set to 10 ppm, and the MS/MS tolerance was set to 0.5 Da. Carbamidomethyl cysteines (+57 Da) and oxidation of methionine residues (+16 Da) was set as variable modification. The charge states of the peptides were set to +2 and +3. All DAT files produced by MASCOT Daemon were combined using the Scaffold software (version 2.06.00; Proteome Software Inc., Portland, OR, USA) to evaluate the MS/MS-based peptide and protein identifications. The probabilistic threshold of protein identification was set at $\geq 95\%$ and the peptide probability was set at $\geq 95\%$. The confidence of protein identification was based on the assignment of at least two identified unique peptides. The false discovery rate (FDR) was calculated from comparison of the spectra assigned to a decoy database that the spectra (random database) versus those assigned to a normal database. The decoy database was generated by Mascot of the same size (i.e., number of amino acids) and also the same number of proteins as the original normal database [67]. The GeLC-MS/MS label-free spectral counts were applied to determine protein ratios and further compare protein expression levels, using the previously described algorithm [68, 69]. Briefly, to use the spectral report from the scaffold to quantify protein expression, the spectra of each protein were normalized by all spectra detected, and the normalized values from the malignant and benign parts were expressed as a ratio. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium ([\[proteomecentral.org\]\(http://proteomecentral.org\)\) via the PRIDE \[70\] partner repository with the dataset identifier PXD007532.](http://proteomecentral.</p></div><div data-bbox=)

Bioinformatic analysis

The proteins identified from the conditioned media of PTC cell lines and FNA cystic fluid samples were analyzed using the SignalP 4.0 and SecretomeP 2.0 servers to predict the secretory pathways used for proteins with or without signal peptides, respectively [29, 30]. The TMHMM 2.0 program was used to predict transmembrane proteins that may have secretory potential [31]. The DAVID Bioinformatics Resources v6.8 [71, 72] was applied to functional annotation of proteins selected from the discovery experiments, including the biological process, molecular function and cellular component.

Meta-analysis of two PTC tissue mRNA microarray datasets

Two public thyroid cancer tissue microarray datasets, E-GEOD-3678 and GDS1665, were obtained from ArrayExpress of the European Bioinformatics Institute (EBI) and the Gene Expression Omnibus (GEO) of the National Center for Biotechnology Information (NCBI) [32]. The tissue samples used for gene expression profiling were obtained from seven and nine independent PTC patients for the E-GEOD-3678 and GDS1665 datasets, respectively. The two-sample *t*-test was utilized to determine genes whose expression levels were significantly different between PTC and paired normal tissues (p -value ≥ 0.05). The mean intensity of each gene probe was measured from healthy and cancerous groups, the tumor/normal (T/N) ratio was calculated, the ratios were ranked, and the top 5% up- and down-regulated genes were selected. To unify the ID names with those used in the FNA cystic fluid proteome and cell line secretome datasets, the selected gene probe IDs were converted to Swiss-Prot IDs, and comparison was used to select the candidates that most consistently showed high-level expression in cancerous tissues and FNA cystic fluid.

Western blot analysis

Cell extracts and conditioned media (50 μg protein) were resolved by SDS-PAGE, transferred to PVDF membranes (pore size, 0.45 μm ; Millipore), and probed with antibodies against agrin (AGRN) (1:1000 dilution; sc-6166, Santa Cruz Biotechnology, Santa Cruz, CA, USA), CD26/DPP4 (1:1000 dilution; BAF1180, R&D Systems, Minneapolis, MN, USA), NPC2 (1:1000 dilution, sc-166321, Santa Cruz Biotechnology), GPNMB (1:500 dilution; BAF2550, R&D Systems), galectin-3 (LGALS3; 1:1000 dilution; MAB4033, Chemicon International, Temecula, CA, USA), ERAP2 (1:1000 dilution; AF3830, R&D

Systems), CTSC (1:1000 dilution, sc-74590, Santa Cruz Biotechnology), SFTPB (PSPB; 1:1000 dilution, sc-133143, Santa Cruz Biotechnology), tubulin (1:3000 dilution; sc-8035, Santa Cruz Biotechnology), and actin (1:5000 dilution; MAB1501, Millipore). The blots were washed three times with TTBS (20 mM Tris-HCl, pH 7.4, 0.5 M NaCl, and 0.05% Tween 20) and incubated with the appropriate secondary antibody (1:3000 dilution of alkaline phosphatase-conjugated anti-rabbit, -mouse or -goat IgG antibodies; Santa Cruz Biotechnology) for 1 h at room temperature. The blots were then washed three times with TTBS and incubated with the CDP-Star Western Blot Chemiluminescence Reagent (PerkinElmer, Boston, MA, USA) to detect the proteins of interest. For re-blotting, the membranes were stripped with 2% SDS, 100 mM 2-mercaptoethanol and 62.5 mM Tris at 56°C for 45 min with occasional agitation, washed three times with TTBS, and then re-blotting using other antibodies.

Immunohistochemical analysis

Formalin-fixed and paraffin-embedded tissue specimens from 121 PTC patients were obtained from Chang Gung Memorial Hospital and sliced into 4- μ m-thick sections for immunohistochemical (IHC) staining, which was performed using an automatic IHC staining system according to the manufacturer's instructions (Bond; Vision BioSystems, USA). The antibodies used for IHC staining included those against AGRN (1:50 dilution, sc-25528, Santa Cruz Biotechnology), CTSC (1:40 dilution, sc-13986, Santa Cruz Biotechnology), ERAP2 (1:30 dilution, AF3830, R&D Systems), and GPNMB (1:50 dilution, BAF2550, R&D Systems). The IHC staining intensity and percentage in each section were evaluated by an experienced pathologist. Intensity scores of 0, 1, 2, and 3 indicated negative, weak, moderate, and strong staining, respectively, and the percentage score ranged from 0 (0%) to 100 (100%). The two scores were multiplied to obtain the IHC staining score (0 to 300).

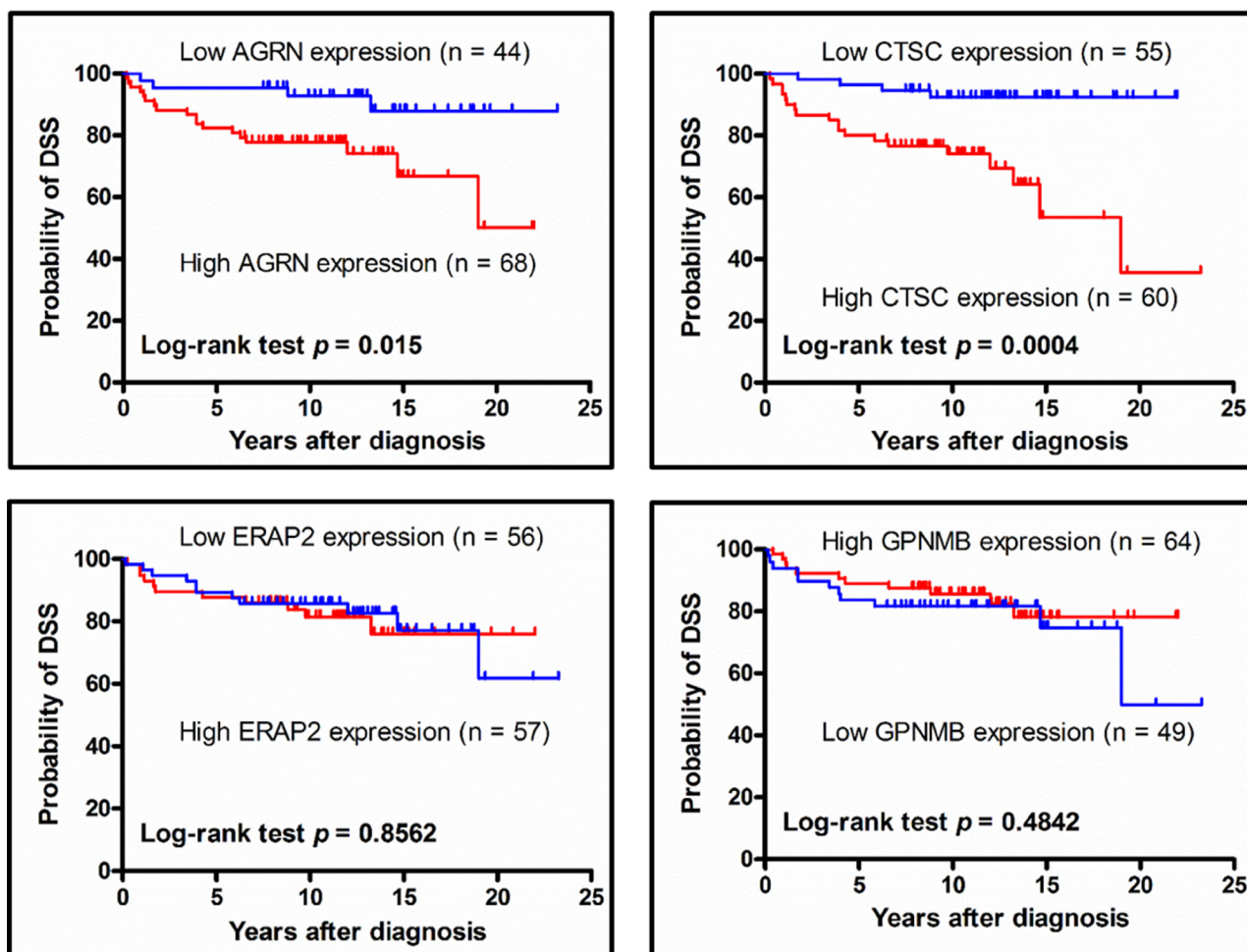


Figure 7: Association between disease-specific survival of PTC patients and target protein expression levels, as analyzed using Kaplan-Meier plots. PTC patient subgroups were stratified by high- and low-level expression of AGRN, CTSC, ERAP2, or GPNMB, and then analyzed for their disease-specific survival using a Kaplan-Meier plot. The log-rank test p -value is denoted in each plot. Two patients died of other causes but not thyroid cancer had been excluded from this analysis. Thus, the numbers of patients used for this analysis are 112, 115, 113 and 113 for AGRN, CTSC, ERAP2 and GPNMB, respectively.

Statistical analysis

The Wilcoxon signed-rank test was used to compare the differences in the IHC staining scores of the target proteins (AGRN, CTSC, ERAP2, and GPNMB) between paired tissue specimens. The percentage of strong (2+/3+) or weak (1+) IHC staining intensity was calculated with McNemar's test. The Chi-square test was used to estimate the correlation between the expression level of the target protein in the tumor part of the tissue sample with various clinicopathological parameters. The Kaplan-Meier method was used for the disease-specific survival (DSS) analysis, and the differences in DSS were assessed by Log-rank test. For all statistical analyses, a two-tailed p -value ≥ 0.05 was considered statistically significant. Calculations were performed using the SPSS 20.0 software (SPSS Inc., Chicago, IL, USA) and the presented diagrams were generated using GraphPad Prism 5.0 (GraphPad Software, Inc., San Diego, CA, USA).

Abbreviations

EBI, European Bioinformatics Institute; ELISA: enzyme-linked immunosorbent assay; FDR, false discovery rate; FNAC, fine-needle aspiration cytology; FPLC, fast protein liquid chromatography; GeLC-MS/MS: in-gel tryptic digestion coupled with liquid chromatography-tandem mass spectrometry; GEO, Gene Expression Omnibus; MARS, multiple affinity removal system; IHC, immunohistochemistry; iTRAQ, isobaric tags for relative and absolute quantitation; LC-MS/MS, liquid chromatography-tandem mass spectrometry; LTQ-Orbitrap, linear ion-trap mass spectrometer; MALDI-TOF-MS: matrix assisted laser desorption / ionization-time of flight mass spectrometer; MRM-HR, high-resolution multiple reaction monitoring; NCBI, National Center for Biotechnology Information; PTC, papillary thyroid carcinoma; SWATH-MS, sequential windowed acquisition of all theoretical fragment ion mass spectra; TM, transmembrane; USG/FNA, ultrasound-guided fine-needle aspiration.

Author contributions

Wu CC, Lin JD, and Yu JS conceived and designed experiments; Wu CC, Chen JT and Chang CM conducted the experiments and analyzed data; Weng HF collected clinical samples; Hsueh C and Chien HP performed IHC scoring. Wu CC, Chen JT and Yu JS wrote the manuscript.

CONFLICTS OF INTEREST

The authors declare no potential conflicts of interest.

FUNDING

This work was supported by the Ministry of Education, Taiwan (to Chang Gung University), by

the Ministry of Science and Technology, Taiwan (grants MOST 103-2325-B-182-003 and 104-2325-B-182-003); and by Chang Gung Memorial Hospital, Taiwan (grants CMRPD1B0531, CIRPD3B0012, CMRPD1C0283, CMRPD1D0271-3, CMRPD1D0301-3 and CLRPD190016).

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. *CA Cancer J Clin.* 2017; 67:7–30. <https://doi.org/10.3322/caac.21387>.
2. Nikiforov YE, Nikiforova MN. Molecular genetics and diagnosis of thyroid cancer. *Nat Rev Endocrinol.* 2011; 7:569–80. <https://doi.org/10.1038/nrendo.2011.142>.
3. Cochand-Priollet B, Guillausseau PJ, Chagnon S, Hoang C, Guillausseau-Scholer C, Chanson P, Dahan H, Warnet A, Tran Ba Huy PT, Valleur P. The diagnostic value of fine-needle aspiration biopsy under ultrasonography in nonfunctional thyroid nodules: a prospective study comparing cytologic and histologic findings. *Am J Med.* 1994; 97:152–7.
4. Carpi A, Di Coscio G, Iervasi G, Antonelli A, Mechanick J, Sciacchitano S, Nicolini A. Thyroid fine needle aspiration: how to improve clinicians' confidence and performance with the technique. *Cancer Lett.* 2008; 264:163–71. <https://doi.org/10.1016/j.canlet.2008.02.056>.
5. Gharib H, Goellner JR. Fine-needle aspiration biopsy of the thyroid: an appraisal. *Ann Intern Med.* 1993; 118:282–9.
6. Kimura ET, Nikiforova MN, Zhu Z, Knauf JA, Nikiforov YE, Fagin JA. High prevalence of BRAF mutations in thyroid cancer: genetic evidence for constitutive activation of the RET/PTC-RAS-BRAF signaling pathway in papillary thyroid carcinoma. *Cancer Res.* 2003; 63:1454–7.
7. de los Santos ET, Keyhani-Rofagha S, Cunningham JJ, Mazzaferri EL. Cystic thyroid nodules. The dilemma of malignant lesions. *Arch Intern Med.* 1990; 150:1422–7.
8. Kashat L, So AK, Masui O, Wang XS, Cao J, Meng X, Macmillan C, Ailles LE, Siu KW, Ralhan R, Walfish PG. Secretome-based identification and characterization of potential biomarkers in thyroid cancer. *J Proteome Res.* 2010; 9:5757–69. <https://doi.org/10.1021/pr100529t>.
9. Chaker S, Kashat L, Voisin S, Kaur J, Kak I, MacMillan C, Ozelek H, Siu KW, Ralhan R, Walfish PG. Secretome proteins as candidate biomarkers for aggressive thyroid carcinomas. *Proteomics.* 2013; 13:771–87. <https://doi.org/10.1002/pmic.201200356>.
10. Giusti L, Iacconi P, Ciregia F, Giannaccini G, Donatini GL, Basolo F, Miccoli P, Pinchera A, Lucacchini A. Fine-needle aspiration of thyroid nodules: proteomic analysis to identify cancer biomarkers. *J Proteome Res.* 2008; 7:4079–88. <https://doi.org/10.1021/pr8000404>.
11. Dinets A, Pernemalm M, Kjellin H, Sviatoha V, Sofiadis A, Juhlin CC, Zedenius J, Larsson C, Lehtio J, Hoog A. Differential protein expression profiles of cyst fluid from

- papillary thyroid carcinoma and benign thyroid lesions. *PLoS One*. 2015; 10:e0126472. <https://doi.org/10.1371/journal.pone.0126472>.
12. Martinez-Aguilar J, Clifton-Bligh R, Molloy MP. Proteomics of thyroid tumours provides new insights into their molecular composition and changes associated with malignancy. *Sci Rep*. 2016; 6:23660. <https://doi.org/10.1038/srep23660>.
 13. Anderson L. Candidate-based proteomics in the search for biomarkers of cardiovascular disease. *J Physiol*. 2005; 563:23–60. <https://doi.org/10.1113/jphysiol.2004.080473>.
 14. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics*. 2002; 1:845–67.
 15. Fischer S, Asa SL. Application of immunohistochemistry to thyroid neoplasms. *Arch Pathol Lab Med*. 2008; 132:359–72. [https://doi.org/10.1043/1543-2165\(2008\)132\[359:AOITTN\]2.0.CO;2](https://doi.org/10.1043/1543-2165(2008)132[359:AOITTN]2.0.CO;2).
 16. Yilmaz E, Karsidag T, Tatar C, Tuzun S. Serum Galectin-3: diagnostic value for papillary thyroid carcinoma. *Ulus Cerrahi Derg*. 2015; 31:192–6. <https://doi.org/10.5152/UCD.2015.2928>.
 17. Cerrato A, Fulciniti F, Avallone A, Benincasa G, Palombini L, Grieco M. Beta- and gamma-catenin expression in thyroid carcinomas. *J Pathol*. 1998; 185:267–72. [https://doi.org/10.1002/\(SICI\)1096-9896\(199807\)185:3<267::AID-PATH113>3.0.CO;2-C](https://doi.org/10.1002/(SICI)1096-9896(199807)185:3<267::AID-PATH113>3.0.CO;2-C).
 18. Anania MC, Miranda C, Vizioli MG, Mazzoni M, Cleris L, Pagliardini S, Manenti G, Borrello MG, Pierotti MA, Greco A. S100A11 overexpression contributes to the malignant phenotype of papillary thyroid carcinoma. *J Clin Endocrinol Metab*. 2013; 98:E1591–600. <https://doi.org/10.1210/jc.2013-1652>.
 19. Vierlinger K, Mansfeld MH, Koperek O, Nohammer C, Kaserer K, Leisch F. Identification of SERPINA1 as single marker for papillary thyroid carcinoma through microarray meta analysis and quantification of its discriminatory power in independent validation. *BMC Med Genomics*. 2011; 4:30. <https://doi.org/10.1186/1755-8794-4-30>.
 20. Barresi G, Tuccari G. Iron-binding proteins in thyroid tumours. An immunocytochemical study. *Pathol Res Pract*. 1987; 182:344–51. [https://doi.org/10.1016/S0344-0338\(87\)80070-5](https://doi.org/10.1016/S0344-0338(87)80070-5).
 21. Lal G, Padmanabha L, Nicholson R, Smith BJ, Zhang L, Howe JR, Robinson RA, O'Dorisio MS. ECM1 expression in thyroid tumors--a comparison of real-time RT-PCR and IHC. *J Surg Res*. 2008; 149:62–8. <https://doi.org/10.1016/j.jss.2007.10.014>.
 22. Brooks E, Simmons-Arnold L, Naud S, Evans MF, Elhosseiny A. Multinucleated giant cells' incidence, immune markers, and significance: a study of 172 cases of papillary thyroid carcinoma. *Head Neck Pathol*. 2009; 3:95–9. <https://doi.org/10.1007/s12105-009-0110-9>.
 23. Topilko A, Caillou B. Acetylcholinesterase and butyrylcholinesterase activities in human thyroid cancer cells. *Cancer*. 1988; 61:491–9.
 24. Park WS, Chung KW, Young MS, Kim SK, Lee YJ, Lee EK. Differential protein expression of lymph node metastases of papillary thyroid carcinoma harboring the BRAF mutation. *Anticancer Res*. 2013; 33:4357–64.
 25. Kondi-Pafiti A, Smyrniotis V, Frangou M, Papayanopoulou A, Englezou M, Deligeorgi H. Immunohistochemical study of ceruloplasmin, lactoferrin and secretory component expression in neoplastic and non-neoplastic thyroid gland diseases. *Acta Oncol*. 2000; 39:753–6.
 26. Sugeno A, Usuda N, Adachi W, Oohashi M, Nagata T, Iida F. Immunohistochemical studies on the localization of fibronectin in human thyroid neoplastic tissues. *Endocrinol Jpn*. 1988; 35:111–20.
 27. Ito Y, Yoshida H, Tomoda C, Uruno T, Miya A, Kobayashi K, Matsuzuka F, Kakudo K, Kuma K, Miyauchi A. S100A4 expression is an early event of papillary carcinoma of the thyroid. *Oncology*. 2004; 67:397–402. <https://doi.org/10.1159/000082924>.
 28. Lucas SD, Karlsson-Parra A, Nilsson B, Grimelius L, Akerstrom G, Rastad J, Juhlin C. Tumor-specific deposition of immunoglobulin G and complement in papillary thyroid carcinoma. *Hum Pathol*. 1996; 27:1329–35.
 29. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*. 2011; 8:785–6. <https://doi.org/10.1038/nmeth.1701>.
 30. Bendtsen JD, Jensen LJ, Blom N, Von Heijne G, Brunak S. Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng Des Sel*. 2004; 17:349–56. <https://doi.org/10.1093/protein/gzh037>.
 31. Moller S, Croning MD, Apweiler R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*. 2001; 17:646–53.
 32. He H, Jazdzewski K, Li W, Liyanarachchi S, Nagy R, Volinia S, Calin GA, Liu CG, Franssila K, Suster S, Kloos RT, Croce CM, de la Chapelle A. The role of microRNA genes in papillary thyroid carcinoma. *Proc Natl Acad Sci USA*. 2005; 102:19075–80. <https://doi.org/10.1073/pnas.0509603102>.
 33. Mathivanan S, Fahner CJ, Reid GE, Simpson RJ. ExoCarta 2012: database of exosomal proteins, RNA and lipids. *Nucleic Acids Res*. 2012; 40:D1241–4. <https://doi.org/10.1093/nar/gkr828>.
 34. Omenn GS. THE HUPO Human Plasma Proteome Project. *Proteomics Clin Appl*. 2007; 1:769–79. <https://doi.org/10.1002/prca.200700369>.
 35. Ozog J, Jarzab M, Pawlaczek A, Oczko-Wojciechowska M, Wloch J, Roskosz J, Gubala E. [Expression of DPP4 gene in papillary thyroid carcinoma]. [Article in Polish]. *Endokrynol Pol*. 2006 (Suppl A); 57:12–7.
 36. Kim WY, Lee JB, Kim HY, Woo SU, Son GS, Bae JW. Endoplasmic reticulum aminopeptidase 2 is highly expressed in papillary thyroid microcarcinoma with cervical lymph node metastasis. *J Cancer Res Ther*. 2015; 11:443–6. <https://doi.org/10.4103/0973-1482.146060>.

37. Chiu CG, Strugnelli SS, Griffith OL, Jones SJ, Gown AM, Walker B, Nabi IR, Wiseman SM. Diagnostic utility of galectin-3 in thyroid cancer. *Am J Pathol.* 2010; 176:2067–81. <https://doi.org/10.2353/ajpath.2010.090353>.
38. Huang Y, Prasad M, Lemon WJ, Hampel H, Wright FA, Kornacker K, LiVolsi V, Frankel W, Kloos RT, Eng C, Pellegata NS, de la Chapelle A. Gene expression in papillary thyroid carcinoma reveals highly consistent profiles. *Proc Natl Acad Sci USA.* 2001; 98:15044–9. <https://doi.org/10.1073/pnas.251547398>.
39. Lin SJ, Chang KP, Hsu CW, Chi LM, Chien KY, Liang Y, Tsai MH, Lin YT, Yu JS. Low-molecular-mass secretome profiling identifies C-C motif chemokine 5 as a potential plasma biomarker and therapeutic target for nasopharyngeal carcinoma. *J Proteomics.* 2013; 94:186–201. <https://doi.org/10.1016/j.jprot.2013.09.013>.
40. Chang KP, Lin SJ, Liu SC, Yi JS, Chien KY, Chi LM, Kao HK, Liang Y, Lin YT, Chang YS, Yu JS. Low-molecular-mass secretome profiling identifies HMGA2 and MIF as prognostic biomarkers for oral cavity squamous cell carcinoma. *Sci Rep.* 2015; 5:11689. <https://doi.org/10.1038/srep11689>.
41. Williams S, Ryan C, Jacobson C. Agrin and neuregulin, expanding roles and implications for therapeutics. *Biotechnol Adv.* 2008; 26:187–201. <https://doi.org/10.1016/j.biotechadv.2007.11.003>.
42. Tatrai P, Dudas J, Batmunkh E, Mathe M, Zaltnai A, Schaff Z, Ramadori G, Kovalszky I. Agrin, a novel basement membrane component in human and rat liver, accumulates in cirrhosis and hepatocellular carcinoma. *Lab Invest.* 2006; 86:1149–60. <https://doi.org/10.1038/labinvest.3700475>.
43. Karlsson NG, McGuckin MA. O-Linked glycome and proteome of high-molecular-mass proteins in human ovarian cancer ascites: identification of sulfation, disialic acid and O-linked fucose. *Glycobiology.* 2012; 22:918–29. <https://doi.org/10.1093/glycob/cws060>.
44. Benatar M, Blaes F, Johnston I, Wilson K, Vincent A, Beeson D, Lang B. Presynaptic neuronal antigens expressed by a small cell lung carcinoma cell line. *J Neuroimmunol.* 2001; 113:153–62.
45. Klein-Scory S, Kubler S, Diehl H, Eilert-Micus C, Reinacher-Schick A, Stuhler K, Warscheid B, Meyer HE, Schmiegel W, Schwarte-Waldhoff I. Immunoscreening of the extracellular proteome of colorectal cancer cells. *BMC Cancer.* 2010; 10:70. <https://doi.org/10.1186/1471-2407-10-70>.
46. Ishidoh K, Muno D, Sato N, Kominami E. Molecular cloning of cDNA for rat cathepsin C. Cathepsin C, a cysteine proteinase with an extremely long propeptide. *J Biol Chem.* 1991; 266:16312–7.
47. Adkison AM, Raptis SZ, Kelley DG, Pham CT. Dipeptidyl peptidase I activates neutrophil-derived serine proteases and regulates the development of acute experimental arthritis. *J Clin Invest.* 2002; 109:363–71. <https://doi.org/10.1172/JCI13462>.
48. Gocheva V, Joyce JA. Cysteine cathepsins and the cutting edge of cancer invasion. *Cell Cycle.* 2007; 6:60–4.
49. Zhuang Z, Jian P, Longjiang L, Bo H, Wenlin X. Oral cancer cells with different potential of lymphatic metastasis displayed distinct biologic behaviors and gene expression profiles. *J Oral Pathol Med.* 2010; 39:168–75. <https://doi.org/10.1111/j.1600-0714.2009.00817.x>.
50. Ahn SE, Choi JW, Rengaraj D, Seo HW, Lim W, Han JY, Song G. Increased expression of cysteine cathepsins in ovarian tissue from chickens with ovarian cancer. *Reprod Biol Endocrinol.* 2010; 8:100. <https://doi.org/10.1186/1477-7827-8-100>.
51. Shikano S, Bonkobara M, Zukas PK, Ariizumi K. Molecular cloning of a dendritic cell-associated transmembrane protein, DC-HIL, that promotes RGD-dependent adhesion of endothelial cells through recognition of heparan sulfate proteoglycans. *J Biol Chem.* 2001; 276:8125–34. <https://doi.org/10.1074/jbc.M008539200>.
52. Safadi FF, Xu J, Smock SL, Rico MC, Owen TA, Popoff SN. Cloning and characterization of osteoactivin, a novel cDNA expressed in osteoblasts. *J Cell Biochem.* 2001; 84:12–26.
53. Tse KF, Jeffers M, Pollack VA, McCabe DA, Shadish ML, Khramtsov NV, Hackett CS, Shenoy SG, Kuang B, Boldog FL, MacDougall JR, Rastelli L, Herrmann J, et al. CR011, a fully human monoclonal antibody-aurostatin E conjugate, for the treatment of melanoma. *Clin Cancer Res.* 2006; 12:1373–82. <https://doi.org/10.1158/1078-0432.CCR-05-2018>.
54. Loding WT, Lal A, Siu IM, Loney TL, Wikstrand CJ, Marra MA, Prange C, Bigner DD, Strausberg RL, Riggins GJ. Identifying potential tumor markers and antigens by database mining and rapid expression screening. *Genome Res.* 2000; 10:1393–402.
55. Onaga M, Ido A, Hasuike S, Uto H, Moriuchi A, Nagata K, Hori T, Hayash K, Tsubouchi H. Osteoactivin expressed during cirrhosis development in rats fed a choline-deficient, L-amino acid-defined diet, accelerates motility of hepatoma cells. *J Hepatol.* 2003; 39:779–85.
56. Rose AA, Pepin F, Russo C, Abou Khalil JE, Hallett M, Siegel PM. Osteoactivin promotes breast cancer metastasis to bone. *Mol Cancer Res.* 2007; 5:1001–14. <https://doi.org/10.1158/1541-7786.MCR-07-0119>.
57. Ott PA, Hamid O, Pavlick AC, Kluger H, Kim KB, Boasberg PD, Simantov R, Crowley E, Green JA, Hawthorne T, Davis TA, Sznol M, Hwu P. Phase I/II study of the antibody-drug conjugate glembatumumab vedotin in patients with advanced melanoma. *J Clin Oncol.* 2014; 32:3659–66. <https://doi.org/10.1200/JCO.2013.54.8115>.
58. Bendell J, Saleh M, Rose AA, Siegel PM, Hart L, Sirpal S, Jones S, Green J, Crowley E, Simantov R, Keler T, Davis T, Vahdat L. Phase I/II study of the antibody-drug conjugate glembatumumab vedotin in patients with locally advanced or metastatic breast cancer. *J Clin Oncol.* 2014; 32:3619–25. <https://doi.org/10.1200/JCO.2013.52.5683>.

59. Saveanu L, Carroll O, Lindo V, Del Val M, Lopez D, Lepelletier Y, Greer F, Schomburg L, Fruci D, Niedermann G, van Endert PM. Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. *Nat Immunol.* 2005; 6:689–97. <https://doi.org/10.1038/ni1208>.
60. Cifaldi L, Lo Monaco E, Forloni M, Giorda E, Lorenzi S, Petrini S, Tremante E, Pende D, Locatelli F, Giacomini P, Fruci D. Natural killer cells efficiently reject lymphoma silenced for the endoplasmic reticulum aminopeptidase associated with antigen processing. *Cancer Res.* 2011; 71:1597–606. <https://doi.org/10.1158/0008-5472.CAN-10-3326>.
61. James E, Bailey I, Sugiyarto G, Elliott T. Induction of protective antitumor immunity through attenuation of ERAAP function. *J Immunol.* 2013; 190:5839–46. <https://doi.org/10.4049/jimmunol.1300220>.
62. Fruci D, Giacomini P, Nicotra MR, Forloni M, Fraioli R, Saveanu L, van Endert P, Natali PG. Altered expression of endoplasmic reticulum aminopeptidases ERAP1 and ERAP2 in transformed non-lymphoid human tissues. *J Cell Physiol.* 2008; 216:742–9. <https://doi.org/10.1002/jcp.21454>.
63. Lin JD, Huang BY. Comparison of the results of diagnosis and treatment between solid and cystic well-differentiated thyroid carcinomas. *Thyroid.* 1998; 8:661–6. <https://doi.org/10.1089/thy.1998.8.661>.
64. Liu CH. A study of staining blood film (Romanowsky system). *J Niigata Med Assoc.* 1956; 70:635–43.
65. LiVolsi VA, Albores-Saavedra J, Asa SL, Baloch ZW, SobrinhoSimoes M, Wenig B, DeLellis RA, Cady B, Mazzaferri EL, Hay I, Fagin JA, Weber AL, Caruso P, et al. Papillary carcinoma. In *Pathology and Genetics of Tumors of the Endocrine Organs*. World Health Organization, Classification of Tumors. pp 57–72. Eds LA DeLellis, RV Lloyd, PU Heitz & C Eng. Lyon: IARC Press, 2004.
66. Sobin LH, Wittekind C. UICC: TNM classification of malignant tumours. New York: Wiley; 2002. pp. 52–6.
67. Wang G, Wu WW, Zhang Z, Masilamani S, Shen RF. Decoy methods for assessing false positives and false discovery rates in shotgun proteomics. *Anal Chem.* 2009; 81:146–59. <https://doi.org/10.1021/ac801664q>.
68. Gokce E, Shuford CM, Franck WL, Dean RA, Muddiman DC. Evaluation of normalization methods on GeLC-MS/MS label-free spectral counting data to correct for variation during proteomic workflows. *J Am Soc Mass Spectrom.* 2011; 22:2199–208. <https://doi.org/10.1007/s13361-011-0237-2>.
69. Carvalho PC, Hewel J, Barbosa VC, Yates JR 3rd. Identifying differences in protein expression levels by spectral counting and feature selection. *Genet Mol Res.* 2008; 7:342–56.
70. Vizcaino JA, Csordas A, del-Toro N, Dianes JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y, Reisinger F, Ternent T, Xu QW, Wang R, Hermjakob H. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* 2016; 44:D447–56. <https://doi.org/10.1093/nar/gkv1145>.
71. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009; 4:44–57. <https://doi.org/10.1038/nprot.2008.211>.
72. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009; 37:1–13. <https://doi.org/10.1093/nar/gkn923>.