공학석사 학위논문

Quantitative proteome profiling of well-differentiated thyroid cancer and anaplastic thyroid cancer using isobaric labeling

분화갑상선암과 비분화갑상선암의 중동체 라벨을 이용한 프로테옴 프로파일링 연구

2019 년 2 월

서울대학교 대학원

협동과정 바이오엔지니어링 전공

왕 인 재

분화갑상선암과 비분화갑상선암의 중동체 라벨을 이용한 프로테옴 프로파일링 연구

지도교수 김 영 수

이 논문을 공학석사 학위논문으로 제출함

2019 년 1 월

서울대학교 대학원
협동과정 바이오엔지니어링 전공
왕 인 재

왕인재의 공학석사 학위논문을 인준함
2019 년 1 월

위 원 장 _____ (인)
부위원장 _____ (인)
위 원 _____ (인)

Quantitative proteome profiling of well-differentiated thyroid cancer and anaplastic thyroid cancer using isobaric labeling

by
Joseph Injae Wang

Thesis
Submitted to the Faculty of Graduate School of the Seoul National University in partial fulfilment of the requirements for the degree of Master of Engineering
in Bioengineering

Approved: January 2019

Chair _____(Seal)
Vice Chair _____(Seal)
Examiner _____(Seal)

Abstract

Joseph Injae Wang
Interdisciplinary Program for Bioengineering
The Graduate School
Seoul National University

Thyroid cancer is the most common endocrine cancers that is expected to see more than 56,000 new cases in the United States in 2018. Annually, thyroid cancer claims more than 2,000 lives in the United States alone. However, despite its prevalence, the high-survival rate has caused research into thyroid cancer to stagnate compared to other cancers as it is deemed relatively innocuous. Consequently, the thyroid cancer proteome remains largely unexplored despite identifying oncogenes and their associated mutations at the mRNA level via microarrays. Since benign types of thyroid cancer have the propensity to devolve into malignant forms, it is imperative to profile the thyroid cancer proteome to achieve a comprehensive understanding of the disease. In this study, both differentiated and undifferentiated variants of thyroid carcinoma were studied using mass spectrometry. Samples obtained from 14 patients were analyzed using an Easy-nLC 1000 coupled with a Q-Exactive mass spectrometer. A total of 7071 proteins were identified, of which 6215 were quantifiable. EIF2 and Rac Signaling Pathway were found to be significantly altered as WDTCs progress into ATCs in RAS and BRAF, respectively.

Keyword : Thyroid cancer, Proteomics, Mass Spectrometry

Student Number : 2017-25653

Table of Contents

Chapter 1. Introduction **1**

Chapter 2. Methods **7**

Chapter 3. Results **13**

Chapter 4. Discussion **41**

Chapter 5. Conclusion **44**

References **46**

Abstract in Korean **49**

List of Figures

Figure 1. Experiment design.....3

Figure 2. Overall workflow of the experiment Introduction 8

Figure 3. Overview of identified & quantified proteins 14

Figure 4. Comparison with past studies & existing literature 15

Figure 5. Reproducibility of analysis 17

Figure 6. Thyroid Differentiation Score 19

Figure 7. Hierachical clustering of experimental groups.....21-23

Figure 8. IPA analysis of BRAF cluster 1 26-27

Figure 9. IPA analysis of BRAF cluster 2 28-29

Figure 10. IPA analysis of RAS cluster 1.....30-31

Figure 11. IPA analysis of RAS cluster 2.....32-33

Figure 12. IPA analysis of RAS cluster 3.....34-35

Figure 13. EIF2 Signaling pathway.....37-38

Figure 14. Rac Signaling pathway.....39-40

List of Tables

Table 1. Sample information 7

Table 2. Top 5 significant terms from IPA organized by cluster..... 36

Introduction

Thyroid Cancer

Thyroid cancer is the most common endocrine cancers that is expected to see more than 53,000 new cases in the United States in 2018. Annually, thyroid cancer claims more than 2,000 lives in the United States alone. [1] Despite its prevalence and risk, the high-survival rate of benign cases has perpetuated an erroneous perception that thyroid cancer is relatively innocuous. In turn, this false sense of security has caused research into thyroid cancer to stagnate and thyroid cancer remains the least funded type of cancer. Consequently, comparatively few studies have investigated the thyroid cancer proteome and a significant portion of it remains unexplored.

Although the exact molecular mechanisms of onco-genesis remain unknown, the wide array of associated driver-mutations confers distinct Clusters of gene expression, which in turn manifest varying histopathological characteristics. [2] Consequently, accurate subtyping is crucial to proper prognosis as it dictates which treatments are effective. Typically, the differentiation degree of thyroid cancer is inversely correlated with aggressiveness of cancer and patient outcome. [3] Thus, thyroid cancers are often classified according to their differentiation status. Well-differentiated thyroid carcinoma (WDTC) arises from thyroid follicular epithelium and comprises the vast majority of thyroid carcinoma cases. The two main forms of follicular cancers include Papillary Thyroid Cancer (PTC) and Follicular Thyroid Cancer (FTC), which accounts for approximately 80-84% and 6-10% of all thyroid carcinomas, respectively. [4] WDTCs rarely metastasizes and are typically responsive to resection and radioactive iodine (RAI) treatment with the exception of

BRAF mutated PTCs, which are often unresponsive. [5] Typically, the iodine absorbing properties of follicular cells is retained in carcinoma and exposure to RAI effectively destroys tumors while sparing the surrounding region. [4] In recent years, it was discovered that the genes involved in iodine uptake are suppressed as MPAK-pathway output is upregulated by BRAF V600E oncogene in PTC. [6] Poorly differentiated Thyroid Carcinoma (PDTC) and Anaplastic Thyroid Cancer (ATC) are comparatively rare tumors that also arise from follicular cells and are highly aggressive. ATC accounts for less than 1% of all thyroid carcinomas and is characterized by aggressive metastasis and high mortality rate as evidenced by the average survival period of 3-4 months after diagnosis. [1] However, in recent works, it is suggested that PDTCs (poorly differentiated thyroid carcinoma) and undifferentiated thyroid carcinoma more often arise from pre-existing WDTC rather occurring in *de novo*. [7] With the knowledge that benign thyroid cancers have the propensity to devolve into malignant forms, it is imperative to profile the thyroid cancer proteome achieve a comprehensive understanding of the disease.

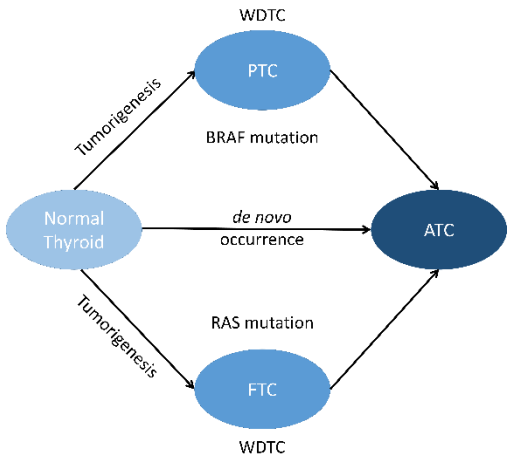


Figure 1: Simple visual representation of experiment design. ATC rarely occurs de novo from normal tissue. Instead, ATC generally progresses from WDTCs such as PTC or FTC. The vast majority of PTCs are typically driven by BRAF mutations while FTCs are generally driven by RAS mutations.

Key Oncogenes & Associated Mutations

Past studies have identified several key targets and their associated mutations such as BRAF, RAS, and TP 53 at the genome and transcriptome level. [2] In particular, BRAF is notable in that $BRAF^{V600E}$ mutants are present in approximately 45% of PTC cases. Numerous studies unvaryingly support that BRAF mutations correlate with poor clinicopathological patient outcome - increase in aggressive pathological features and increased recurrence. [2, 8] At the present, it is unclear whether BRAF mutation initiates PTC tumorigenesis or is a result of it. Only second to BRAF in prevalence is the RAS mutation. In particular, among the various isoforms of RAS, NRAS is predominantly found in RAS mutants. While RAS can activate both MAPK and PI3K-AKT pathway, the latter is upregulated in thyroid tumorigenesis. [8] However, as mRNA expression often deviates from protein

expression, it is necessary to investigate these pathways at the protein level as well.

Application of mass spectrometry in proteomics studies

Due to their higher sensitivity and ease of access to pre-made kits and protocols, immunoassays such as ELISA remain the gold standard amongst researchers when quantifying a specific protein. However, despite their widespread use, the inability to effectively mitigate cross-reactivity between reporter antibodies can result in inaccurate or even false results when it comes to multiplexed protein assay. These demerits severely limit their application in multiplexed assays, where there is a growing need for omics research. In this regard, mass spectrometry provides a significant advantage over immunoassays in that it can quantify thousands of proteins in a single run. In addition to an increase in coverage, proteins are detected based on mass to charge (m/z) ratio, which makes the detection more objective as targets are not selected prior to analysis as in the case with immunoassays. In light of these advantages, I employed a standardized liquid chromatography and mass spectrometry techniques with report ions to simultaneously quantify expression profiles of various thyroid carcinoma.

Previous proteomic studies of thyroid cancer

In previous works, investigators have primarily focused on profiling the proteome of common thyroid cancers such as PTC, FTC, or follicular adenoma. While prevalent, these diseases are often survivable and usually reach potency when developed into deadly and malignant cancers, such as ATC. As shown in the works of Uyy et al. and Ban et al, the composition of sample groups is deficient or limited to common forms, such as FA or PTC. [9, 10] Even in the works of Martinez-Aguilar

et al., ATC, the chief culprit of thyroid cancer deaths, is omitted. [11, 12] In contrast, a study conducted by Gawin et al. used a wide variety of samples, including ATC. [13] However, without proper grouping of samples, their study produces cursory analysis that fails to target the core issue of thyroid cancer – its propensity to devolve into deadly malignancy. Furthermore, the vast majority of these works are performed in gels or in label-free. For methods such as 2D-PAGE, proteins with extreme qualities in size, acidity or hydrophobicity poorly represented, limiting the functional range of identifiable proteins. [14]

Significance of labeling in quantitative proteomics

The wide array of labelling techniques used in mass spectrometry can be divided into two major categories - *in vivo* and *in vitro* labelling. For methods such as SILAC (Stable Isotope Labeling by Amino acids in Cell culture) or SILAM (Stable Isotope Labeling by Amino acids in Mammals), i*n vivo* labeling is accomplished by heavy isotopes are metabolically incorporated into the organism. Consequently, the labeling efficiency often varies and the requirement of a controlled diet makes it unfeasible for clinical samples. In contrast, *in vitro* labeling methods feature uniform labeling efficacy, which makes it more suitable for clinical studies. Isotope-coded-affinity-tag (ICAT), which uses chemically identical probes with distinct masses, was the first to introduce the concept of *in vitro* labelling. However, despite its novelty at its inception, it was discovered that ICAT-deuterated peptides elute earlier than its counterpart when separated in a reverse-phase column. This shift in retention time was addressed with the development of isobaric tags, which does not alter the retention time of eluents. [15] One such isobaric tag is tandem mass tag (TMT), which has been continually developed since its introduction

in 2003. [16] As of 2013, TMT now supports up to 10 channels, greatly increasing its quantification capabilities. [17] By substituting isotopes in its mass-normalization group and reporter ion region, a mass difference of 00.63 Da retained and subsequently resolved in MS^2. This increase in capacity is relevant as it eliminates run-to-run variation that result from having multiple datasets. In light of previous works, this study aims to produce a credible dataset by using isobaric tags and analyse the progression in how less malignant forms devolve into lethal cancers.

Methods

Reagents

All chemical reagents were purchased from Sigma-Aldrich (St. Louis, MO, USA) unless specified otherwise. Sequencing grade trypsin was purchased from Promega. Oasis HLB sorbents were purchased from Waters Corporation (Milford, MA, USA). BCA Kit and TMT10plex™ Isobaric Label Reagent Set was purchased from Thermo Fischer Scientific (Waltham, MA, USA). Corning® Costar® Spin-X® Plastic Centrifuge Tube Filters were purchased from Merck KGaA (Darmstadt, Germany).

Table 1. Sample information including subtype, age, sex, tumor origin, and mutations.

Sample Designation	Subtype	Age	Sex	Tumor origin	Oncogene Mutation
PRO_C1	normal thyroid				
PRO_C2	normal thyroid				
PRO_C3	normal thyroid				
PRO_C4	normal thyroid				
PRO1	Focal ATC/PDTC (10% PD)	63	Female	PTC	BRAF (V600E)
PRO2	Focal ATC/PDTC (10% PD)	91	Female	PTC	BRAF (V600E)
PRO3	Focal ATC/PDTC (5% ATC)	75	Female	PTC	BRAF (V600E)
PRO4	ATC	65	Female	PTC	BRAF (V600E)
PRO5	ATC	55	Female	PTC	BRAF (V600E)
PRO6	ATC	71	Female	PTC	BRAF (V600E), PIK3CA (H1047R)
PRO7	wiFTC	68	Male	FTC	NRAS (Q61K)
PRO8	wiFTC	76	Female	FTC	NRAS (Q61R)
PRO9	wiFTC	59	Male	FTC	NRAS (Q61R)
PRO10	ATC	68	Male	FTC	HRAS (G13R), AKT1 (E17K), EIF1AX (c.338-2A>G, splicing)
PRO11	ATC	54	Female	FTC	NRAS (Q61R), EIF1AX (G8E)
PRO12	ATC	55	Female	FTC	NRAS (Q61R), EIF1AX (G8E)
PRO13	miFTC				Oncogene mutation
PRO14	miFTC				

Abbreviations: Anaplastic Thyroid Cancer (ATC); widely invasive Follicular Thyroid Cancer (wiFTC); minimally invasive Follicular Thyroid Cancer (miFTC) ,

Thyroid Tissue Samples

Thyroid tissue samples were obtained from Seoul National University Hospital. Patients samples were collected in accordance with Institutional Review Board guidelines (IRB 1802-067-922). The cohort consisted of 4 normal tissue, 3 PTC, 3 FTC, and 6 ATC samples (Table 1).

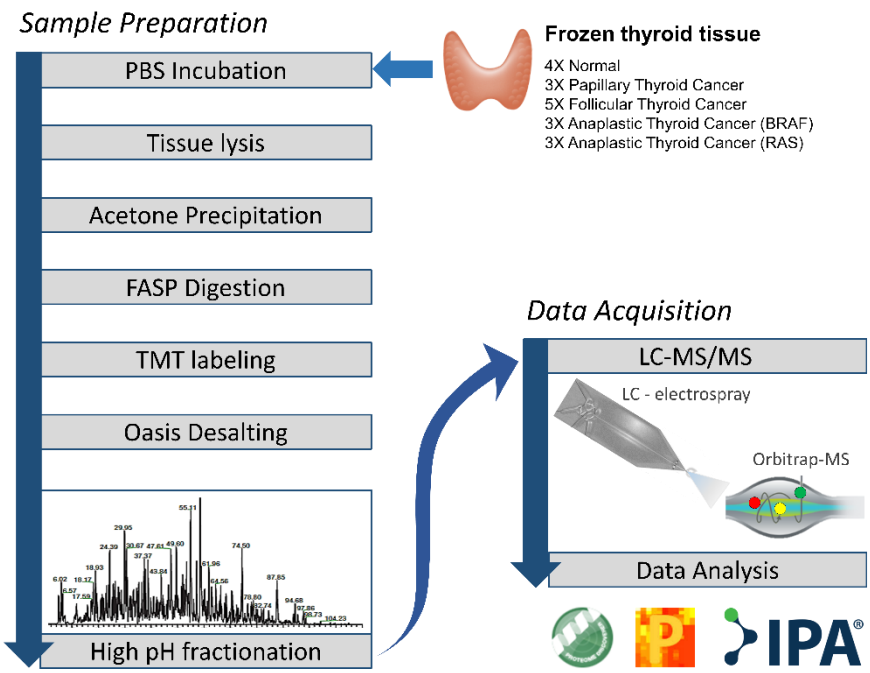


Figure 2. Overall workflow of the experiment. Peptide samples were extracted form 18 frozen thyroid tissues. Samples were analyzed via LC-MS.

Preparation of Samples

The frozen thyroid tissue samples (Total n=18) were transferred into Eppendorf tubes prior to overnight incubation in ice cold PBS at 4 °C. The supernatant was discarded and the samples were sonicated (28% amplitude) in

varying volumes of lysis buffer (4% SDS, 2mmol TCEP, 0.1 mol 0.1 M Tris–Cl, pH 7.4). The lysed samples were then boiled for 30 minutes at 100 °C and filtered using Spin-X plastic centrifuge tubes (Merck). Protein concentrations were measured by BCA kit (Thermo Fisher Scientific). The concentrations were used to precipitate 140 µg of starting amount by adding cold acetone (Thermo Fischer Scientific). at a ratio of 5:1 (v/v). The samples were incubated overnight at -20 °C. The samples were pelleted at 15,000 rpm and washed again with 500 µL acetone. After discarding the supernatant, the pellet was air-dried for 2 hours and reconstituted in a buffer (2% SDS, 0.1 M DTT, 0.1 M Tris–Cl, pH 7.4) and boiled for 30 minutes at 100 °C.

Protein Digestion

Protein samples were purified and digested following FASP protocols that are described in detail in previous works. [18, 19] To briefly summarize, the dissolved samples were transferred to 30-kDa cutoff-filters (Amicon® Ultra, Millipore, USA) loaded with 300 µL urea (8 M urea, Merck, USA). and centrifuged (14,000g, 15 min, 20 °C). The initial wash was followed by two additional urea buffer washes (300 µL) under the same conditions. Excess eluents were discarded between washes. The filters were loaded with 200 µL IAA (50 mM IAA, 8 M urea, Tris-Cl, pH 8.5) solution and incubated for 45 minutes in dark room temperature. After incubation, the filters were centrifuged once under the same condition without adding reagents. Then, the filters were washed twice with 300 µL urea buffer (14,000g, 15 min, 20 °C) and then washed three times with 300 µL 40 *mM* triethyl ammonium bicarbonate (TEAB) under identical conditions. After the wash, the filter-units containing the protein were transferred into new tube. After loading 100 µL TEAB buffer to each filter, sequencing-grade trypsin (0.1 µg/µL) was added at a

ratio of 50:1 (*w*/*w*) of the initial protein amount. The samples were incubated in a shaker at 37 °C overnight (18 hr). Samples were eluted in three centrifuge cycles: once without additional liquid (14,000g, 15 min, 20 °C) followed by a 100 µL TEAB wash (14,000g, 15 min, 20 °C), and finally with 50 µL NaCl (0.5 *M*) wash (14,000g, 15 min, 20 °C). Eluted peptides were transferred to Eppendorf tubes and their concentrations were measured by tryptophan fluorescence assay. [20]

Tandem Mass Tag Labelling

Since the number of samples (n=18) exceeded the capacity of a single TMT 10-Plex kit, the samples were distributed between two 10-Plex kits with each set containing 9 samples and 1 pooled sample as control. The pooled sample was created by mixing 4.4 ug portions from each sample. The mixed sample was then placed at the 10th channel of each TMT set to serve as inter-set control. Samples were assigned to channels via randomization using the functions of Excel in order to avoid potential bias. 40 ug aliquots of each peptide sample were prepared for tandem mass tag labelling. The differing sample volumes were matched to the highest volume by adding 40 m*M* TEAB buffer. Equal volumes of TMT reagent and ovalbumin were added to each sample along with ACN to reach a final concentration of 30%. Samples of each set were combined into a 5 mL tube and incubated for 90 minutes. The reaction was subsequently quenched by adding 13 µL of 0.3% quenching solution to the two tubes and snap-frozen at -80 °C.

Solid-Phase Extraction (SPE) & High pH Fractionation

The labeled peptides were desalted using HLB Oasis columns (Waters). The manufacturer's manual was followed and lyophilized in a SpeedVac. The dried

samples were subsequently reconstituted with ACN (15 m*M* ammonium formate). Samples were fractionated offline using a 1260 Infinity II Bio-Inert LC system (Agilent, Santa Clara, CA). The fractions were eluted onto a 96-well plate along a 5-40% ACN gradient over 60 minutes. The 96 fractions were concatenated by row to produce a total of 12 vials per TMT set. The vials were then lyophilized in a SpeedVac and stored at -80 °C.

LC-MS/MS Analysis

Prior to analysis, the samples were dissolved in Solvent A (2% ACN, 0.1% formic acid). A total of 24 vials were analysed by a set-up consisting of an Easy-nLC 1000 (Thermo Fisher Scientific, Waltham, MA) attached with a nano-electrospray ion source (Thermo Fisher Scientific, Waltham, MA) coupled with a Q Exactive mass spectrometer (Thermo Fisher Scientific, Waltham, MA). All samples were analysed in duplicate under the following conditions: 240 minutes non-linear gradient ranging from 8% to 60% Solvent B (0.1% formic acid in ACN), spray voltage of 2.2 kV in positive ion mode, heated capillary temperature of 320°C, data-dependent acquisition mode with top 15, precursor ions within 300 – 1,650 m/z, resolution of 70,000 at 200 m/z, automatic gain control (AGC) target value of 3 x 10^6, isolation window of 1.2 m/z, HCD scan resolution of 35,000, and normalized collision energy (NCE) of 32. [21]

Data Processing

Raw MS data was searched in Proteome Discoverer (version 2.2.0.338) (Thermo Fisher Scientific). SEQUEST HT algorithm was used against the Uniprot database with the following parameters: up to 2 missed cleavages, minimum peptide

length of 6, maximum peptide length of 144, precursor mass tolerance of 10 ppm, and fragment mass tolerance of 0.02 Da. Static modifications included carboxymethylation (C) and TMT 6-plex at N-terminus lysine residues along with dynamic modifications of methionine oxidation and deamidation at N-terminus. All statistical tests were conducted in Perseus (1.6.2.2) and Excel (2016). [22] Pathway analysis was performed using Ingenuity Pathway Analysis (version 01-04) by Qiagen (Venlo, Netherlands).

Statistical analysis

For hierarchical clustering, k value of 300 for k-means pre-processing and average linkage were used. Significance level was set to 0.05 (a=0.05) for ANOVA. Subsequently, fold changes of 1.25 for upregulation and 0.8 for down regulation were applied to the list of DEPs to generate the final candidates due for analysis via IPA. For integrative analysis, terms were filter based on two categories: activation score and Fisher's exact test. Only terms whose activation scores were greater than 1 in magnitude were retained. Likewise, a p value of 0.05 was applied to filter statistically insignificant terms.

Results

Overview

The surgically resected thyroid tissues were classified into 5 histological groups; Healthy normal, PTCs with BRAF mutation, FTCs with RAS mutation, and ATC. The frozen tissue samples were processed using a previously established TMT-based quantification method. The resulting peptide samples were analysed via LC-MS/MS and the raw data was processed using Proteome Discoverer. Applying a FDR of 1%, 7071 proteins were identified, of which 6215 were quantified (Figure 3A). For peptides, 89859 groups were identified. As evidenced by the large overlap shown in Fig. 3B, the vast majority of identified proteins were common to all histological groups. Only a single protein was uniquely identified between PTC and FTC as opposed to the 6 proteins for ATC. Proteins were deemed to be quantified if detected in all 18 samples. The number of quantified proteins was relatively uniform across biological replicates between 6000 and 7000 (Figure 3C). The low variation in the number of quantified proteins across channels is further supported by the low Coefficient of Variation (CV) value of 1.19% (Figure 3C). Thus, it was possible to cursorily confirm labeling efficiency was comparable across channels. In 2015, Liu et al.'s work produced the largest and most comprehensive proteomic dataset of the thyroid (Figure 4B). When comparing the quantified proteins, more than 70% of Liu et al.'s dataset (3097 proteins), was encompassed by this dataset (Figure 4A), making it the largest proteome dataset to my knowledge. [23]

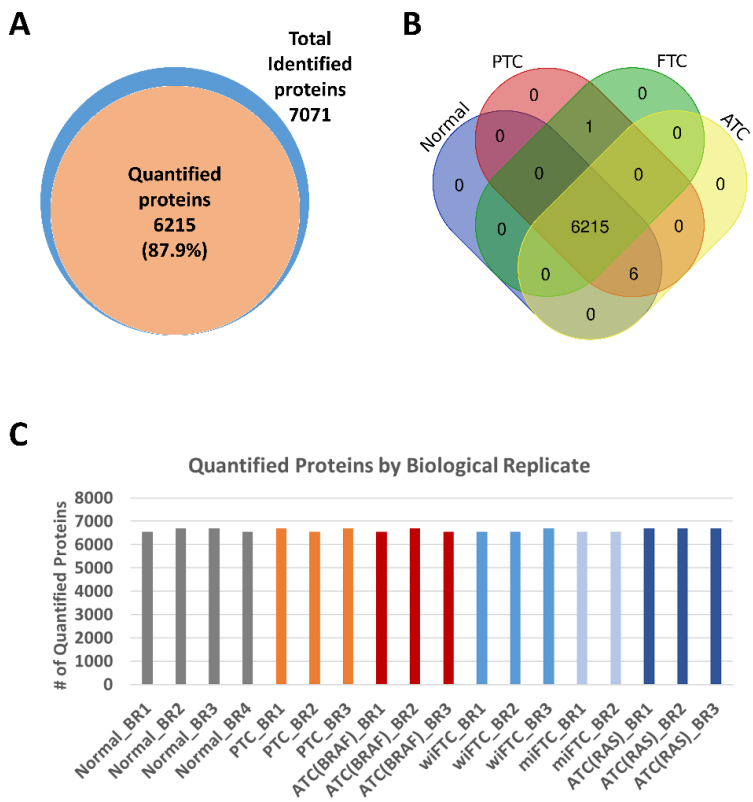


Figure 3. (**A)** Venn diagram of identified proteins and quantified proteins. Proteins were considered to be identified if it were detected with high confidence and did not register in the contaminants list. Proteins were deemed to be quantified only if detected in all 18 biological replicates. (**B)** Comparison of quantified proteins in each biological group (Normal, PTC, FTC, and ATC). Proteins were deemed to be quantified if detected in every biological replicate. The vast majority (99.9%) of proteins were detected in all biological replicates. (**C)** Graphical representation of quantified proteins in all samples. The number of identified proteins remains stable across replicates and MS runs as evidenced by the low Coefficient of Variation value of 1.19%.

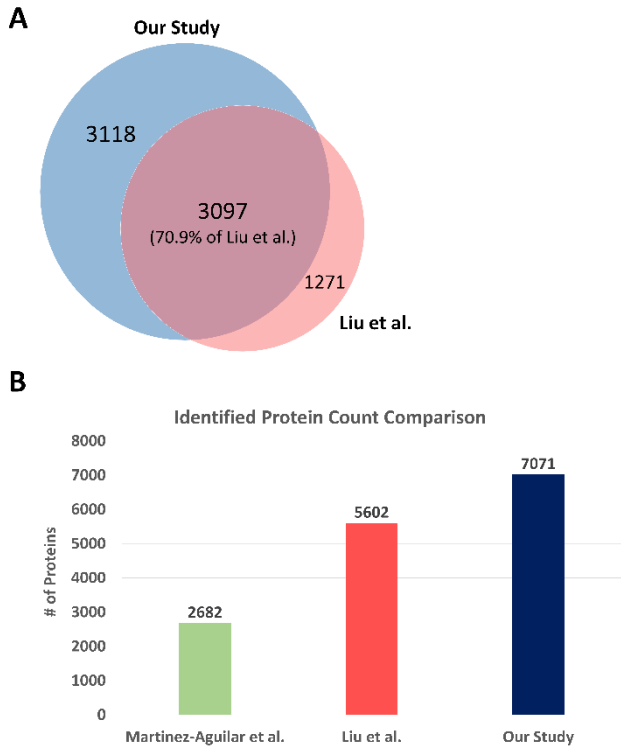


Figure 4. (A) Venn diagram of quantified proteins in this study and that of Liu et al. Over 70% of Liu et al.'s dataset overlapped. **(B)** Comparison of identified proteins in LC-MS/MS based studies of thyroid proteome. More than 7,000 proteins were identified, surpassing Liu et al.'s dataset by more than 1,000 proteins.

Assessment of reproducibility

Inter- and intra-batch abundances were normalized based on ovalbumin, a non-homologous external standard, or total abundance using Proteome Discoverer's built-in function. The two normalization methods were then assessed by comparing the two resulting datasets. For both datasets, a CV value was calculated for each protein identified in all technical replicates. Then, the distribution of CV values was compared across biological groups by comparing the medians. The average of medians was lower when normalized by total abundance (26.89%) than by ovalbumin (31.85%). Thus, the raw abundances were normalized based on total abundance. After normalization, the abundance of detected ovalbumin verified that a similar amount was injected to each sample (Figure 5C). The box and whisker plots (Figure 5A & 5B) show a decrease in the median CV values of each sample group after normalization. Each biological group experienced an average reduction in CV value by a margin of ~16.3%. Plotting the abundances of the two TMT sets against one another showed that the results were highly reproducible as supported by the high coefficient of correlation value of 0.99 (Figure 5D). Therefore, the quantitative results of two TMT sets can be joined without compromising the validity and quality of the dataset. Furthermore, it is apparent that the differences of protein expression among thyroid cancer types originate from physiological differences, rather than technical variations.

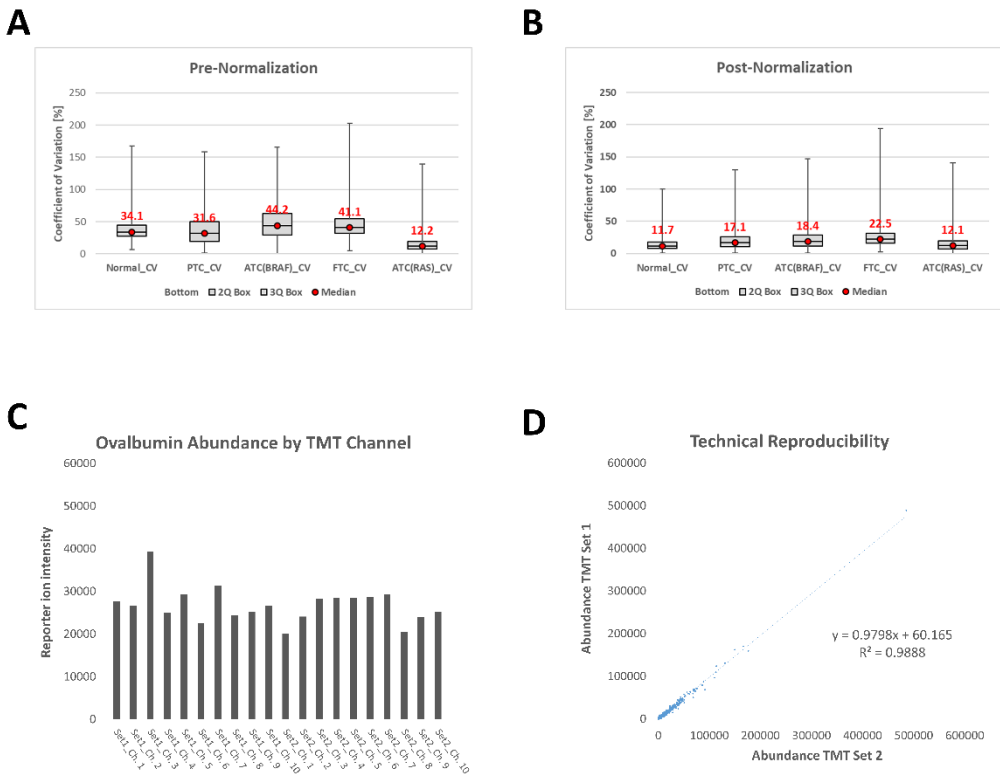
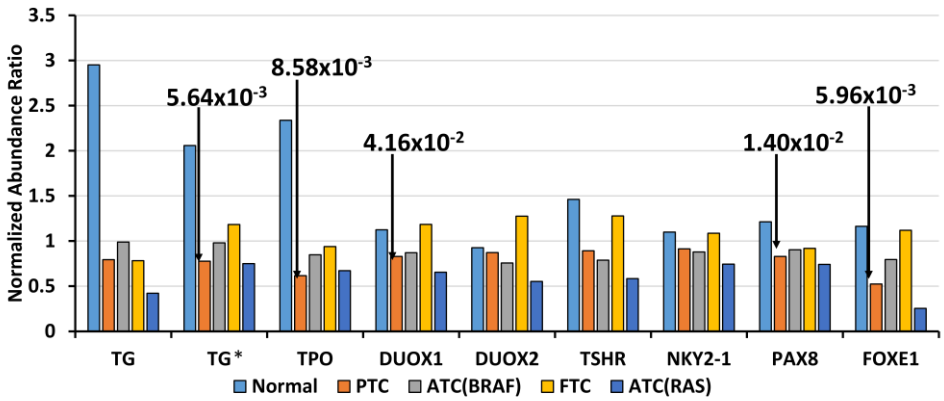


Figure 5. (A) & **(B)** Box and whiskers plot of CV values of quantified proteins within histological groups with medians annotated in red. Median CV decrease drastically post-normalization. (**C**) With the exception of Set 1 Channel 3, all reporter ion intensities of ovalbumin generally ranged between 26,000 to 30,0000. (**D**) Scatterplot of abundance values of TMT Set 1 & Set 2 on each axis. High coefficient of determination can be observed between the two TMT sets.

Thyroid Differentiation Score

In a landmark study of integrated genomics of PTC by Agrawal et al., thyroid differentiation degree was represented by a single metric known as thyroid differentiation score (TDS). In their study, thyroid differentiation score was calculated by averaging the log2-normalized fold change values of 16 key genes related to thyroid function and metabolism. [24] When cross-referenced, 8 of the aforementioned 16 genes were identified in the dataset. Each gene had one corresponding protein with the exception of TG, which produced thyroglobulin as well as its isoform. When averaged by histological group, TG, TPO, DUOX1, PAX8, and FOXE1 passed a student's t-test between normal group and PTC group. Following Agrawal et al.'s method, TDS was calculated for each sample in this study (Figure 6B). On the whole the Normal sample had the highest TDS scores while ATC(RAS) had the lowest TDS scores.

A



B

Sample	Normal				PTC			ATC (BRAF)			miFTC		wiFTC			ATC (RAS)		
BR	BR1	BR2	BR3	BR4	BR1	BR2	BR3	BR1	BR2	BR3	BR1	BR2	BR1	BR2	BR3	BR1	BR2	BR3
TDS	0.691	0.757	0.704	0.824	0.074	-0.103	-0.611	-0.229	-0.569	0.467	0.589	0.016	0.266	0.214	0.042	-0.632	-0.493	-0.694

Figure 6. (A) Bar graph representation of normalized abundance ratio of key gene-derived proteins averaged by histological groups. Arrows indicate significance level (p=0.05) by student's t-test between Normal group and PTC. Asterisk denotes isoform in TG*. (**B)** Calculated TDS values based on the 8 genes common to Agrawal et al.'s list. For the purpose of statistical analysis, miFTC and wiFTC were treated as a singular group.

Hierarchical clustering

On the whole, clustering of the quantified proteins grouped demonstrated that the expression profiles within histological groups were similar. All samples were adjacent to at least one other sample within the same group. Interestingly, ATC(BRAF) group and ATC(RAS) group were clustered apart from one another despite sharing some upregulated proteins (Figure 7A). Two distinct patterns were observed with a cluster of upregulated proteins in Normal group and another cluster of upregulated proteins in ATC group. For the ATC-upregulated cluster, it is notable that the upregulation is more distinct in ATC(RAS) group than in ATC(BRAF) group. When analyzed by DAVID 6.8, it was found that the most enriched biological process in the Normal-upregulated cluster was Oxygen transport (160-fold) followed by Hydrogen peroxide catabolic process (120-fold). In the ATC-upregulated cluster, the majority of the terms were related to protein folding and assembly. The top 2 most enriched terms were Cytosol to ER transport (357-fold) followed by Negative regulation of post-translational protein modification (238-fold).

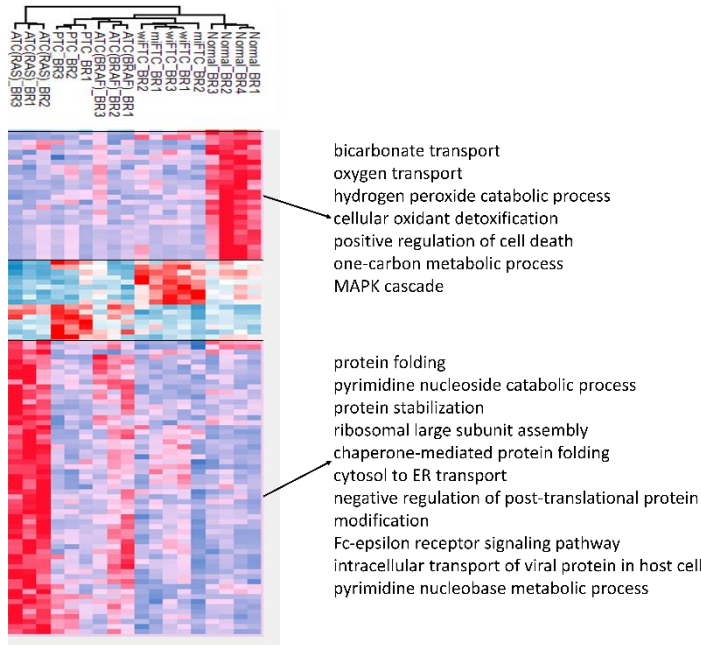


Figure 7A. Hierarchical clustering of DEPs. A total of 101 proteins were obtained from ANOVA test (Benjamin-Hochberg FDR=0.05). A cluster of 26 proteins and a cluster of 59 proteins were annotated using DAVID 6.8. A *p* value of 0.05 was deemed significant for gene ontology analysis. Significant Biological Process terms are listed next to their respective clusters.

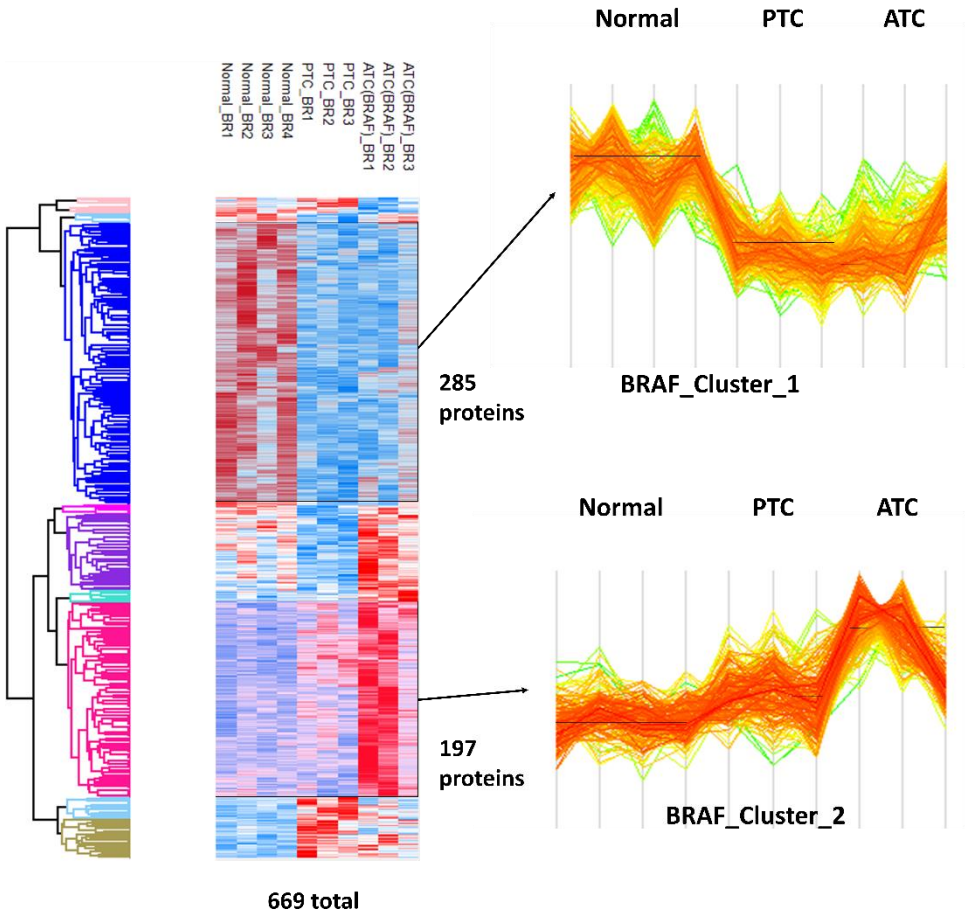


Figure 7B. DEP candidate selection for RAS. Multiple-sample T-test was conducted on Normal, FTC, and ATC(RAS) samples with a p-value of 0.05. The resulting 1623 DEPs were clustered based on their expression Clusters across the biological groups. A total of 879, 258, and 109 DEPs were selected from Cluster 2, Cluster 8, and Cluster 10, respectively.

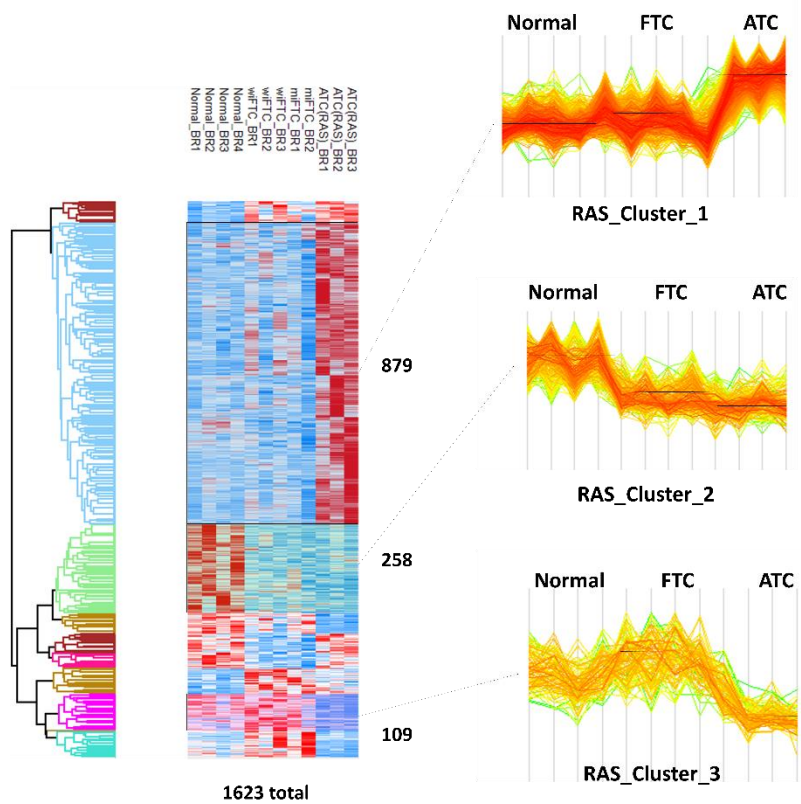


Figure 7C. DEP candidate selection for RAS. Multiple-sample T-test was conducted on Normal, FTC, and ATC(RAS) samples with a p-value of 0.05. The resulting 1623 DEPs were clustered based on their expression Clusters across the biological groups. A total of 879, 258, and 109 DEPs were selected from Cluster 2, Cluster 8, and Cluster 10, respectively.

Selection of DEPs

Applying ANOVA tests ($p = 0.05$) found 669 and 1623 significant proteins for BRAF group and RAS group, respectively (Figure 5A & Figure 5B). Significant proteins were then grouped based on expression trends across Normal→WDTC (PTC/FTC)→ATC (BRAF/RAS). Clusters that sequentially increased or decreased linearly with malignancy progression were selected as Differentially Expressed Protein (DEP) candidates (Figure 7B & 7C). In addition, a cluster of 109 (RAS Cluster 3) proteins was chosen despite displaying a Cluster that increases from Normal to FTC and then decreases in ATC. This Cluster was deemed significant as it may be indicative of upregulation of compensatory functions that are subsequently deregulated in ATC as dedifferentiation progresses. Although a similar Cluster was found for BRAF, the cluster contained was too small (39 proteins) and was therefore excluded from the analysis. Clustering of quantified proteins for BRAF showed two major expression Clusters that satisfy the criterion of sequential expression. Cluster 1 of BRAF shows an increasing trend in expression levels as malignancy increases (Figure 7B). Cluster 2 of BRAF shows an opposite Cluster where expression levels of proteins decrease in relation to malignancy. Likewise, Cluster 1 and Cluster 2 of RAS follow the same tendencies as those of BRAF with the exception of Cluster 3, which displays an increase followed by a sharp decrease.

Pathway Analysis

The selected clusters were analyzed in Ingenuity Pathway Analysis. In detail, each cluster was analyzed separately in order to utilize the program's built-in comparison analysis function. The resulting Canonical Pathway and Disease & Function terms were filtered based on significance score ($p < 0.05$) and activation

score (z-score > 1). The top 5 significant terms are listed in Table 2.

Pathway analysis was conducted on the 285 proteins and 197 proteins that belonged to Cluster 1 & Cluster 2, respectively. Canonical pathway analysis showed upregulation of key pathways such as Rac Signaling (Z=3.46, *p*=3.19E-09), NF-kB Signaling (Z=1.34, *p*=1.44) and ERK/MAPK Signaling (Z=1.34, *p* =1.59E-02) (Figure 8A). Functional analysis found down-regulation of terms such as apoptosis, cell death, and cell death of tumor cell lines (Figure 8A-2). For Cluster 2, relatively fewer pathways and terms were found to be altered by integrative analysis. Notably, oxidative phosphorylation pathway was significantly downregulated (Z=-3.61, *p* =2.40E-09) (Figure 8B-1). For disease and function terms, infection-related terms were found to be downregulated (Figure 8B-2).

For RAS Cluster 1, EIF2 Signaling pathway was the most significantly upregulated term (Z=5.30, *p*=2.99E-46). Similar to BRAF Cluster 1, RAS Cluster 1 displayed upregulation of Rac Signaling (Z=4.36, *p*=5.36E-08). Other well-established such as ERK/MAPK Signaling (Z=0.22, *p*=6.81E-07), PTEN Signaling (Z=-2.31, *p*=7.23E-03), and p53 Signaling (Z=-1.63, *p*=5.01E-03) were also identified but were not as strongly regulated or downregulated (Figure 9A-1). Likewise, tumor related functions such as apoptosis and cell death were down-regulated in RAS Cluster 1 (Figure 9A-2). The down-regulation of these terms is stronger in ATC than in FTC. RAS 2 Cluster showed downregulation of GP6 Signaling Pathway (Z=-1.63, *p*=5.42E-03), which was also identified in BRAF Cluster 2. Neoplasia of cells (Z=-1, P=1.76E-05) was the most down-regulated Disease & Function term for RAS Cluster 2.

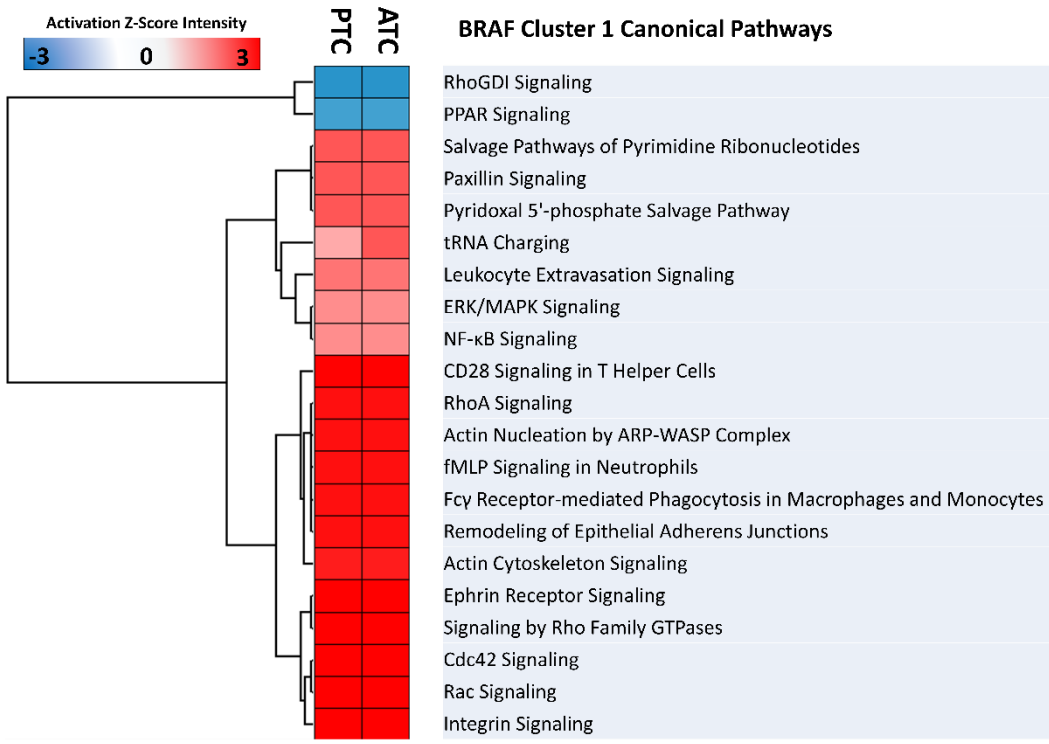


Figure 8A. Hierarchical clustering of activated terms in BRAF Cluster 1. (|Z|>1, *p<0.05)*

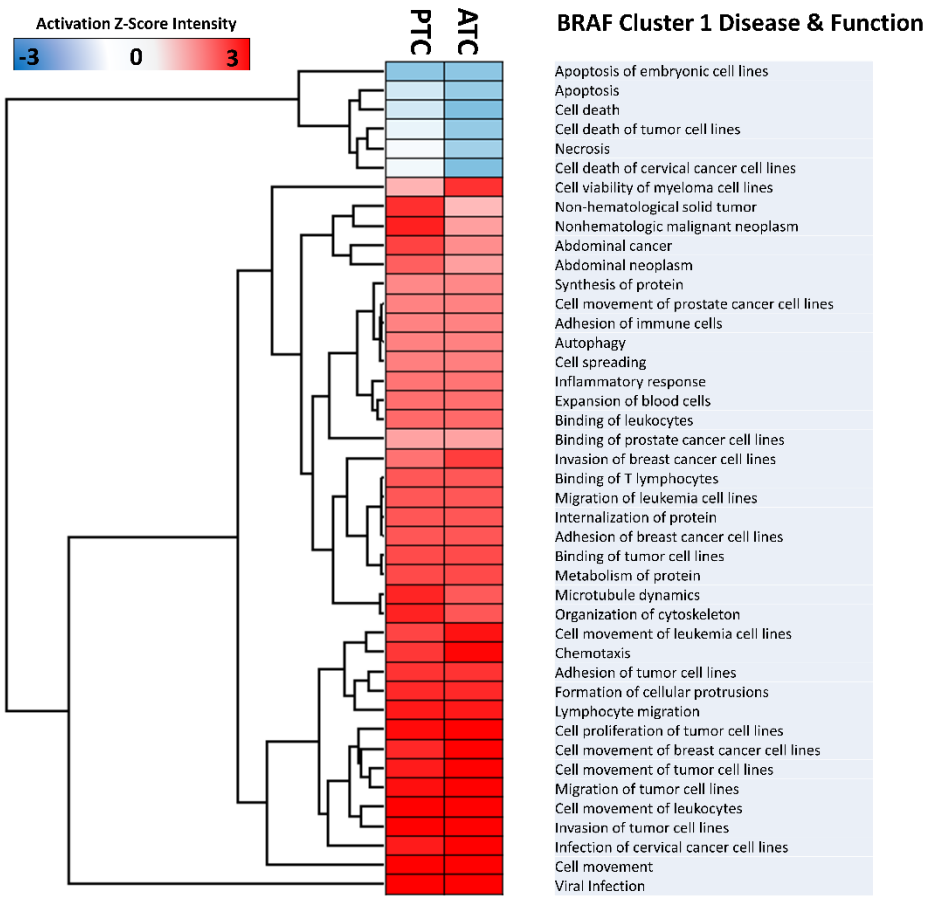


Figure 8B. Hierarchical clustering of activated terms in BRAF Cluster 1. (|Z|>1, *p<0.05)*

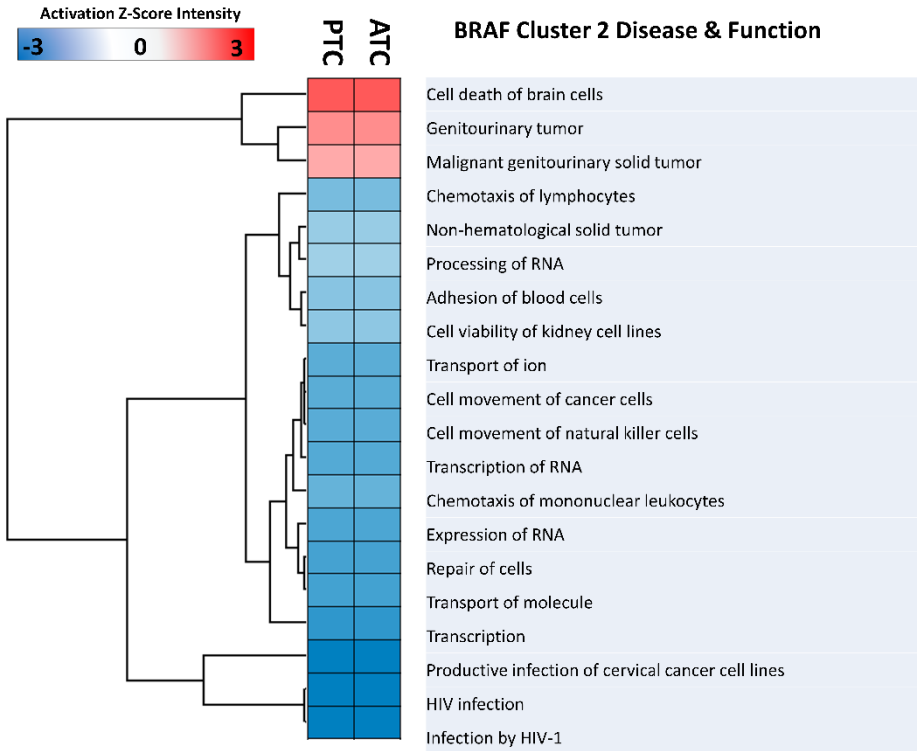


Figure 9A. Hierarchical clustering of activated terms in BRAF Cluster 2. (|Z|>1, *p<0.05)*

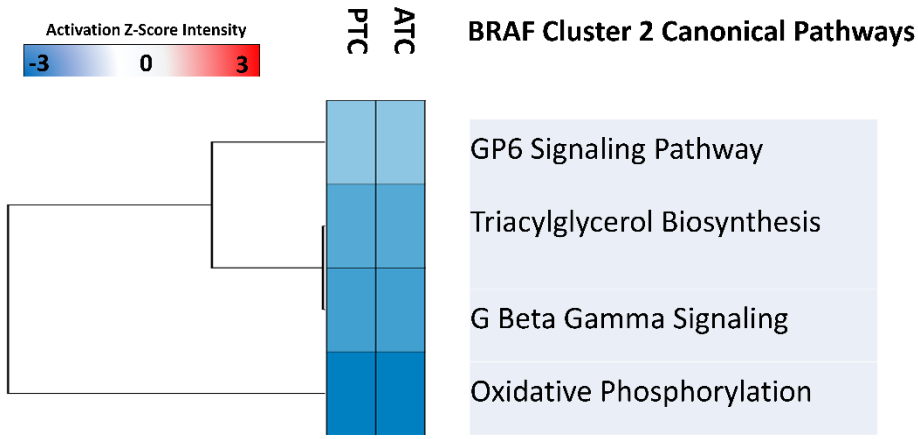


Figure 9B. Hierarchical clustering of activated terms in BRAF Cluster 2. (|Z|>1, *p<0.05)*

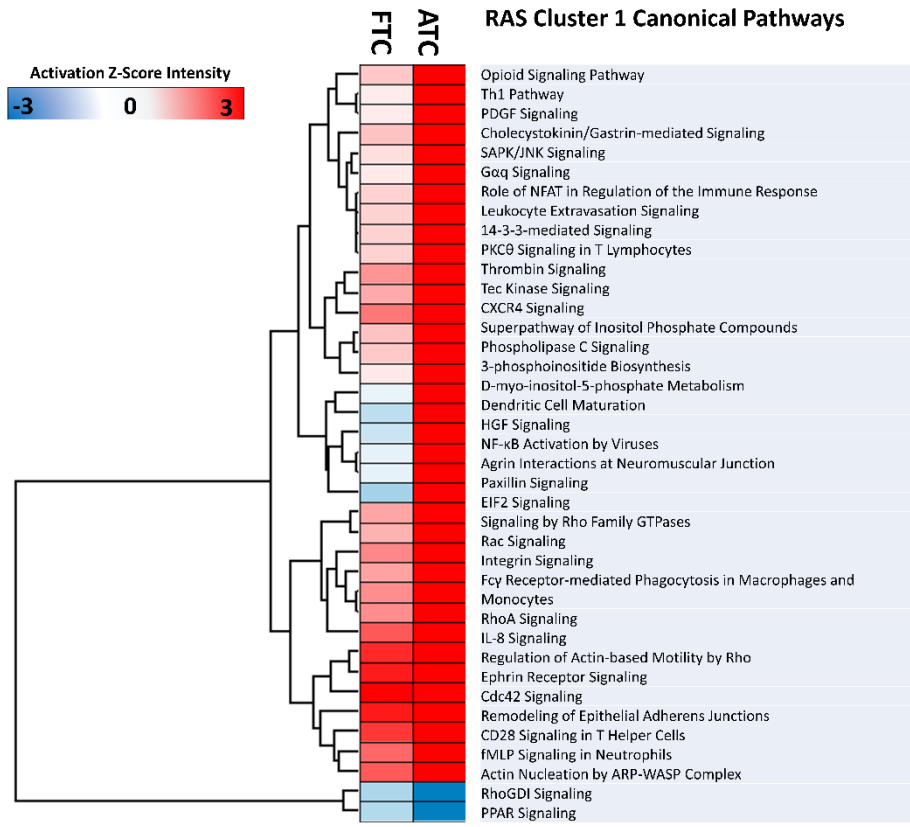


Figure 10A. Hierarchical clustering of activated terms in RAS Cluster 1. (|Z|>1, *p<0.05)*

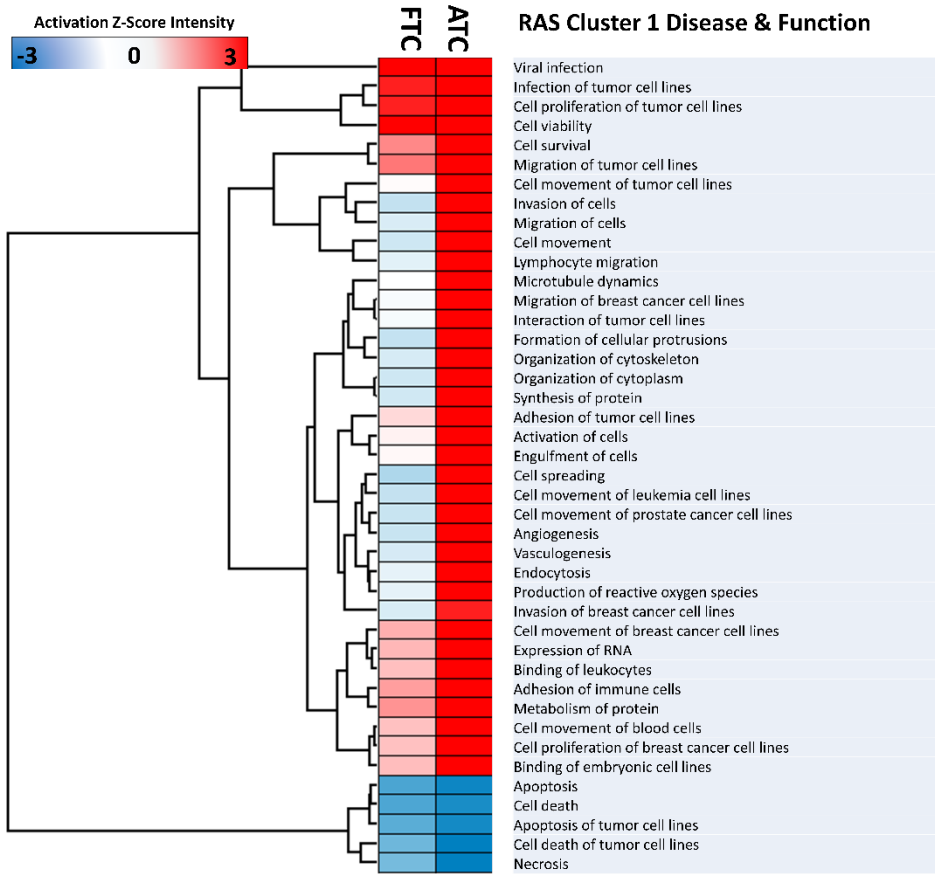


Figure 10B. Hierarchical clustering of activated terms in RAS Cluster 1. (|Z|>1, *p<0.05)*

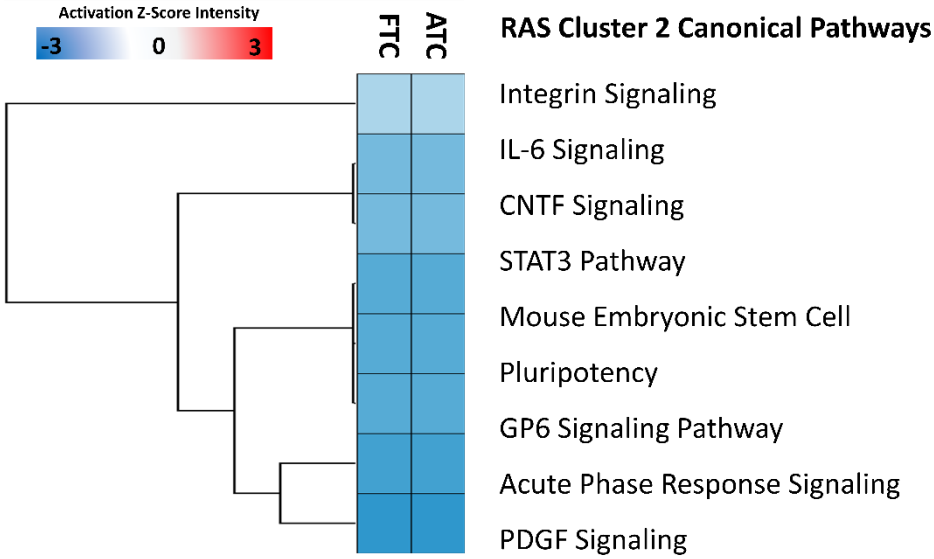


Figure 11A. Hierarchical clustering of activated terms in RAS Cluster 2. (|Z|>1, *p<0.05)*

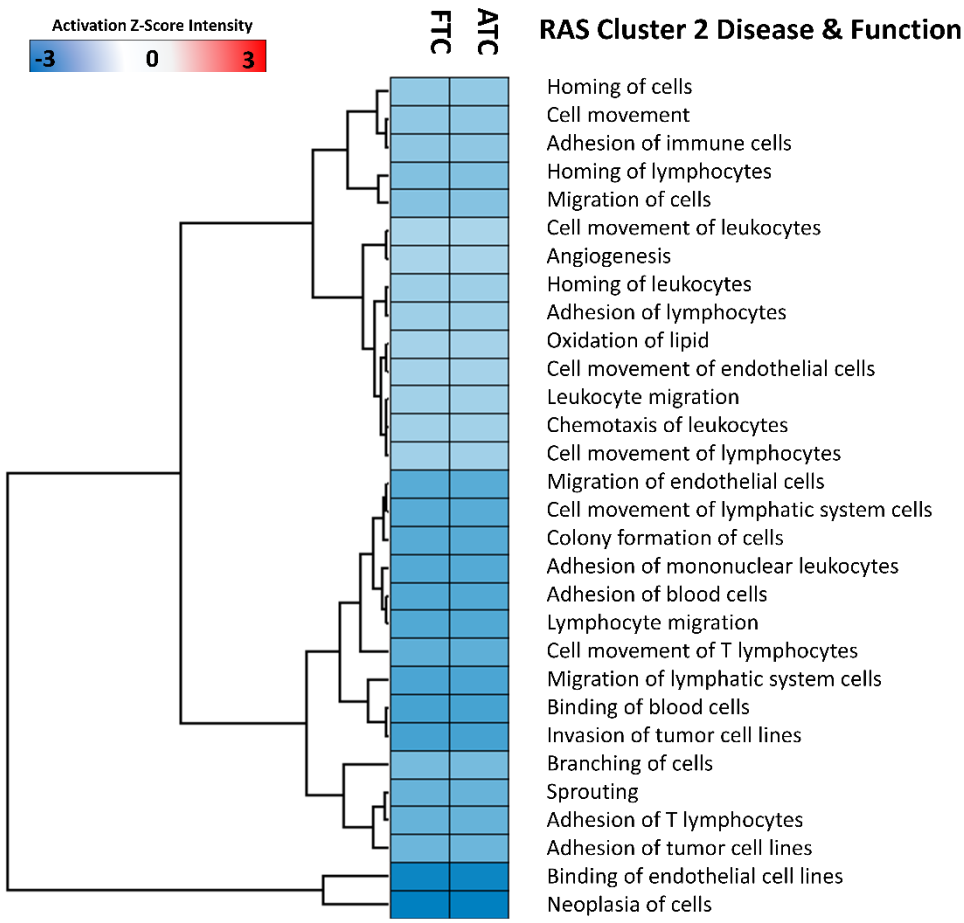


Figure 11B. Hierarchical clustering of activated terms in RAS Cluster 2. ($|Z|>1$, $p<0.05$)

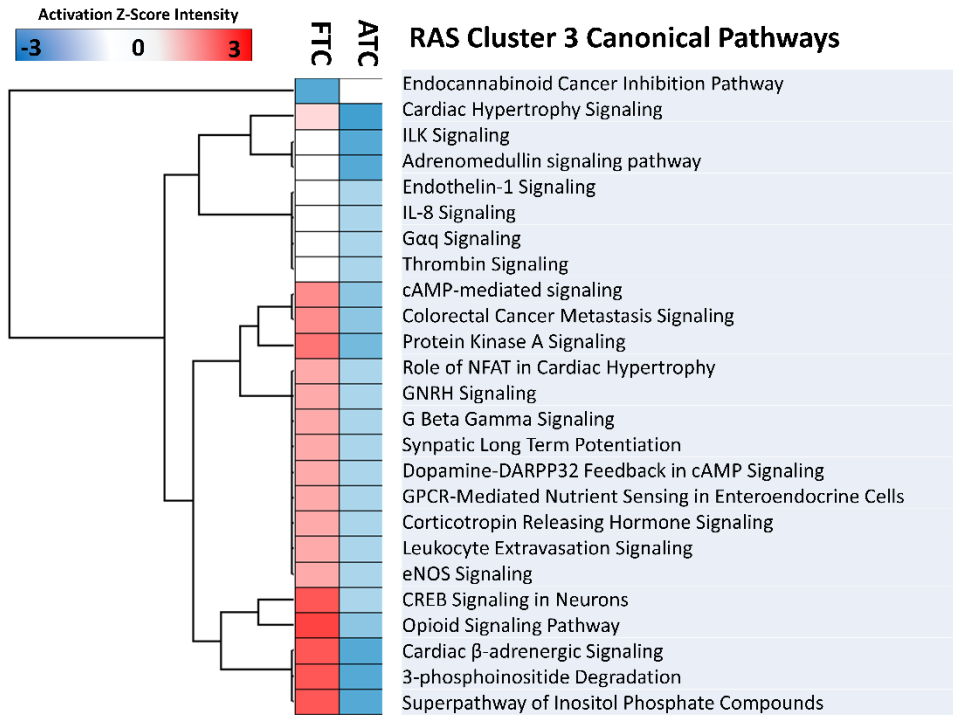


Figure 12A. Hierarchical clustering of activated terms in RAS Cluster 3. (|Z|>1, *p<0.05)*

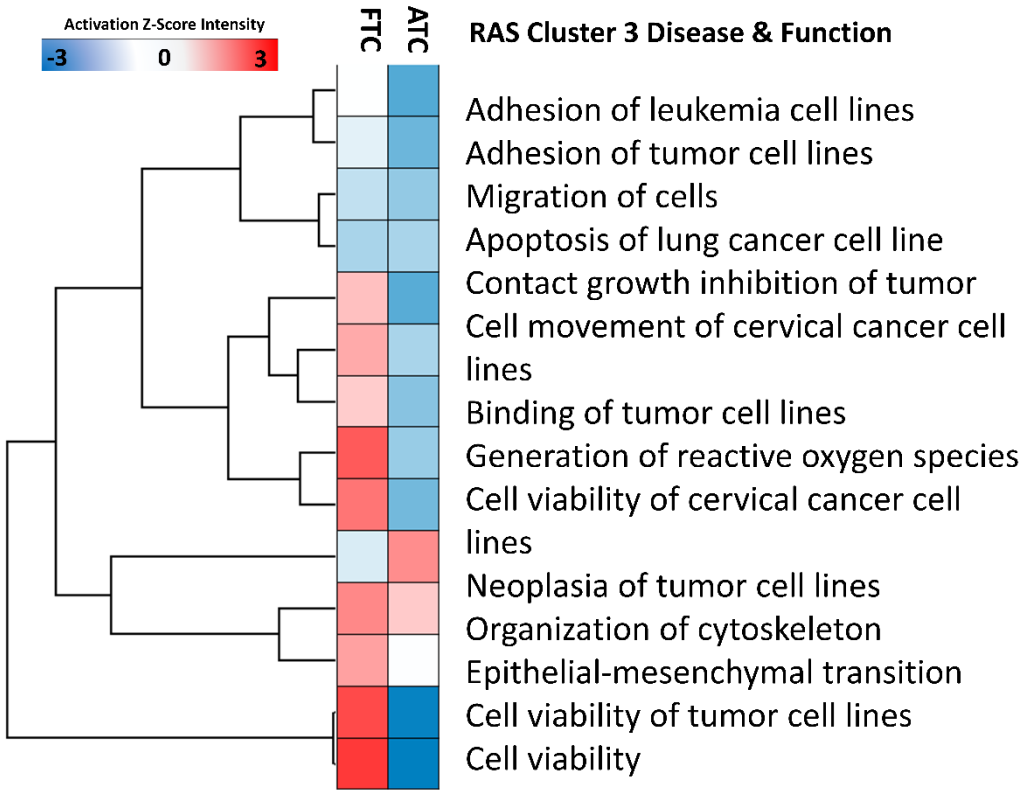


Figure 12B. Hierarchical clustering of activated terms in RAS Cluster 3. (|Z|>1, *p<0.05)*

Table 2 – Top 5 significant terms determined by IPA organized by cluster.

Canonical Pathways				
RAS			**BRAF**	
Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 1**	**Cluster 2**
EIF2 Signaling	GP6 Signaling Pathway	Colorectal Cancer Metastasis	Cdc42 Signaling	Oxidative Phosphorylatio n
Regulation of eIF4 and p70S6K Signaling	CNTF Signaling	GPCR-Mediated Nutrient Sensing in Enteroendocrine Cells	Remodeling of Epithelial Adherens Junctions	Triacylglycerol Biosynthesis
mTOR Signaling	Acute Phase Response Signaling	G Beta Gamma Signaling	Rac Signaling	G Beta Gamma Signaling
Remodeling of Epithelial Adherens Junctions	IL-6 Signaling	Synaptic Long Term Potentiation	Actin Nucleation by ARP-WASP Complex	GP6 Signaling Pathway
Actin Cytoskeleton Signaling	PDGF Signaling	IL-8 Signaling	Regulation of Actin-based Motility by Rho	
Diseases & Functions				
RAS			**BRAF**	
Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 1**	**Cluster 2**
Metabolism of protein	Leukocyte migration	Binding of tumor cell lines	Viral Infection	Proceessing of RNA
Synthesis of protein	Migration of cells	Neoplasia of tumor cell lines	Infection of cells	Transport of molcule
Cell death	Cell movement	Adhesion of tumor cell lines	Infection by RNA virus	Transcription
Viral Infection	Cell movementt of lymphatic system cells	Contact growth inhibition of tumor cell lines	Metabolism of protein	Cell movement of natural killer cells
Apoptosis	Neoplasia of cells	Migration of cells	Organization of cytoplasm	Adhesion of blood cells

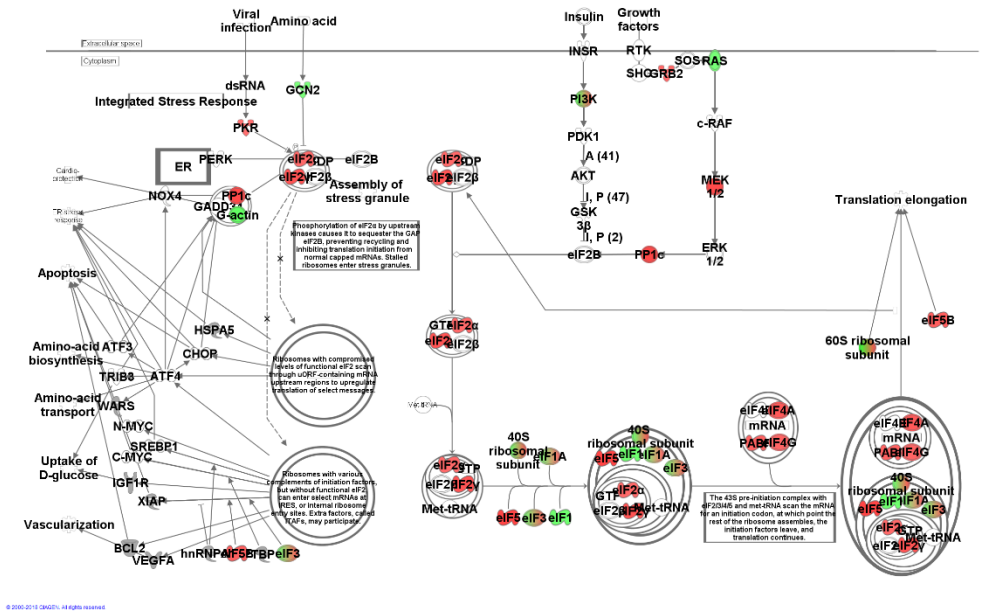


Figure 13A. Interaction map of EIF2 Signaling Pathway drawn with IPA using fold change values of FTC to Normal. Identified proteins are shown in grey. Red signifies upregulation while green denotes downregulation.

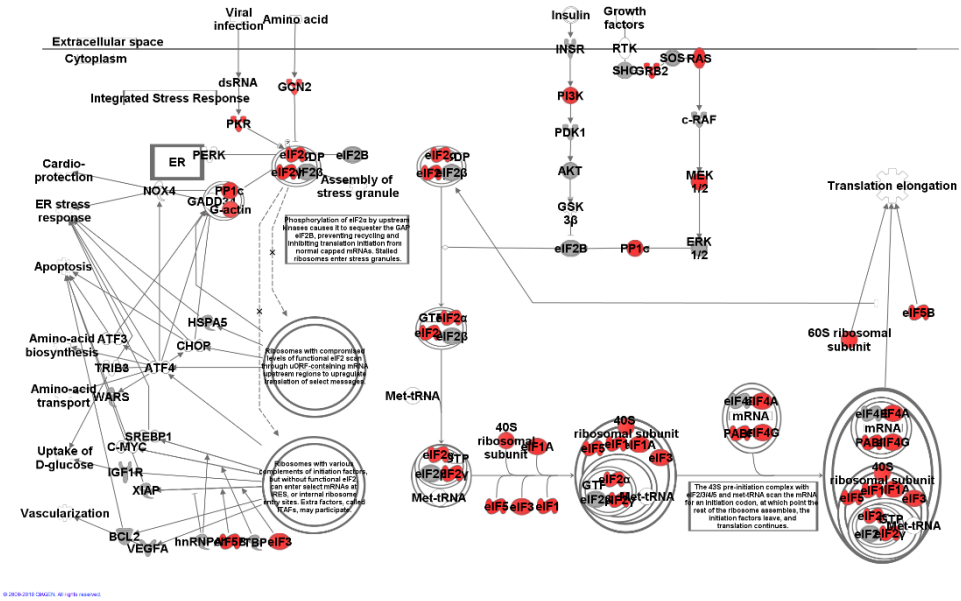


Figure 13B. Interaction map of EIF2 Signaling Pathway drawn with IPA using on fold change values of ATC(RAS) to Normal. Identified proteins are shown in grey. Red signifies upregulation while green denotes downregulation.

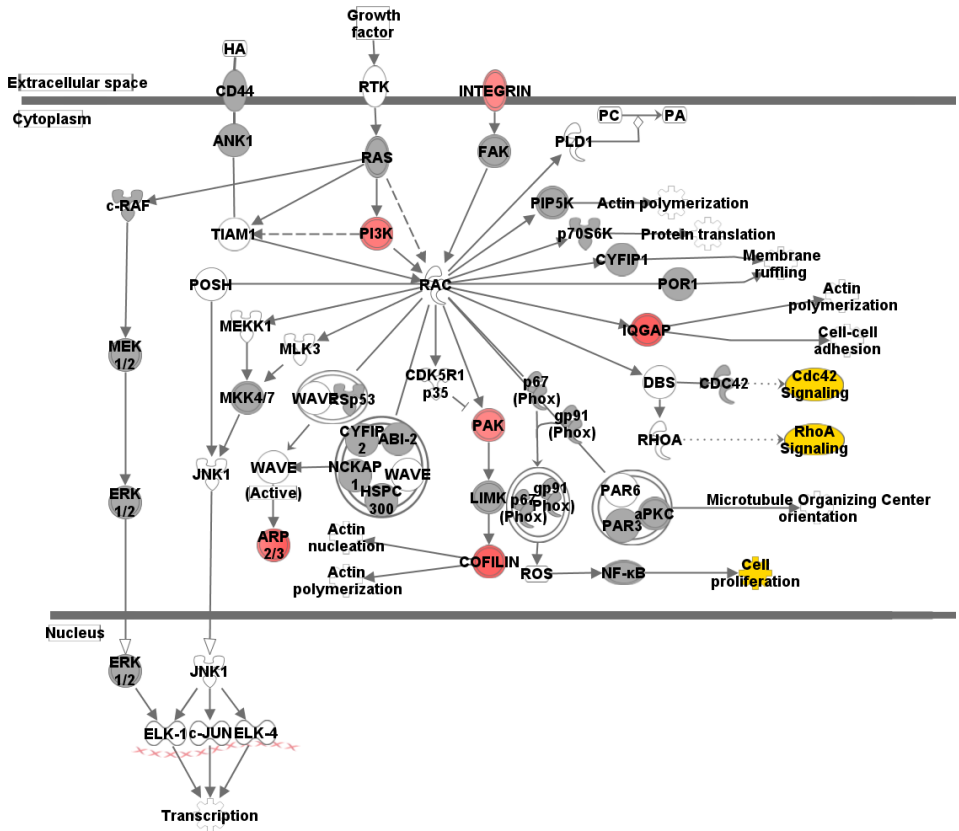


Figure 14A. Interaction map of Rac Signaling Pathway drawn with IPA using fold change values of PTC to Normal. Identified proteins are shown in grey. Red signifies upregulation while green denotes downregulation. Canonical pathways identified in BRAF clusters are highlighted in yellow. MKK4/7 was identified in PTC but not in ATC(BRAF).

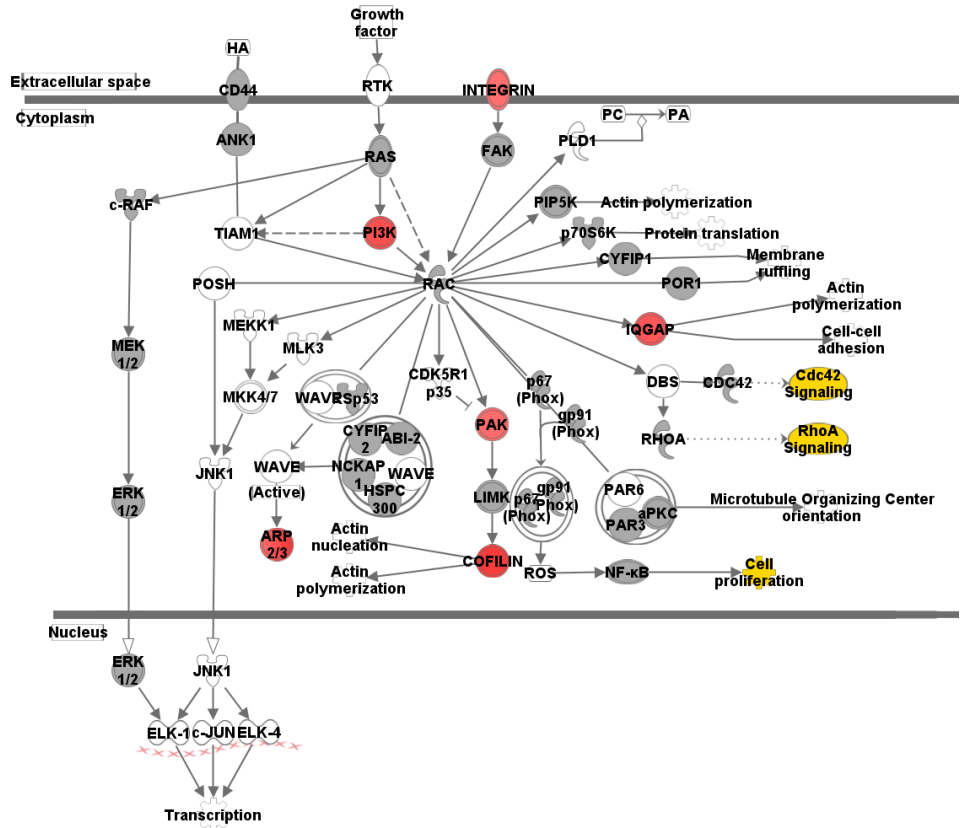


Figure 14B. Interaction map of Rac Signaling Pathway drawn with IPA using fold change values of ATC(BRAF) to Normal. Identified proteins are shown in grey. Red signifies upregulation while green denotes downregulation. Canonical pathways identified in BRAF clusters are highlighted in yellow. Rho A was identified in ATC(BRAF) but not in PTC.

Discussion

Assessment of Acquired Data

Considering the differences in LC-MS setups and operating parameters, it is difficult to exactly pinpoint why more proteins were quantified in this study than in other studies. One factor to this increase in count may be attributed to the diversity of the sample cohort. In comparison, Martinez-Aguilar et al.'s study only analyzed normal tissue and follicular adenoma whereas Liu et al. only analysed normal tissue. Additionally, the samples were analyzed over a longer gradient (3 hrs) compared to Martinez-Aguilar et al. (2 hrs) and Liu et al. (50 min). In conjunction, these factors may account for the increased coverage of nearly 1,500 proteins in this dataset with respect to Liu et al.'s.

Comparison with past literature

When considering the expression patterns of well-known protein markers such as thyroglobulin and galectin-3, the dataset was mostly consistent with established findings. Expression of thyroglobulin, the defining protein of the thyroid, was significantly downregulated in caner groups. Considering that functionality decreases in carcinoma, this finding is coherent as thyroglobulin is integral to thyroid function. [9] Galectin-3, another extensively studied candidate, was quantified at elevated levels in PTC, which is also consistent with a previous finding. [25] It is notable that catabolism of hydrogen peroxide was highly enriched (120-fold) in Normal group as hydrogen peroxide is a known downstream effectors of Ras GTPases. [26] When considering that TDS reflects the differentiation degree of thyroid tissue, the general trend of TDS from normal to ATC is sensible as

dedifferentiation is a hallmark of ATC. However, as Agrawal et al.'s study exclusively analyzed PTC, it is unknown whether this decreasing trend in FTC and ATC would be replicated in their study.

Cluster Analysis

Through analysis of the significantly upregulated and downregulated proteins, it was found that EIF2 Signalling pathway was significantly activated in RAS progression of thyroid carcinoma (Table 2). Although not in the top 5, the PI3K/AKT Signalling pathway, which is associated with RAS variant carcinoma, was also significantly activated in RAS Cluster 1. Likewise, ERK/MAPK Signalling pathway, which is known to have a fundamental role in regulating cell proliferation and tumorigenesis, was significantly activated in BRAF Cluster 1. NF-κB Signalling was found in both RAS and BRAF pathways, which is expected since the link between NF-κB Signalling's role in activating hallmark features of tumorigenesis, such as proliferation, migration, and resistance to apoptosis. [8] In addition Rac Signaling, an upstream regulator of NF- κB Signalling, was also found to upregulated in both BRAF and RAS.

As for the significantly up/down-regulated pathways, comparing the EIF2 Signalling pathway between FTC and ATC(RAS) showed that key pathways, such as RAS and PI3K, were significantly more upregulated in ATC than in FTC (Figure 9A & Figure 9B). This was accompanied by altered expression of PKR, PP1c, and eIF2. The upregulation of is eIF2 is associated with the folding and maturation of thyroglobulin, which may support Uyy et al.'s finding that chaperone macromolecules are up-regulated in cancer to compensate for the reduced availability of thyroglobulin in tumour tissue. [9] This is further supported by how

terms such as Protein folding (14-fold) Riobosomal large subunit assembly (51), and Chaperone-mediated protein folding (37-fold) were strongly enriched in RAS (Figure 8A).

A similar comparison between PTC and ATC(BRAF) was conducted for the Rac Signalling pathway. In contrast to Ras, the significantly up/down-regulated pathways in BRAF Clusters, such as Cdc42 Signalling and RhoA Signalling, were inter-connected as highlighted in Figure 11A & 11B. Surprisingly, PI3K was found to be up-regulated while ERK1/2 and other components of the MAPK pathway were not. However, considering that ERK/MAPK Signalling pathway also had passing scores for significance and activation, it is difficult to gauge the contribution of either pathway in the progression of de-differentiation.

Conclusion

Concluding remarks

Due to the rarity of the samples and the lack of interest from researchers, the mapping and identification of the thyroid carcinoma proteome is still at its infancy in comparison other cancers. As the findings of genome and RNA based studies do not necessarily correlate with those of proteome studies, it is imperative to validate those findings at the protein level. This dataset is notable in that it identified the largest number of proteins belonging to the thyroid cancer proteome. Furthermore, the dataset includes a wide variety of thyroid carcinoma, including the extremely rare ATC, which comprises less than 5% of thyroid cancer incidences. Considering that an overwhelming majority of thyroid cancer datasets only contain PTC, I expect this dataset to be an invaluable contribution to researchers studying malignant thyroid carcinomas. In addition, it is the first dataset to quantify a wide array of thyroid cancer subtypes via reporter ions. The use of reporter ions is significant as TMT labelling yields numerous advantages: simultaneous quantification and a significant decrease in run-time. As mass spectrometers are delicate instruments, they require frequent maintenance as their sensitivity deteriorates over time when samples inevitably contaminate the detector. Thus, simultaneous quantification, which drastically reduces the number of runs compared to analysing each individual fraction of every sample, improves the quality of the acquired data by reducing run-to-run variation.

Future works

One of the limitations of this study is the low number of samples per histological group. As thyroid cancers are heterogeneous and abundant in subtypes, having a large number of samples is instrumental in procuring the necessary statistical power to override individual variation. In addition, employing a MARS 14 column to deplete abundant proteins may help further increase the coverage of the thyroid proteome.

References

1. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2018.* CA Cancer J Clin, 2018. **68**(1): p. 7-30.
2. Fagin, J.A. and S.A. Wells, Jr., *Biologic and Clinical Perspectives on Thyroid Cancer.* N Engl J Med, 2016. **375**(11): p. 1054-67.
3. Liska, J., et al., *Thyroid tumors: histological classification and genetic factors involved in the development of thyroid cancer.* Endocr Regul, 2005. **39**(3): p. 73-83.
4. Paschke, R., et al., *The Treatment of Well-Differentiated Thyroid Carcinoma.* Dtsch Arztebl Int, 2015. **112**(26): p. 452-8.
5. McHenry, C.R. and R. Phitayakorn, *Follicular adenoma and carcinoma of the thyroid gland.* Oncologist, 2011. **16**(5): p. 585-93.
6. Chakravarty, D., et al., *Small-molecule MAPK inhibitors restore radioiodine incorporation in mouse thyroid cancers with conditional BRAF activation.* J Clin Invest, 2011. **121**(12): p. 4700-11.
7. Pita, J.M., et al., *Gene expression profiling associated with the progression to poorly differentiated thyroid carcinomas.* Br J Cancer, 2009. **101**(10): p. 1782-91.
8. Xing, M., *Molecular pathogenesis and mechanisms of thyroid cancer.* Nat Rev Cancer, 2013. **13**(3): p. 184-99.
9. Uyy, E., et al., *Endoplasmic Reticulum Chaperones Are Potential Active Factors in Thyroid Tumorigenesis.* J Proteome Res, 2016. **15**(9): p. 3377-87.
10. Ban, Y., et al., *Proteomic profiling of thyroid papillary carcinoma.* J

Thyroid Res, 2012. **2012**: p. 815079.

11. Martinez-Aguilar, J., R. Clifton-Bligh, and M.P. Molloy, *Proteomics of thyroid tumours provides new insights into their molecular composition and changes associated with malignancy.* Sci Rep, 2016. **6**: p. 23660.
12. Martinez-Aguilar, J., R. Clifton-Bligh, and M.P. Molloy, *A multiplexed, targeted mass spectrometry assay of the S100 protein family uncovers the isoform-specific expression in thyroid tumours.* BMC Cancer, 2015. **15**: p. 199.
13. Gawin, M., et al., *Proteome profiles of different types of thyroid cancers.* Mol Cell Endocrinol, 2018. **472**: p. 68-79.
14. Molloy, M.P., *Two-dimensional electrophoresis of membrane proteins using immobilized pH gradients.* Anal Biochem, 2000. **280**(1): p. 1-10.
15. Zhang, L. and J.E. Elias, *Relative Protein Quantification Using Tandem Mass Tag Mass Spectrometry*. Methods Mol Biol, 2017. **1550**: p. 185-198.
16. Thompson, A., et al., *Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS.* Anal Chem, 2003. **75**(8): p. 1895-904.
17. Rauniyar, N. and J.R. Yates, 3rd, *Isobaric labeling-based relative quantification in shotgun proteomics.* J Proteome Res, 2014. **13**(12): p. 5293-309.
18. Do, M., et al., *Quantitative proteomic analysis of pancreatic cyst fluid proteins associated with malignancy in intraductal papillary*

mucinous neoplasms. Clin Proteomics, 2018. **15**: p. 17.

19. Woo, J., et al., *Quantitative Proteomics Reveals Temporal Proteomic Changes in Signaling Pathways during BV2 Mouse Microglial Cell Activation.* J Proteome Res, 2017. **16**(9): p. 3419-3432.
20. Wisniewski, J.R. and F.Z. Gaugaz, *Fast and sensitive total protein and Peptide assays for proteomic analysis.* Anal Chem, 2015. **87**(8): p. 4110-6.
21. Kim, D.K., et al., *Molecular and functional signatures in a novel Alzheimer's disease mouse model assessed by quantitative proteomics.* Mol Neurodegener, 2018. **13**(1): p. 2.
22. Tyanova, S., et al., *The Perseus computational platform for comprehensive analysis of (prote)omics data.* Nat Methods, 2016. **13**(9): p. 731-40.
23. Liu, X., et al., *Comprehensive Map and Functional Annotation of Human Pituitary and Thyroid Proteome.* J Proteome Res, 2017. **16**(8): p. 2680-2691.
24. Cancer Genome Atlas Research, N., *Integrated genomic characterization of papillary thyroid carcinoma.* Cell, 2014. **159**(3): p. 676-90.
25. Chiu, C.G., et al., *Diagnostic utility of galectin-3 in thyroid cancer.* Am J Pathol, 2010. **176**(5): p. 2067-81.
26. Ferro, E., et al., *The Interplay between ROS and Ras GTPases: Physiological and Pathological Implications.* J Signal Transduct, 2012. **2012**: p. 365769.

국문 초록

분화갑상선암과 비분화갑상선암의 중동체 라벨을 이용한 프로테옴 프로파일링 연구

왕인재

서울대학교 대학원

협동과정 바이오엔지니어링 전공

갑상선암은 매년 미국의 인구 기준으로 56,000 건의 새로운 사례가 발병하고 2000 명의 인명피해를 끼치는 종양질환이다. 높은 유병률을 보유하고 있음에도 불구하고 양성질환의 생존률이 높기 때문에 갑상선암의 위험이 낮게 평가되고 있다. 기존에 유전체학 연구를 통해 종양 형성 유전자들과 그에 관련된 돌연변이들을 발견했지만 사상자가 많이 없기 때문에 다른 암에 비해 갑상선과 갑상선암에 관한 단백체 연구가 적극적으로 진행되고 있지 않다. 양성 질환의 갑상선 암이라도 악성 질환으로 변할 가능성을 배제할

수 없기 때문에 프로테움 프로파일링을 통해 갑상선암을 포괄적인 측면에서 연구할 필요가 있다. 본 연구에서는 분화갑상선암과 비분화 갑상선암을 질량 분석기를 사용하여 분석하였다. 총 18명의 환자에게로부터 얻은 시료들을 Easy-nLC 1000와 Q-Exactive 질량분석기를 이용하여 7071개의 단백질들을 동정하였고 6215개의 단백질들을 적량하였다. 분화갑상선에서 비분화갑상선으로 진행되는 과정에 RAS Progression에서 EIF2 Signaling Pathway와 BRAF Progression에서 RAC Signaling Pathway가 유의적으로 변하는 것을 확인하였다.

주요어: 갑상선암, 단백체학, 질량 분석기
학번: 2017-25653