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

XIANYIN LAI1,2* and SHAOXIONG CHEN3*

1Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, U.S.A.; 2Department of Cellular & Integrative Physiology, Indiana University School of Medicine, Indianapolis, IN, U.S.A.; 3Department 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 TGFβ1, 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.

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 \textit{BRAF}, \textit{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

\textbf{Materials.} Urea, DL-dithiothreitol, triethylphosphine, iodoethanol, and ammonium bicarbonate (NH\textsubscript{4}HCO\textsubscript{3}) were purchased from Sigma-Aldrich (St. Louis, MO, USA). LC-MS grade water (H\textsubscript{2}O), LC-MS grade 0.1% formic acid in acetonitrile (ACN), and 0.1% formic acid in water (H\textsubscript{2}O) were obtained from Burdick & Jackson (Muskegon, MI, USA). Modified sequencing grade porcine trypsin was purchased from Princeton Separations (Freehold, NJ, USA).

\textbf{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).

\textbf{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\textsubscript{4}HCO\textsubscript{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.

\textbf{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).

\textbf{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).
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).

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 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.
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.
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 \[ FD \times 10 + FC \times 10 + FD \times FC \times 10 + NP + SS + SD + NC \], where FD, FC, FD \times 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 upgraded 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, deoxyguuridine, 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 3. The upstream regulators of TGFβ1 and its target proteins from Category 1. Chemical drugs or reagents highlighted in blue inhibit TGFβ1, while TGFβ1 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 TGFβ1 is increased as well, but the amount of the chemicals should be decreased. If more chemicals are applied, the expression of TGFβ1 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 TGFβ1. 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 TGFβ1.
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.
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.
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, 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 TGFβ1 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 TGFβ1 > 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


