Identification of Novel Biomarker and Therapeutic Target Candidates for Diagnosis and Treatment of Follicular Adenoma

XIANYIN LAI[1,2*] and SHAOXIONG CHEN[3*]

[1]*Department of Biochemistry and Molecular Biology,*
Indiana University School of Medicine, Indianapolis, IN, U.S.A.;
[2]*Department of Cellular & Integrative Physiology,*
Indiana University School of Medicine, Indianapolis, IN, U.S.A.;
[3]*Department of Pathology and Laboratory Medicine,*
Indiana University School of Medicine, Indianapolis, IN, U.S.A.

Abstract. *Follicular adenoma is a type of benign and encapsulated nodule in the thyroid gland, but some adenomas have the potential to progress to follicular carcinoma. Therefore, it is important to monitor the state and progress of follicular adenoma in the clinic and discover drug development targets for the treatment of follicular adenoma to prevent its worsening to follicular carcinoma. Currently, the study of biomarkers and therapeutic targets lacks applications of up-to-date technologies, including proteomics and bioinformatics. To discover novel protein biomarker and therapeutic target candidates, a liquid chromatography-tandem mass spectrometry approach was applied to directly compare follicular adenoma with normal thyroid tissue samples. The proteomics analysis revealed 114 protein biomarker candidates out of 1,780 identified and quantified proteins. A comprehensive approach to prioritize the biomarker candidates by category and rank revealed CD63, DDB1, TYMP, VDAC2, and DCXR as the top five biomarker candidates. Upstream regulator analysis using the Ingenuity Pathway Analysis (IPA) software discovered four therapeutic target candidates for follicular adenoma, including TGFB1, MYC, ANGPT2, and NFE2L2. This study provided biomarker and therapeutic target candidates for a follow-up study, which will facilitate monitoring and treatment of follicular adenoma.*

*These authors contributed equally to this study.

Abbreviations: FNA, Fine-needle aspiration; IPA, Ingenuity Pathway Analysis; PIBAP, prioritization index of biomarker candidates for assay of plasma/serum specimens; TPP, Trans-Proteomic Pipeline.

Correspondence to: Xianyin Lai, Ph.D., Department of Biochemistry and Molecular Biology, Department of Cellular and Integrative Physiology, Indiana University School of Medicine, 635 Barnhill Drive, Room 0044, Indianapolis, IN 46202, U.S.A. Tel: +1 3172780084, Fax: +1 3172744686, e-mail: xlai@iu.edu and Dr. Shaoxiong Chen, IU Health Pathology Laboratory, 350 West 11th Street, Room 4088, Indianapolis, IN 46202, U.S.A. Tel: +1 3174916443, Fax: +1 3174916114, e-mail: chen251@iupui.edu

Key Words: Adenoma, biomarker, follicular, mass spectrometry, therapeutic target.

A lump within the thyroid gland is called a thyroid nodule and consists of an abnormal growth of thyroid cells. Most are asymptomatic. Occasionally, a lump in the neck is noticed by the patient. More often, a lump is discovered incidentally during a medical procedure for completely unrelated reasons. However, thyroid nodules are among the most common endocrine complaints in the United States. Approximately 10-20 million Americans have clinically-detectable thyroid nodules (1) and nearly 50% of the population have thyroid nodules at autopsy (2).

According to the American Thyroid Association guidelines on thyroid nodules and differentiated thyroid cancer published 2009, a series of medical history investigation, physical examination, laboratory tests including measurement of thyroid-stimulating hormone, ultrasonography, and fine needle biopsy should be performed with the discovery of a thyroid nodule (3). Fortunately, over 90% of thyroid nodules are benign (1).

Generally, benign nodules are not cancerous and do not need to be removed, but should be watched closely with ultrasound examination every 6-12 months and annual physical examination (3, 4). However, among various types of benign nodules, follicular adenoma has an increased risk of malignancy, leading to follicular carcinoma (5, 6). Therefore, it is extremely important to monitor the state and development of follicular adenoma in clinical management and find drug targets for the treatment of follicular adenoma to prevent its worsening to follicular carcinoma.

Currently, fine-needle aspiration (FNA) biopsy is the most widely used and most accepted diagnostic test for a thyroid nodule. However, several issues have been raised, including the inherent inadequacy of cytological specimens in assessing capsular and vascular involvement of thyroid follicular lesions (7). Because the malignant nature of follicular lesions requires histologic proof of capsular or vascular invasion, any FNA diagnosis of a follicular lesion is inherently uncertain (8). Alternatively, the analysis of *BRAF*, *RAS*, and other mutations in cytological samples has been applied to follicular adenoma management (8, 9). However, no marker or group of markers has yet been sufficiently validated, and measurements of molecular markers remain investigational (7). The lack of convenient biomarkers, such as proteins in plasma, makes monitoring of follicular adenomas difficult. Thus, the current study aimed to profile the proteomic differences between follicular adenoma and normal tissue in order to explore the possible mechanisms and identify potential biomarkers for the development of novel monitoring techniques and therapeutic approaches.

Through literature search, we found approximately 10 articles reporting on proteomic profiling of follicular adenoma (10-12). Among these, five have directly compared follicular adenoma with normal tissue (13-17). However, only an outdated proteomic technique, two-dimensional electrophoresis technique, has been applied. Proteomic technology has dramatically improved over the last decade. A shotgun proteomics approach provides better profiling of proteins because of its sensitivity and high-throughput capability (18). Therefore, a liquid chromatography-tandem mass spectrometry (LC-MS/MS) approach should be investigated on the proteomic profiling of follicular adenoma.

In the present study, fresh-frozen tissue samples of normal and follicular adenoma were analyzed using a quantitative LC-MS/MS approach. Profiling of the two groups of tissues was carried-out for direct comparison. The results revealed possible mechanisms and potential biomarkers involved in follicular adenoma, significantly expanding the number of meaningful biomarker candidates for monitoring follicular adenoma and developing drugs to treat follicular adenoma before it worsens to follicular carcinoma.

Materials and Methods

Materials. Urea, DL-dithiothreitol, triethylphosphine, iodoethanol, and ammonium bicarbonate (NH_4HCO_3) were purchased from Sigma-Aldrich (St. Louis, MO, USA). LC-MS grade water (H_2O), LC-MS grade 0.1% formic acid in acetonitrile (ACN), and 0.1% formic acid in water (H_2O) were obtained from Burdick & Jackson (Muskegon, MI, USA). Modified sequencing grade porcine trypsin was purchased from Princeton Separations (Freehold, NJ, USA).

Protein extraction. The study was approved by the Indiana University Institutional Review Board (IRB #1206008865). Follicular adenoma and normal tissues were obtained from the Indiana University Health Methodist Research Institute (Indianapolis, IN, USA). Protein extraction from the tissue samples was performed according to a published procedure (19). After each tissue sample was weighed, it was placed in a 50-ml beaker, and 8 volumes of lysis buffer were added. The tissue was thoroughly minced with needle-nose surgical scissors. The buffer/minced tissue solution was transferred to a ground-glass homogenizer tube. The material was grinded-up by twisting with an up/down motion until no solid tissue remained. The homogenates were transferred into a 2-ml microcentrifuge tube and centrifuged at $15,000 \times g$ for 10 min at -4 °C to remove insoluble materials. Fully-solubilized samples were then stored at -80 °C until analysis. Protein concentration was determined by the Bradford Protein Assay using Bio-Rad protein assay dye reagent concentrate (20).

Protein reduction, alkylation, and digestion. Protein reduction, alkylation, and digestion were carried-out using a method previously published by the author (21). Briefly, a 100-μg aliquot of protein sample was placed in a 2 ml tube and then adjusted to 200 μl of 4 M urea by adding water. Two hundred μL of the reduction/alkylation cocktail consisted of triethylphosphine and iodoethanol were added to each of the tubes. The sample was incubated at 35°C for 60 min, dried by SpeedVac, and reconstituted with 100 μL of 100 mM NH_4HCO_3 at pH 8.0. A 150 μL aliquot of a 20 μg/mL trypsin solution was added to the sample and incubated at 35°C for 3 h, after which another 150 μL of trypsin was added, and the solution incubated at 35°C for 3 h.

LC-MS/MS analysis. The digested samples were analyzed using a Thermo Scientific linear ion-trap (LTQ) mass spectrometer coupled with a Surveyor autosampler and MS HPLC system (Thermo Scientific, Waltham, MA, USA). Tryptic peptides were injected onto the C18 RP column (TSKgel ODS-100V, 1.0 mm x 150 mm) at a flow rate of 50 μl/min. The mobile phases A and B were 0.1% formic acid in water and 50% ACN with 0.1% formic acid in water, respectively. The gradient elution profile was as follows: 10.0% B (90.0% A) for 7 min, 10.0-20.6% B (90.0-79.4% A) for 5 min, 20.6-65.6% B (79.4-34.4% A) for 148 min, 65.6-100.0% B (34.4-0.0% A) for 10 min, 100.0% B for 10 min. The data were collected in the "Data-dependent MS/MS" mode with the ESI interface using normalized collision energy of 35%. Dynamic exclusion settings were set to repeat count 1, repeat duration 30 s, exclusion duration 120 s, and exclusion mass width 0.6 m/z (low) and 1.6 m/z (high).

Protein identification and quantification. Acquired data were searched against the UniProt protein sequence database of HUMAN (released on 10/03/2012) using SEQUEST (v.28 rev.12) algorithms in Bioworks (v.3.3). General parameters were set to: mass type set as "monoisotopic precursor and fragments", enzyme set as "trypsin (KR)", enzyme limits set as "fully enzymatic - cleaves at both ends", missed cleavage sites set at 2, peptide tolerance 2.0 amu, fragment ion tolerance 1.0 amu, fixed modification set as +44 Da on Cysteine, and no variable modifications used. The searched peptides and proteins were validated by PeptideProphet (22) and ProteinProphet (23) in the Trans-Proteomic Pipeline (TPP, v.3.3.0) (http://tools.proteomecenter.org/software.php). Only proteins and peptides with protein probability ≥0.9000 and peptide probability ≥0.8000 were reported. Protein quantification was performed using a label-free quantification software package, IdentiQuantXL™ (Indianapolis, IN, USA) (24).

Table I. *The ranking of biomarker candidates in the Category 1.*

Rank	ID	Gene	PFN	PFF	FD	FC	NP	SS	SD	NC	PIBAP
1	F8W022	CD63	0.2	0.2	0.0	5.2	2	7	3	6	70.0
2	F5GY55	DDB1	0.2	0.2	0.0	3.8	2	1	3	14	58.0
3	H9KVA0	TYMP	0.3	0.2	−0.2	3.3	2	10	9	5	51.9
4	P45880-1	VDAC2	0.3	0.7	0.3	2.5	2	4	3	5	50.7
5	Q7Z4W1	DCXR	0.2	0.2	0.0	4.3	2	4	0	1	50.0
6	B4DLZ5	NAGK	0.2	0.0	−0.2	4.0	3	4	0	9	47.7
7	P31040	SDHA	0.3	0.7	0.3	1.7	5	4	9	2	46.0
8	H0Y9B6	HEXB	0.0	0.3	0.3	2.3	2	4	0	5	45.0
9	P00739	HPR	0.3	0.8	0.5	1.6	2	10	3	1	45.0
10	H7C2Y0	SEPT2	0.2	0.2	0.0	3.0	2	4	0	6	42.0
11	Q9BXN1	ASPN	0.3	0.5	0.2	1.6	7	10	3	1	41.3
12	P34932	HSPA4	0.3	0.0	−0.3	3.0	4	4	3	12	39.7
13	G3XA98	FBLN5	0.5	0.7	0.2	1.5	3	10	3	3	38.2
14	P54652	HSPA2	0.2	0.3	0.2	1.5	3	4	9	3	38.2
15	F8VPE8	RPLP0	0.2	0.2	0.0	2.6	2	4	3	3	38.0
16	F5GWA7	PHB2	0.2	0.3	0.2	1.5	2	4	3	8	36.2
17	B4DY47	PHB	0.5	0.5	0.0	1.5	2	1	9	9	36.0
18	E7EN95	FLNB	0.5	0.3	−0.2	2.6	2	4	3	6	35.0
19	H0YIZ1	NPC2	0.3	0.3	0.0	1.9	2	10	0	4	35.0
20	P00915	CA1	0.7	0.5	−0.2	1.9	11	4	3	1	33.2
21	P84243	H3F3A	0.7	0.8	0.2	1.5	5	1	3	3	31.2
22	E9PRY8	EEF1D	0.3	0.3	0.0	1.8	3	4	3	3	31.0
23	B1AR63	HK1	0.0	0.2	0.2	1.5	2	4	3	2	30.2
24	B7Z4L4	RPN1	0.3	0.5	0.2	1.5	4	4	0	3	30.2
25	B4DZM8	PSMD5	0.2	0.0	−0.2	2.9	2	0	3	2	29.5
26	P14923	JUP	0.3	0.0	−0.3	2.1	2	7	3	6	28.7
27	F5GWX2	HEBP1	0.2	0.3	0.2	1.6	2	4	0	2	28.3
28	Q99832-3	CCT7	0.5	0.3	−0.2	2.0	2	4	3	3	27.0
29	O00232	PSMD12	0.3	0.2	−0.2	1.9	2	4	3	3	26.2
30	Q99714-2	HSD17B10	0.3	0.0	−0.3	1.5	4	4	9	2	25.7
31	E7EX44	CALD1	0.8	0.5	−0.3	2.0	4	4	3	4	25.0
32	Q9H3G5	CPVL	0.2	0.2	0.0	1.5	3	4	0	2	24.0
33	Q96C19	EFHD2	0.2	0.0	−0.2	2.7	2	0	0	1	23.8
34	P50502	ST13	0.5	0.3	−0.2	1.8	2	4	0	4	23.3
35	H0YL72	IDH3A	0.3	0.2	−0.2	1.5	3	4	3	1	21.9
36	O94874	UFL1	0.2	0.0	−0.2	1.7	3	4	0	1	20.5
37	E7ERW2	GOT2	0.5	0.3	−0.2	1.6	2	4	0	2	19.7
38	Q9UBV8	PEF1	0.3	0.0	−0.3	2.4	2	4	0	1	19.7
39	B2RDM2	TXNDC5	0.3	0.0	−0.3	1.5	2	4	3	4	19.7
40	P00492	HPRT1	0.2	0.0	−0.2	1.5	2	4	0	1	17.8
41	B4DXP9	ACTR1A	0.3	0.0	−0.3	1.7	2	4	0	3	17.0
42	E7EW08	SHMT1	0.3	0.0	−0.3	1.7	2	4	0	3	17.0
43	O43684-2	BUB3	0.2	0.0	−0.2	1.5	2	1	0	3	16.8
44	P38919	EIF4A3	0.2	0.0	−0.2	1.7	2	1	0	1	16.5
45	B3KST5	COPS4	0.5	0.0	−0.5	1.8	2	4	3	3	16.0
46	Q9Y678	COPG1	0.5	0.0	−0.5	1.7	5	4	0	2	14.5
47	H7BZT7	ESD	0.3	0.0	−0.3	1.5	2	4	0	1	13.7

PFN, Protein identification frequency in normal tissues; PFF, protein identification frequency in follicular adenoma tissues; FD, protein identification frequency difference between normal and follicular adenoma tissues. FC, Fold change; NP, number of peptides; SS, score of subcellular location; SD, score of involvement in targeted diseases and disorders; NC, network connectivity; and PIBAP, prioritization index of biomarker candidates for assay of plasma/serum specimens. The PIBAP is calculated using a formula of FD × 10 + FC × 10 + FD × FC × 10 + NP + SS + SD + NC.

Results and Discussion

Protein identification and quantification. From the normal and follicular adenoma tissue samples, 1,812 protein groups (unique proteins) with a probability ≥0.9 were identified by 6,853 peptides with a probability ≥0.8. In the identification process, proteins identified with completely identical peptides are placed into a single protein group. Among the 1,812 proteins, 1,114 proteins were identified with at least two distinct peptides. To obtain more accurate

quantification, multiple filters were applied to eliminate unqualified peptides for protein quantification (24). The complete list of quantified proteins is available in Table I (to be provided upon request from the authors), including 1,780 protein groups quantified by 6,421 peptides. Among them, 1,089 proteins were quantified with at least two distinct peptides.

Biomarker candidates. Once numerous proteins are identified and quantified in the biomarker discovery phase, the challenging task is to determine which proteins should be chosen for further validation using alternative approaches. A comprehensive approach to prioritizing and ranking proteins has been published by the authors (25). One of the aims of the present study was to discover biomarker candidates to monitor follicular adenoma in clinical management by testing the blood level of certain proteins. An ideal biomarker should be low abundance in normal tissue while high abundance in follicular adenoma tissue. Volcano plots in Figure 1 illustrate the protein changes between normal and follicular tissues. Only proteins highlighted in the red square are up-regulated by at least 1.5-fold and considered potential biomarker candidates for further prioritization and rank. Excluding protein isoforms, 114 unique proteins were considered biomarker candidates.

Prioritizing the biomarker candidates by categorizing. To have a better chance of success in biomarker validation, we categorized the candidates with different priorities according to the most critical factors. Because residual blood could be present in the tissue samples, proteins from blood rather than the tissue cells would be identified as well. However, these proteins should not be easily excluded from the candidate list, since they are usually expressed in multiple types of tissues and these proteins may be from blood and/or the analyzed tissues. Therefore, common plasma proteins, such as fibrinogen (26), in the candidate list were labeled as Category 3, in which proteins have lowest priority in the candidate list for further validation (Supplementary Table II, to be provided upon request from the Authors).

Proteins identified and quantified by one peptide with high probability should not be rejected, because potentially important biological information or novel biomarkers may be discarded before they are even provided the opportunity of being validated (27-29). However, proteins with more than one peptide should have higher priority than proteins with only one peptide. Based on this criterion, proteins with one peptide are put into Category 2, in which proteins have low priority for further validation (Supplemental Table III, to be provided upon request from the authors). Other proteins are put into Category 1, in which proteins have high priority among the entire list (Table I).
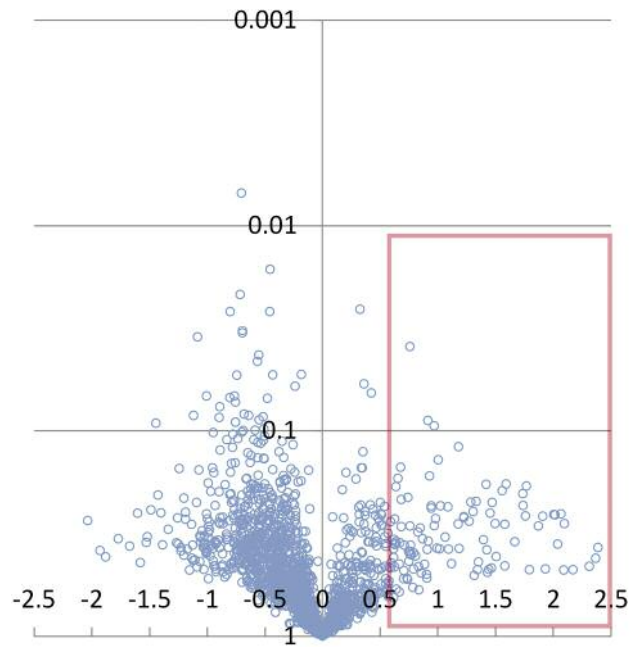

Figure 1. *Volcano plots of the protein changes between the follicular adenoma and normal thyroid tissues. Only proteins highlighted in the red square are up-regulated by at least 1.5-fold and considered potential biomarker candidates for further prioritization and rank. Excluding protein isoforms, 114 unique proteins were considered as biomarker candidates.*

Ranking the biomarker candidates inside each category. To further improve success in biomarker validation, the candidates were ranked inside each category based on multiple factors. According to the comprehensive approach of prioritizing and ranking (25), frequency difference, fold change, the consistency between frequency difference and fold change, and the number of peptides were the basic components for ranking. These numbers were directly available from Supplemental Table I. Besides these basic components, subcellular location, involvement in targeted diseases and disorders, and network connectivity from informatics analysis were also scored and included.

In order to assign their subcellular locations, involvement in diseases and disorders, and network connectivity, proteins in each category were individually submitted to QIAGEN's Ingenuity® Pathway Analysis (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity) in which the Ingenuity Pathways Knowledge Base is used. Besides subcellular location, involvement in diseases and disorders proteins, and network connectivity, IPA provided important information, such as canonical pathways, upstream regulators, regulator effects, *etc*. To affiliate ranking of the biomarker candidates, only the highly related results are presented in this publication. The network connectivity of proteins in
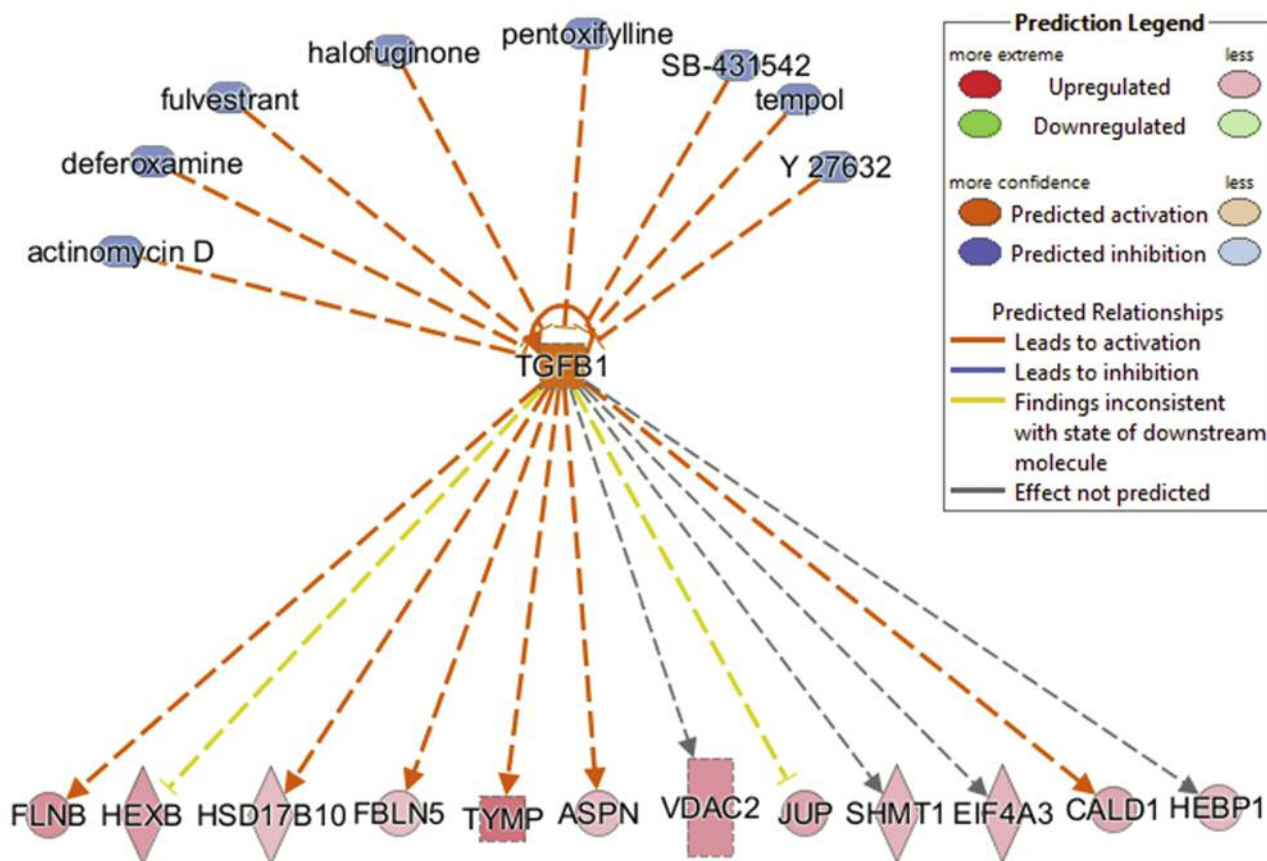
Figure 2. *A merged network from three networks generated by IPA using all 47 proteins in Category 1. The gray lines show the connectivity inside each network and the pink lines indicates new connections generated by merging the three networks into one. The connections of each protein can be counted from the network.*

Categories 1, 2, and 3 are presented in Figure 2 and Supplemental Figure 1 and 2 respectively (to be provided upon request from the Authors). The number of a protein's connections in a network is counted and reported in Table I, Supplemental Table I, and Supplemental Table II. The subcellular location and the involvement in diseases and disorders proteins of each protein are included in Table I, Supplemental Table I, and Supplemental Table II. In this

Table, scores of 10, 7, 4, 1, and 0 were assigned for extracellular space, plasma membrane, cytoplasm, nucleus, and other (unknown or unassigned), respectively. For the involvement in targeted diseases and disorders, proteins involved in endocrine disorders were assigned 6 points, proteins related to cancer were assigned 3 points, and proteins involved in both disorders were assigned 9 points.

275

Figure 3. *The upstream regulators of TGFB1 and its target proteins from Category 1. Chemical drugs or reagents highlighted in blue inhibit TGFB1, while TGFB1 activates 10 out of the 12 proteins highlighted in red. Based on our results that the expression of proteins at the bottom is increased, IPA predicted that the expression of TGFB1 is increased as well, but the amount of the chemicals should be decreased. If more chemicals are applied, the expression of TGFB1 will decrease and so does the expression of the 10 proteins at the bottom. The orange lines show activation effect of six proteins from the 10 proteins was positively confirmed by IPA. The gray lines means the other four proteins lacked literature support to predict the activation effect of TGFB1. The yellow lines indicates our results and the literature finds in IPA are not consistent. From the IPA knowledge database, 10 chemical drugs or reagents have been found to inhibit TGFB1.*

Finally, a PIBAP (prioritization index of biomarker candidates for assay of plasma/serum specimens) score was calculated based on all numbers from the above using a formula of FD × 10 + FC × 10 + FD × FC × 10 + NP + SS + SD + NC, where FD, FC, FD × FC, NP, SS, SD , and NC represent frequency difference, fold change, consistency between frequency difference and fold change, number of peptides used for protein identification, score of protein subcellular location, score of the involvement in targeted diseases and disorders, and connectivity in a network, respectively. Proteins were ranked according to their PIBAP score within each category.

Top biomarker candidates. After categorizing and ranking, the top five biomarker candidates in Category 1 were CD63, DDB1, TYMP, VDAC2, and DCXR. CD63 is a cellular

membrane protein and strongly expressed in the early stages of melanoma (30, 31). Our data showed it was 5.2-fold up-graded in the follicular adenoma and it had the greatest fold change among the top five candidates. DDB1 is involved in the nucleotide excision repair pathway as a component of the damaged DNA binding protein complex. It plays a role in the transcriptional regulation of UV-induced genes (32). TYMP catalyzes the reversible dephosphorylation of thymidine, deoxyguridine, and their analogs. Its expression is elevated in many solid tumors where it is likely involved in mechanisms that regulate cell proliferation, apoptosis, and angiogenesis. It has been intensively investigated to ascertain whether TYMP is a biological marker in breast carcinoma (33, 34). VDAC2 mediates the exchange of metabolites through the mitochondrial membrane. Genetic and biochemical studies have indicated that VDAC2 is anti-
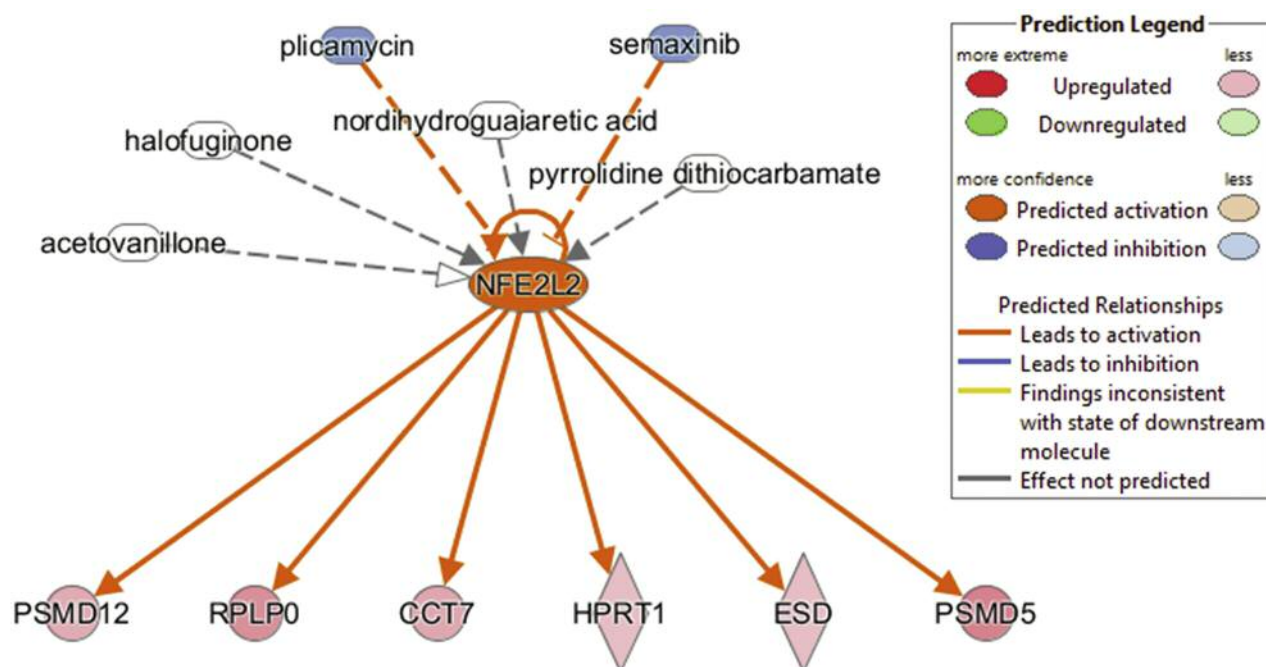
Figure 4. *The upstream regulators of NFE2L2 and its target proteins from Category 1. Based on the expression of the six proteins, IPA predicted that NFE2L2 was activated in the follicular adenoma tissue. To inhibit the expression of NFE2L2, two chemical drugs, plicamycin and semaxinib, have been revealed by IPA to decrease the express of NFE2L2.*
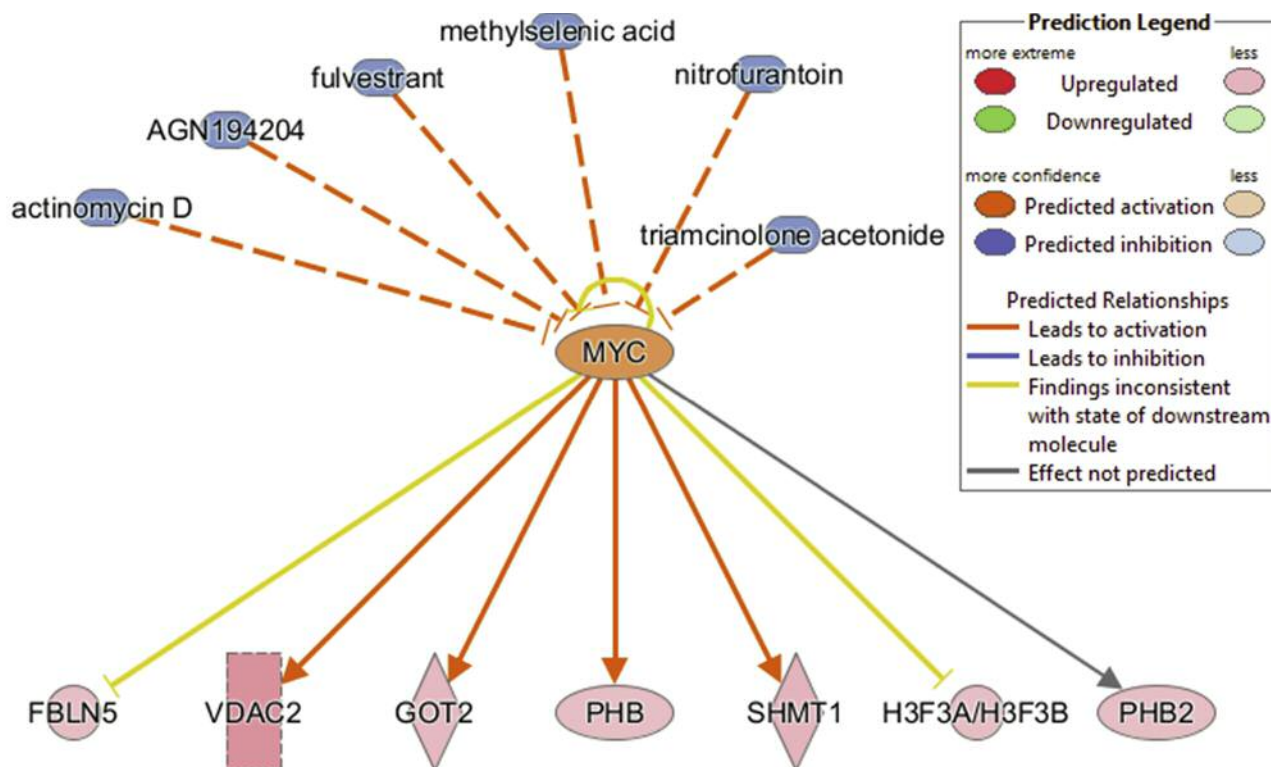

Figure 5. *The upstream regulators of MYC and its target proteins from Category 1. According to literature, five out of the seven proteins are activated by MYC. The activation effect of four of these five proteins was positively confirmed by IPA, which are shown with an orange line with an arrow at the end. The remaining protein lacked literature support to predict the activation effect of MYC, which is shown with a gray line with an arrow at the end. From the IPA knowledge database, six chemical drugs or reagents have been found to inhibit MYC.*
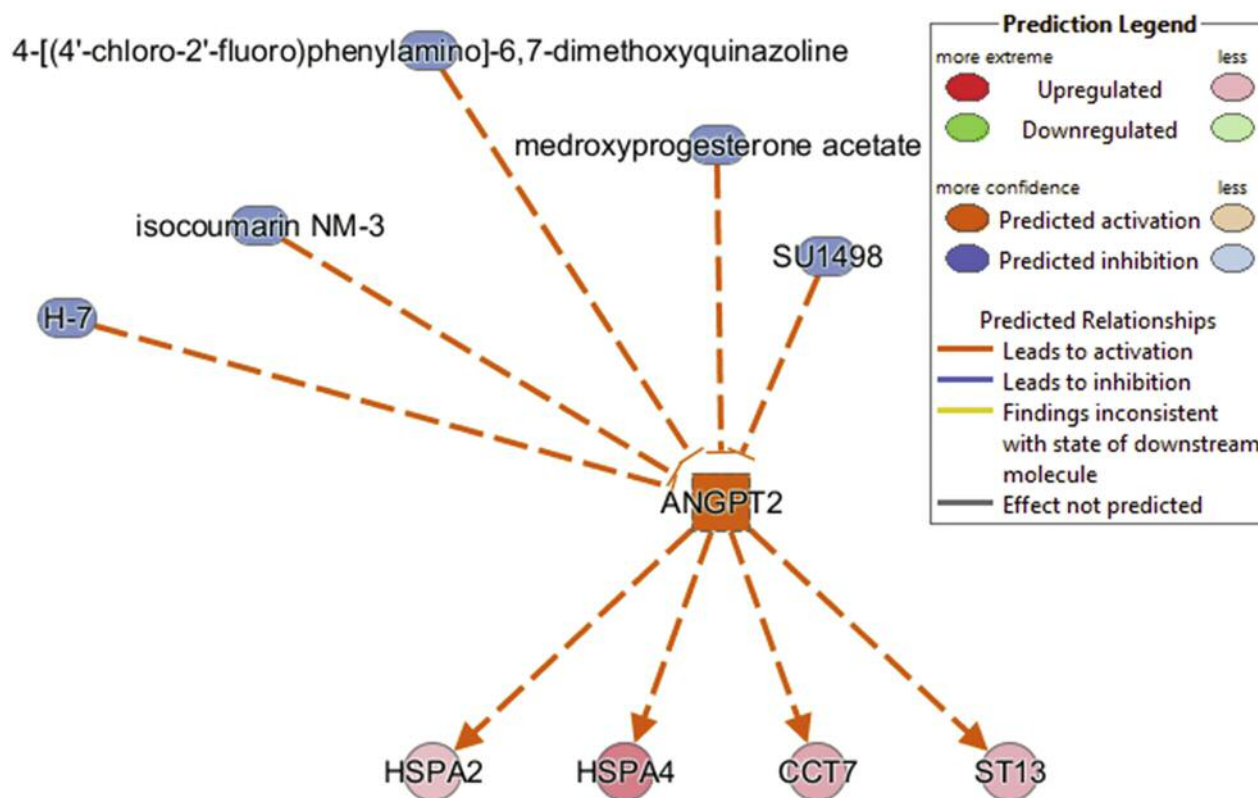
Figure 6. *The upstream regulators of ANGPT2 and its target proteins from Category 1. It can be inhibited by five chemical drugs or reagents from the IPA knowledge database and activated four proteins in Category 1, including CCT7, HSPA2, HSPA4, and ST13.*

apoptotic through binding and inhibition of a pro-apoptotic multi-domain protein (35). DCXR showed a 4.3-fold upgraded change in the follicular adenoma, the second greatest fold change among the five proteins. It is involved in different processes in the organism: regulation of the osmotic state of the cellular environment, detoxification of the cellular environment, cell to cell interaction, *etc*. (36). Among the five proteins, the first four protein have been reported to be involved in cancer. Although with no involvement in cancer, the last one, DCXR, has extensive functions and had a great fold changes between follicular adenoma and normal tissues. According to the proteomics results and the literature, they are all considered important biomarker candidates for monitoring follicular adenoma in clinic management in case that it develops into follicular carcinoma. They are all highly preferred biomarker candidates for further validation.

There are many other proteins listed in Category 1. The other top 10 proteins have been involved in cancers, possess great-fold change, or show good consistency between frequency difference and fold change, making them good candidates for further biomarker validation. Also, some proteins in Categories 2 and 3 had good numbers in these factor values, presenting sound reason for them to be biomarker candidates. Although these proteins were not in Category 1 of prioritization, the top five proteins in Category 2 and the top three proteins in Category 3 are still valuable biomarker candidates. The category and prioritization were only intended to validate protein candidates with the best chance of success in a quicker and less expensive way.

Drug development target candidates for the treatment of follicular adenoma. Using Ingenuity Pathway Analysis, a list of upstream regulators of proteins in each category has been discovered. For proteins in Category 1, 130 upstream regulators show a *p*-value less than 0.01. Among them, eight regulators had an activation z-score. Among the eight regulators, four were activated proteins. Only these four proteins, TGFB1, NFE2L2, MYC, and ANGPT2, were considered drug development target candidates. In theory, any chemical that is able to inhibit the expression of the four proteins has potential to be a drug for the treatment of follicular adenoma to prevent its worsening to follicular carcinoma.

Rank	Upstream Regulator			
	TGFB1	**NFE2L2**	**MYC**	**ANGPT2**
1	CD63	CD63	CD63	CD63
3	**TYMP**	TYMP	TYMP	TYMP
4	**VDAC2**	VDAC2	**VDAC2**	VDAC2
8	**HEXB**	HEXB	HEXB	HEXB
11	**ASPN**	ASPN	ASPN	ASPN
12	HSPA4	HSPA4	HSPA4	**HSPA4**
13	**FBLN5**	FBLN5	**FBLN5**	FBLN5
14	HSPA2	HSPA2	HSPA2	**HSPA2**
15	RPLP0	**RPLP0**	RPLP0	RPLP0
16	PHB2	PHB2	**PHB2**	PHB2
17	PHB	PHB	**PHB**	PHB
18	**FLNB**	FLNB	FLNB	FLNB
19	NPC2	NPC2	NPC2	NPC2
21	H3F3A	H3F3A	**H3F3A**	H3F3A
22	EEF1D	EEF1D	EEF1D	EEF1D
23	HK1	HK1	HK1	HK1
25	PSMD5	**PSMD5**	PSMD5	PSMD5
26	**JUP**	JUP	JUP	JUP
27	**HEBP1**	HEBP1	HEBP1	HEBP1
28	CCT7	**CCT7**	CCT7	**CCT7**
29	PSMD12	**PSMD12**	PSMD12	PSMD12
30	**HSD17B10**	HSD17B10	HSD17B10	HSD17B10
31	**CALD1**	CALD1	CALD1	CALD1
34	ST13	ST13	ST13	**ST13**
37	GOT2	GOT2	**GOT2**	GOT2
40	HPRT1	**HPRT1**	HPRT1	HPRT1
42	**SHMT1**	SHMT1	**SHMT1**	SHMT1
44	**EIF4A3**	EIF4A3	EIF4A3	EIF4A3
47	ESD	**ESD**	ESD	ESD

Figure 7. *Correspondence between the priority of regulators and the rank of proteins targeted by the regulators. The Figure shows that a regulator targeting higher-ranked proteins has a higher candidacy priority as a drug development target, indicating that the ranking of proteins is meaningful.*

TGFB1 is a growth factor and controls proliferation, differentiation, and other functions in many cell types (37). It is an upstream regulator of 12 proteins in Category 1, including ASPN, CALD1, EIF4A3, FBLN5, FLNB, HEBP1, HEXB, HSD17B10, JUP, SHMT1, TYMP, and VDAC2 (Figure 3). According to literature, 10 out of the 12 proteins are activated by TGFB1. The activation effect of six proteins from the 10 proteins has been positively confirmed by IPA, which are shown with an orange line with an arrow at the end. The other four proteins lack literature support to predict the activation effect of TGFB1 to them, which is shown with a gray line with an arrow at the end. From the IPA knowledge database, 10 chemical drugs or reagents have been found to inhibit TGFB1. Any of them can be used in future experiments to test their effect on TGFB1, which will validate the discovery of this study. In the Figure, all the orange lines indicate that the discovery of this study is consistent with and confirmed by literature reports. The more lines with orange color, the more confident the discovery of this study is. Thus, Figure 3 shows TGFB1 is indeed a good drug development target candidate.

NFE2L2 is a transcription regulator and controls the expression of genes for several anti-oxidant enzymes, metal-binding proteins, drug-metabolizing enzymes, drug transporters, and molecular chaperones (38). It activates six proteins in Category 1: CCT7, ESD, HPRT1, PSMD12, PSMD5, and RPLP0 (Figure 4). Based on the expression of these six proteins, IPA predicted that NFE2L2 has been activated in the follicular adenoma tissue. According to the literature, the activation of NRF2 may stimulate the development of de novo cancerous tumors (39), that is consistent with our findings. To inhibit the expression of NFE2L2, two chemical drugs, plicamycin and semaxinib, have been revealed by IPA to decrease the expression of NFE2L2, leading to inhibiting proliferation and tumorigenicity of cancer cells (40, 41). Therefore, NFE2L2 is considered as a drug target candidate.

MYC is a transcription regulator and orchestrates transcriptional regulatory pathways underlying cell growth, cell-cycle progression, metabolism, and survival (42). It is an upstream regulator of seven proteins in Category 1, including FBLN5, GOT2, H3F3A, PHB, PHB2, SHMT1, and VDAC2 (Figure 5). IPA showed that five of the seven proteins were activated by MYC and the activation effect of four of the five was positively confirmed by the literature. Also, IPA revealed six chemical drugs or reagents that were able to inhibit MYC. MYC is overexpressed in many types of cancers and necessary for the rapid proliferation of cancer cells. Strategies aimed at inhibiting MYC have emerged as effective cancer treatments (43). Small molecules targeting MYC are promising in anticancer therapeutics (44) and multiple therapeutic strategies have been developed to inhibit MYC (42). These reports strongly support the candidacy of MYC as a drug development target.

ANGPT2 is another growth factor that plays a role in tumor angiogenesis and growth (45). It activates four proteins in Category 1: CCT7, HSPA2, HSPA4, and ST13 (Figure 6). It can be inhibited by five chemical drugs or reagents from the IPA knowledge database. ANGPT2

signaling has been targeted in cancer therapy (46, 47). It has been studied as a therapeutic target in hepatocellular carcinoma treatment (48) and epithelial ovarian cancer (49). The expression of ANGPT2 is normally low in quiescent mature vessels but is strongly increased in many inflammatory and angiogenic settings. Inhibiting ANGPT2 is to benefit the course of the disease, whether it is airway inflammation, lung injury, or solid tumors (45). According to our results and literature, it should be a therapeutic target for follicular adenoma.

Although all four upstream regulators are considered as strong candidates, their priorities in future experiments for validation and further study should be determined in order to improve the success. Comprehensively, considering the *p*-value, activation z-score, and target molecules in dataset provided by IPA for each regulator, the priority of TGFB1 is much greater than the other three regulators. The priorities of NFE2L2, MYC, and ANGPT2 are similar, but that of NFE2L2 is slightly greater than MYC and MYC is slightly greater than ANGPT2. To inspect the correspondence between the priority of regulators and the rank of proteins targeted by the regulators, a plot is generated by listing regulators in the column and proteins in the row (Figure 7). Figure 7 shows that a regulator targeting higher-ranked proteins has a higher candidacy priority as a drug development target, indicating that the ranking of proteins is meaningful. Based on the discovery, an adjusted priority of the regulators is TGFB1 > MYC > ANGPT2 > NFE2L2.

Conclusion

The objective of the present study was to discover diagnostic biomarker candidates in the clinical management by measuring their levels in plasma, to monitor the state and development of follicular adenoma and find drug development targets for the treatment of follicular adenoma, to prevent its worsening to follicular carcinoma. As a result of the comprehensive mass spectrometry-based analysis used in the study, we identified biomarker and therapeutic target candidates. The comprehensive approach to prioritize the biomarker candidates by categorizing and ranking reveals CD63, DDB1, TYMP, VDAC2, and DCXR as the top five biomarker candidates. Upstream regulator analysis using IPA discovered four therapeutic targets for follicular adenoma. Validation of the biomarker candidates and therapeutic targets in a follow-up study will facilitate monitoring and treatment of follicular adenoma.

Conflicts of Interest

The Authors declare that there exist no conflicts of interests regarding the publication of this paper.

References

1 Nell S, Kist JW, Debray TP, de Keizer B, van Oostenbrugge TJ, Borel Rinkes IH, Valk GD and Vriens MR: Qualitative elastography can replace thyroid nodule fine-needle aspiration in patients with soft thyroid nodules. A systematic review and meta-analysis. Eur J Radiol *84*: 652-661, 2015.

2 Moon WJ, Jung SL, Lee JH, Na DG, Baek JH, Lee YH, Kim J, Kim HS, Byun JS, Lee DH, Thyroid Study Group KSoN, Head and Neck R: Benign and malignant thyroid nodules: US differentiation--multicenter retrospective study. Radiology *247*: 762-770, 2008.

3 American Thyroid Association Guidelines Taskforce on Thyroid N, Differentiated Thyroid C, Cooper DS, Doherty GM, Haugen BR, Kloos RT, Lee SL, Mandel SJ, Mazzaferri EL, McIver B, Pacini F, Schlumberger M, Sherman SI, Steward DL and Tuttle RM: Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. Thyroid *19*: 1167-1214, 2009.

4 Hoang JK, Nguyen XV and Davies L: Overdiagnosis of Thyroid Cancer: Answers to Five Key Questions. Acad Radiol *22*: 1024-1029, 2015.

5 Nikiforova MN and Nikiforov YE: Molecular genetics of thyroid cancer: implications for diagnosis, treatment and prognosis. Expert Rev Mol Diagn *8*: 83-95, 2008.

6 Welker MJ and Orlov D: Thyroid nodules. Am Fam Physician *67*: 559-566, 2003.

7 Lundgren CI, Zedenius J and Skoog L: Fine-needle aspiration biopsy of benign thyroid nodules: an evidence-based review. World J Surg *32*: 1247-1252, 2008.

8 Mathur A, Olson MT and Zeiger MA: Follicular lesions of the thyroid. Surg Clin North Am *94*: 499-513, 2014.

9 Kratky J, Vitkova H, Bartakova J, Telicka Z, Antosova M, Limanova Z and Jiskra J: Thyroid nodules: pathophysiological insight on oncogenesis and novel diagnostic techniques. Physiol Res *63(Suppl 2)*: S263-275, 2014.

10 Trojanowicz B, Sekulla C, Lorenz K, Kohrle J, Finke R, Dralle H and Hoang-Vu C: Proteomic approach reveals novel targets for retinoic acid-mediated therapy of thyroid carcinoma. Mol Cell Endocrinol *325*: 110-117, 2010.

11 Netea-Maier RT, Hunsucker SW, Hoevenaars BM, Helmke SM, Slootweg PJ, Hermus AR, Haugen BR and Duncan MW: Discovery and validation of protein abundance differences between follicular thyroid neoplasms. Cancer Res *68*: 1572-1580, 2008.

12 Braunschweig T, Kaserer K, Chung JY, Bilke S, Krizman D, Knezevic V and Hewitt SM: Proteomic expression profiling of thyroid neoplasms. Proteomics Clin Appl *1*: 264-271, 2007.

13 Sofiadis A, Becker S, Hellman U, Hultin-Rosenberg L, Dinets A, Hulchiy M, Zedenius J, Wallin G, Foukakis T, Hoog A, Auer G, Lehtio J and Larsson C: Proteomic profiling of follicular and papillary thyroid tumors. Eur J Endocrinol *166*: 657-667, 2012.

14 Krause K, Prawitt S, Eszlinger M, Ihling C, Sinz A, Schierle K, Gimm O, Dralle H, Steinert F, Sheu SY, Schmid KW and Fuhrer D: Dissecting molecular events in thyroid neoplasia provides evidence for distinct evolution of follicular thyroid adenoma and carcinoma. Am J Pathol *179*: 3066-3074, 2011.

15 Krause K, Karger S, Schierhorn A, Poncin S, Many MC and Fuhrer D: Proteomic profiling of cold thyroid nodules. Endocrinology *148*: 1754-1763, 2007.

16 Torres-Cabala C, Bibbo M, Panizo-Santos A, Barazi H, Krutzsch H, Roberts DD and Merino MJ: Proteomic identification of new biomarkers and application in thyroid cytology. Acta Cytol *50*: 518-528, 2006.

17 Srisomsap C, Subhasitanont P, Otto A, Mueller EC, Punyarit P, Wittmann-Liebold B and Svasti J: Detection of cathepsin B up-regulation in neoplastic thyroid tissues by proteomic analysis. Proteomics *2*: 706-712, 2002.

18 Abdallah C, Dumas-Gaudot E, Renaut J and Sergeant K: Gel-based and gel-free quantitative proteomics approaches at a glance. Int J Plant Genomics *2012*: 494572, 2012.

19 Lai XY and Witzmann FA: Gel-Based and Gel-Free Sample Preparation for LC-MS/MS Analysis. *In*: Sample Preparation in Biological Mass Spectrometry (Ivanov AR, Lazarev AV eds.): Springer Netherlands, pp. 3-17, 2011.

20 Bradford MM: A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem *72*: 248-254, 1976.

21 Lai X: Reproducible method to enrich membrane proteins with high purity and high yield for an LC-MS/MS approach in quantitative membrane proteomics. Electrophoresis *34*: 809-817, 2013.

22 Keller A, Nesvizhskii AI, Kolker E and Aebersold R: Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem *74*: 5383-5392, 2002.

23 Nesvizhskii AI, Keller A, Kolker E and Aebersold R: A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem *75*: 4646-4658, 2003.

24 Lai X, Wang L, Tang H and Witzmann FA: A novel alignment method and multiple filters for exclusion of unqualified peptides to enhance label-free quantification using peptide intensity in LC-MS/MS. J Proteome Res *10*: 4799-4812, 2011.

25 Lai X and Chen S: Identification of novel biomarker candidates for immunohistochemical diagnosis to distinguish low-grade chondrosarcoma from enchondroma. Proteomics *15*: 2358-2368, 2015.

26 Anderson NL and Anderson NG: The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics *1*: 845-867, 2002.

27 Gupta N and Pevzner PA: False discovery rates of protein identifications: a strike against the two-peptide rule. J Proteome Res *8*: 4173-4181, 2009.

28 Veenstra TD, Conrads TP and Issaq HJ: What to do with "one-hit wonders"? Electrophoresis *25*: 1278-1279, 2004.

29 Cottingham K: Two are not always better than one. J Proteome Res *8*: 4172, 2009.

30 Lupia A, Peppicelli S, Witort E, Bianchini F, Carloni V, Pimpinelli N, Urso C, Borgognoni L, Capaccioli S, Calorini L and Lulli M: CD63 tetraspanin is a negative driver of epithelial-to-mesenchymal transition in human melanoma cells. J Investig Dermatol Symp Proc *134*: 2947-2956, 2014.

31 Fu E, Pan L, Xie Y, Mu D, Liu W, Jin F and Bai X: Tetraspanin CD63 is a regulator of HIV-1 replication. Int J Clin Exp Pathol *8*: 1184-1198, 2015.

32 Iovine B, Iannella ML and Bevilacqua MA: Damage-specific DNA binding protein 1 (DDB1): a protein with a wide range of functions. Int J Biochem Cell Biol *43*: 1664-1667, 2011.

33 Yang Q, Yoshimura G, Mori I, Sakurai T and Kakudo K: Thymidine phosphorylase and breast carcinoma. Anticancer Res *22*: 2355-2360, 2002.

34 Bonotto M, Bozza C, Di Loreto C, Osa EO, Poletto E and Puglisi F: Making capecitabine targeted therapy for breast cancer: which is the role of thymidine phosphorylase? Clin Breast Cancer *13*: 167-172, 2013.

35 Plotz M, Gillissen B, Hossini AM, Daniel PT and Eberle J: Disruption of the VDAC2-Bak interaction by Bcl-x(S) mediates efficient induction of apoptosis in melanoma cells. Cell Death Differ *19*: 1928-1938, 2012.

36 Akintayo A, Legare C and Sullivan R: Dicarbonyl L-xylulose reductase (DCXR), a "moonlighting protein" in the bovine epididymis. PLoS One *10*: e0120869, 2015.

37 Poniatowski LA, Wojdasiewicz P, Gasik R and Szukiewicz D: Transforming growth factor Beta family: insight into the role of growth factors in regulation of fracture healing biology and potential clinical applications. Mediators Inflamm *2015*: 137823, 2015.

38 Merikallio H, Paakko P, Kinnula VL, Harju T and Soini Y: Nuclear factor erythroid-derived 2-like 2 (Nrf2) and DJ1 are prognostic factors in lung cancer. Hum Pathol *43*: 577-584, 2012.

39 DeNicola GM, Karreth FA, Humpton TJ, Gopinathan A, Wei C, Frese K, Mangal D, Yu KH, Yeo CJ, Calhoun ES, Scrimieri F, Winter JM, Hruban RH, Iacobuzio-Donahue C, Kern SE, Blair IA and Tuveson DA: Oncogene-induced Nrf2 transcription promotes ROS detoxification and tumorigenesis. Nature *475*: 106-109, 2011.

40 Zhang M, Mathur A, Zhang Y, Xi S, Atay S, Hong JA, Datrice N, Upham T, Kemp CD, Ripley RT, Wiegand G, Avital I, Fetsch P, Mani H, Zlott D, Robey R, Bates SE, Li X, Rao M and Schrump DS: Mithramycin represses basal and cigarette smoke-induced expression of ABCG2 and inhibits stem cell signaling in lung and esophageal cancer cells. Cancer Res *72*: 4178-4192, 2012.

41 Bogaard HJ, Natarajan R, Henderson SC, Long CS, Kraskauskas D, Smithson L, Ockaili R, McCord JM and Voelkel NF: Chronic pulmonary artery pressure elevation is insufficient to explain right heart failure. Circulation *120*: 1951-1960, 2009.

42 McKeown MR and Bradner JE: Therapeutic strategies to inhibit MYC. Cold Spring Harb Perspect Med 42014.

43 Huang H, Weng H, Zhou H and Qu L: Attacking c-Myc: targeted and combined therapies for cancer. Curr Pharm Des *20*: 6543-6554, 2014.

44 Chen BJ, Wu YL, Tanaka Y and Zhang W: Small molecules targeting c-Myc oncogene: promising anti-cancer therapeutics. Int J BiolSci *10*: 1084-1096, 2014.

45 Thurston G and Daly C: The complex role of angiopoietin-2 in the angiopoietin-tie signaling pathway. Cold Spring Harb Perspect Med 2: a006550, 2012.

46 Eroglu Z, Stein CA and Pal SK: Targeting angiopoietin-2 signaling in cancer therapy. Expert Opin Investig Drugs 22: 813-825, 2013.

47 Gerald D, Chintharlapalli S, Augustin HG and Benjamin LE: Angiopoietin-2: an attractive target for improved antiangiogenic tumor therapy. Cancer Res 73: 1649-1657, 2013.

48 Bupathi M, Kaseb A and Janku F: Angiopoietin 2 as a therapeutic target in hepatocellular carcinoma treatment: current perspectives. Onco Targets Ther *7*: 1927-1932, 2014.

49 Secord AA, Nixon AB and Hurwitz HI: The search for biomarkers to direct antiangiogenic treatment in epithelial ovarian cancer. Gynecol Oncol *135*: 349-358, 2014.