JBUON 2020; 25(6): 2690-2699
ISSN: 1107-0625, online ISSN: 2241-6293 • www.jbuon.com
Email: editorial_office@jbuon.com

ORIGINAL ARTICLE

# Identification of key pathways and biomarkers in anaplastic thyroid cancer using an integrated analysis

Jialu Zhou[1], Shuhui Dong[2], and Chao Shi[3]

[1]Department of Clinical Medicine, The Second Clinical Medical College, Nanchang University, Nanchang, Jiangxi, China. [2]Department of Clinical Medicine, The Fourth Clinical Medical College, Nanchang University, Nanchang, Jiangxi, China. [3]Department of Oncology, The First Affiliated Hospital of Nanchang University, Nanchang, Jiangxi, China

## Summary

***Purpose:*** *Thyroid carcinoma (THCA) is one of the most common endocrine tumours with high morbidity worldwide. Anaplastic thyroid cancer (ATC) is the most fatal and has the poorest prognosis of the four THCA types, as it lacks effective treatments. Early screening of ATC is problematic and so identifying ATC biomarkers is increasingly crucial.*

***Methods:*** *We performed a systematic search of the thyroid transcriptome in the Gene Expression Omnibus (GEO) database and an integrative analysis of gene expression profiles. Moreover, we conducted a pathway enrichment analysis in ATC using the WEB-based GEne SeT AnaLysis Toolkit. We identified the intersections of all the differentially expressed genes (DEGs) between ATC and normal samples and DEGs between ATC and non-ATC samples in the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING). Finally, we used Cytoscape software to visualize the protein-protein interaction (PPI) network.*

***Results:*** *Six gene expression datasets containing 131 thyroid cancer samples and 98 normal control samples were collected to identify the significant DEGs. A total of 1489 DEGs were identified between ATC and normal samples, and 522 DEGs between ATC and non-ATC samples. ATC showed a greater association with the cell cycle. The Principal component analysis (PCA) results revealed 222 genes with substantial contributions to the identification of ATC.*

***Conclusion:*** *Cell cycle plays a decisive role in the high mortality rate of ATC. TOP2A, NUSAP1, PBK, KIF15, CENPF, CEP55, CDK1, CCNB2, CDCA8 and CDC20 were identified as hub genes.*

***Key words:*** *anaplastic thyroid cancer (ATC); cell cycle; biomarkers; differentially expressed genes (DEGs).*

## Introduction

Thyroid carcinoma (THCA) is the most common endocrine tumour, with an increasing incidence in both developed and developing countries [1,2]. Most cases are pathologically differentiated thyroid cancers (DTC) with a good prognosis. Among DTC, papillary thyroid cancer (PTC) is the most common type, while DTC also includes follicular thyroid cancer (FTC) and medullary thyroid cancer (MTC). Anaplastic thyroid cancer (ATC) retains follicular elements and thyroglobulin production, and is classified as poorly differentiated thyroid cancer (PDTC) [3]. These four types are classified based on the original cancer cell types and their appearance and characteristics. ATC is very rare, as it affects fewer than 2% of patients with thyroid cancer [4]. According to epidemiological investigations, ATC is arguably the solid tumour with the highest mortality rate in humans and is characterized by rapid growth and metastasis in the nonremission clinical stage. ATC does not respond to most treatments and is almost always fatal [5]. The median overall survival (OS) of patients is only 3-5 months [6-8].

*Corresponding author:* Chao Shi, MD. Department of Oncology, The First Affiliated Hospital of Nanchang University, 17 Yong Waizheng Street, Nanchang, Jiangxi 330006, China.
Tel: +86-15170498689, Fax: +86 0791- 86319566, Email: KKZZ023@163.com
Received: 03/08/2020; Accepted: 30/08/2020

The disease poses challenges for clinicians and researchers to develop new systematic therapies [5]. In the management of ATC, multiple treatments, including available surgery, remain the best strategy for improving survival and preventing suffocation-induced death. The addition of chemotherapy has improved survival, but the prognosis remains poor [9]. To date, an effective treatment that prolongs the survival of patients with ATC is unavailable [10]. Therefore, appropriate comfort measures and emotional support should be provided to these unfortunate patients while they are urgently awaiting the development of a new treatment for this type of cancer [11]. Based on insights into the molecular, genetic, and biochemical changes that occur during carcinogenesis, the focus of drug development has shifted from empirical therapy to treatment with specific molecular targets that cause tumours [12]. New targeted therapies aimed at blocking the activity of oncogenes in ATC that promote cell growth and oncogenic transformation are the most promising therapies currently being explored [13]. Hence, a new target of ATC is urgently needed [14]. However, currently, little is known about the exact molecular mechanisms underlying ATC progression [15]. The reintroduction of wild-type p53 into ATC cells has been reported to induce cell differentiation, inhibit cell proliferation, and restore cell reactivity [16-18]. Both autophagy and AKT/mTOR signalling are also involved in ATC [19]. *CDK7*, *PPP1R15A* [20], and *KAT5* [21] were identified as potential biomarkers and therapeutic targets for ATC in two different studies. Nevertheless, most studies of ATC have been conducted in a single facility with a limited number of patients [22]. In other words, the molecular mechanism of ATC remains unclear.

We performed this integrated analysis to understand why the mortality rate of ATC is particularly high among patients with THCA and to identify the key pathways and biomarkers of ATC. Differentially expressed genes (DEGs) are identified by folding changes of expression [23]. The identification of DEGs between ATC tissues and normal tissues along with the identification of DEGs between ATC tissues and non-ATC tissues is an effective method to find a new target for the treatment of thyroid cancer. Combined with existing microarray datasets in data warehouses, the reliability and universality of the statistical results can be improved. Moreover, the statistical ability to produce more robust and accurate findings leads to more accurate judgments of DEGs [23,24]. A pathway enrichment analysis was performed to reveal the functional changes in ATC. ATC has more specific DEGs and shows a stronger association with the cell cycle than THCA and PTC, which corresponds to the extremely high mortality of ATC. Finally, by identifying hub genes, we enriched the study of the molecular mechanisms of ATC. Our study contributes to guiding gene targeted therapy for ATC, which is important for prolonging the survival of patients with ATC.

## Methods

*Selection of microarray datasets for the meta-analysis*

A search was performed in the Gene Expression Omnibus (GEO) database from the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/geo/) with the keywords "thyroid cancer" and "thyroid neoplasms". The inclusion criteria were: (1) all samples were obtained from Homo sapiens; and (2) more than three samples of normal thyroid were available. Studies were excluded if they met the following criteria: (1) samples were obtained from peripheral blood; (2) dual-channel studies; (3) studies examining the miRNA expression profile and DNA methylation; (4) studies that lacked controls and analysed other cancers; (5) studies including datasets without literature traceability; and (6) studies of cell lines.

*Meta-analysis of multiple microarray datasets*

We downloaded files (.CEL) containing microarray datasets of thyroid samples from the GEO database. We performed a meta-analysis of the gene expression profiles of six datasets using R statistical software (http://www.r-project.org/) to reveal the differences in expression between THCA and normal samples and to identify

**Table 1.** Characteristics of the individual studies retrieved from GEO for the meta-analysis

| *Study* | *Samples* | *Cases/ controls* | *ATC/PTC* | *Country* | *PMID* | *Platform* | *Chip type* |
|---|---|---|---|---|---|---|---|
| GSE65144 | 25 | 12/13 | 12/0 | USA | 25675381 | GPL570 | HG-U133_plus_2 |
| GSE33630 | 105 | 60/45 | 11/49 | Belgium | 22266856 | GPL570 | HG-U133_plus_2 |
| GSE29265 | 49 | 29/20 | 9/20 | Belgium | 22828612 | GPL570 | HG-U133_plus_2 |
| GSE6004 | 18 | 14/4 | 0/14 | USA | None | GPL570 | HG-U133_plus_2 |
| GSE3678 | 14 | 7/7 | 0/7 | USA | 17296934 | GPL570 | HG-U133_plus_2 |
| GSE3467 | 18 | 9/9 | 0/9 | USA | None | GPL570 | HG-U133_plus_2 |
| Meta-analysis | 229 | 131/98 | 32/99 | - | 16365291 | - | HG-U133_plus_2 |

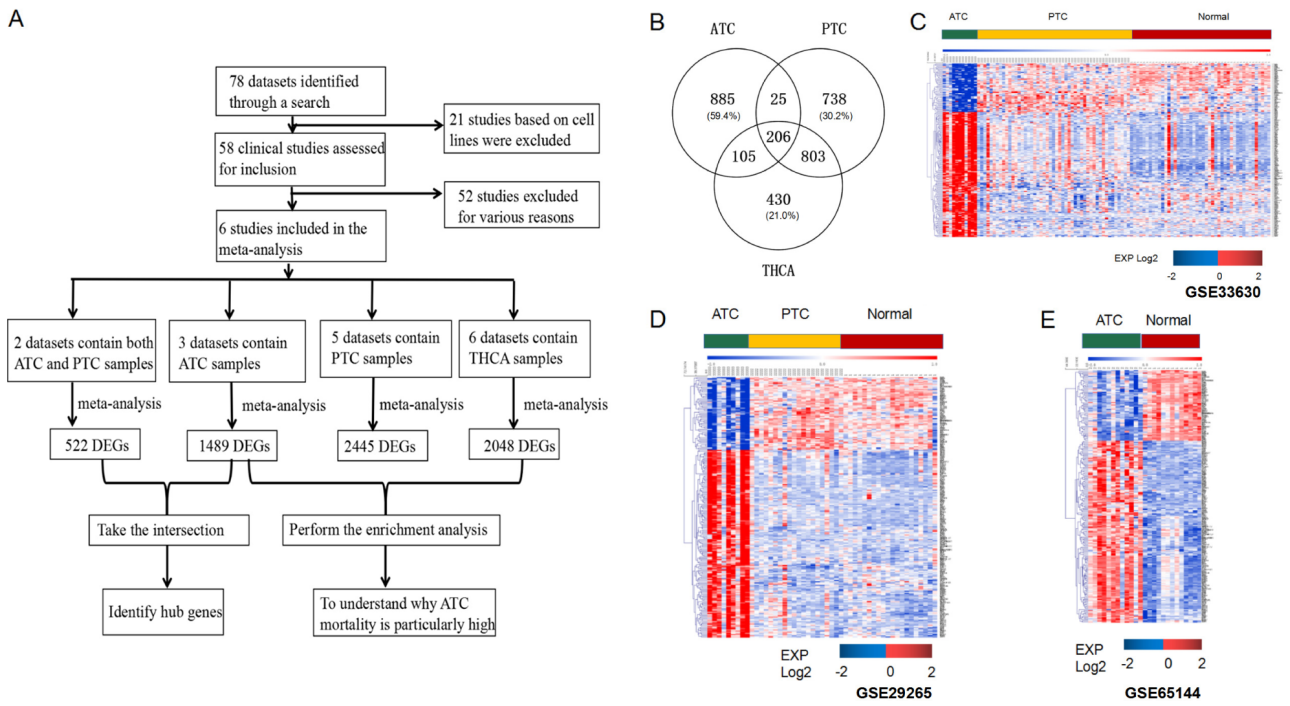
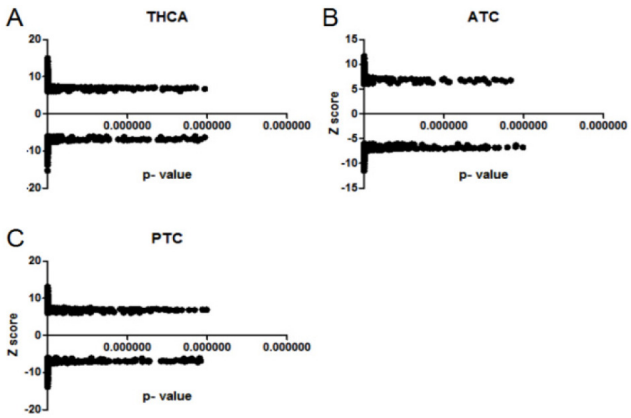



**Figure 1.** Summary of the integrative analysis based on the GEO dataset. **A:** Flow chart of the data collection and analysis processes. **B:** Venn diagram showing the overlapping DEGs in THCA, ATC, and PTC constructed using Venny 2.1.0. Specific genes in ATC accounted for a larger proportion, suggesting that a more specific pathway is involved in ATC development. **C-D:** The 222 genes clustered in ATC versus PTC and normal tissues in GSE33630 **(C)**. Each column indicates a sample, and each row shows the expression level of a gene and GSE29265 **(D)**. **E:** The 222 genes clustered in ATC versus PTC and normal tissues in GSE65144. The colour scale represents the raw Z score ranging from blue (low expression) to red (high expression). The change in the sample group was consistent with the change in colour levels. GEO: Gene Expression Omnibus; DEGs: differentially expressed genes; THCA: thyroid carcinoma; ATC: anaplastic thyroid cancer; PTC: papillary thyroid cancer.




**Supplementary Figure 1.** Scatter diagram of the p values and Z scores of THCA **(A)**, ATC **(B)** and PTC **(C)** based on a p value < 1e-12 and |Z|> 6. The X axis represents p values, while the Y axis represents Z scores. Upregulated DEGs are located above the X axis, while downregulated DEGs are located below the X axis.

a list of DEGs (up- or downregulated). Using the MAMA, mataMA, affyPLM, CLL and RankProd packages, we finally obtained p values and Z scores with the T-test and Z score methods. The Z score (threshold of the absolute value > 6) and p value (value < 1e-12) were used as the DEG screening criteria. In addition, we divided the thyroid cancer samples into two groups, including ATC and PTC samples, to perform the meta-analysis. Finally, we performed a meta-analysis of GSE29265 and GSE33630 to reveal the differences in expression between ATC and non-ATC samples. We took the intersections of all the DEGs between ATC and normal samples and DEGs between ATC and non-ATC samples to identify them as hub genes.

*Principal component analysis*

Principal component analysis (PCA) has become a common method used to reveal the most important changes in proteins [26]. It is a widely used feature extraction tool in the computer vision field [27]. We used it to verify the hub genes we identified to determine whether they do contribute differentiating between ATC and PTC, normal samples. PCA does not quantitatively predict performance, and thus the area under the Receiver Operating Characteristics (ROC) curve was also calculated for verification.

*Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis*

We performed enrichment analyses based on the meta-analysis results to further investigate the biological characteristics of the most significant DEGs. We conducted a KEGG pathway enrichment analysis to identify

the most significant DEGs using the WEB-based GEne SeT AnaLysis Toolkit (http://www.webgestalt.org/option.php) and we drew a bubble chart to visualise the results.

*Protein-protein interaction (PPI) network construction*

The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database is used to collect, score, and integrate all publicly available PPI information and to supply this information by calculating predictions. For the enrichment analysis, STRING not only implements well-known classification systems such as GO and KEGG but also provides additional new classification systems based on high-throughput text mining and hierarchical clustering based on the association network itself. The STRING resource is available online at https://string-db.org/ [27]. Furthermore, we used Cytoscape software to visualize the PPI network through the MCODE plug-in app and cytoHubba as an auxiliary tool.

*Selection of hub genes and modules*

We employed Molecular Complex Detection (MCODE) software with a cut-off value=2, node score cut-off=0.2, k-core=2, and max depth=100 to select the most important PPI network clustering module. The twelve algorithms of the cytoHubba plug-in in Cytoscape software were used to identify hub genes, after which we conducted an analysis of the most significant module. We drew box plots to show the expression of hub genes in GSE33630 and GSE29265 that contain all ATC, PTC and normal samples to reveal the differential expression of hub genes in ATC, PTC and normal samples. In addition, we verified the consistency of the included studies by drawing the forest charts of hub genes expressed in each study. While PCA cannot quantitatively predict performance, so the area of ROC curve was drawn to reveal the efficacy of hub genes. We drew ROC curves of expressions of hub genes in datasets contains ATC samples except GSE65144 for it contains less samples which may not be well represented. We divided samples into ATC, PTC and normal groups to verify whether selected hub genes can successfully distinguish ATC samples from PTC and normal samples. We evaluated the cross-validation performance using average ROC area under the curve (AUC) metrics.

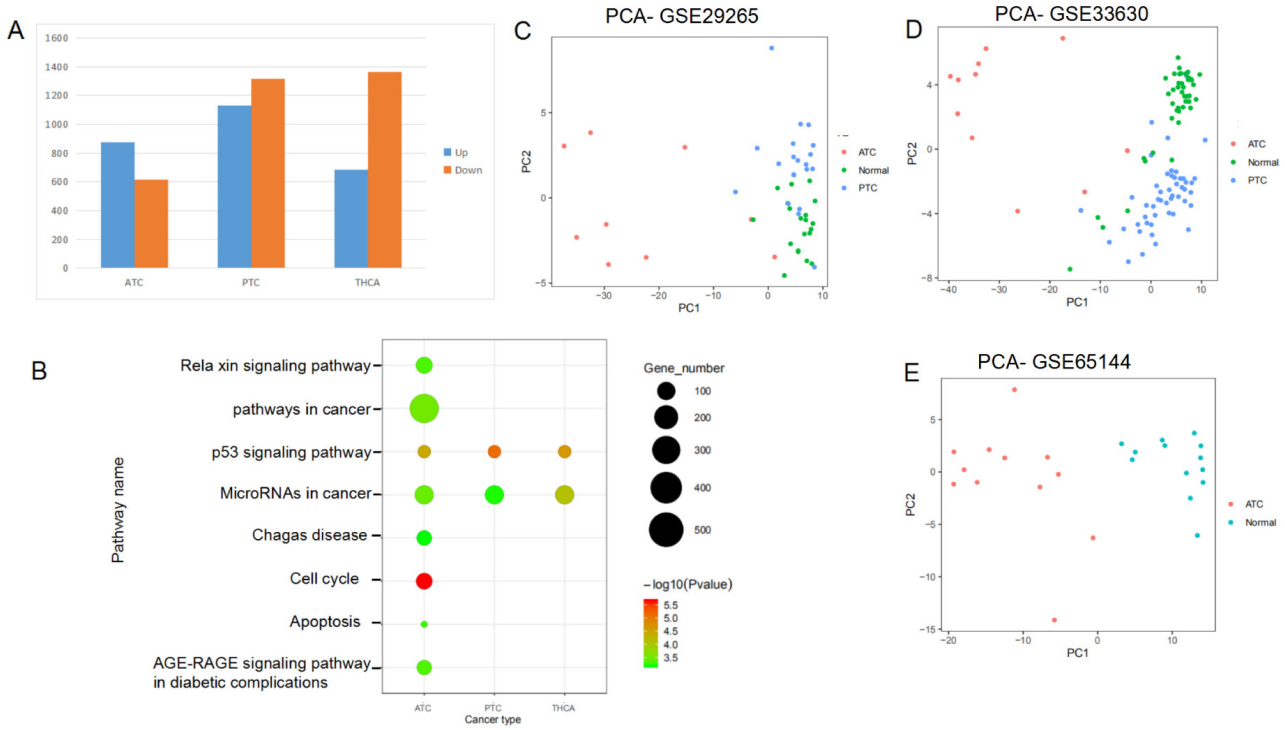



**Figure 2.** Verification of the efficacy and pathway analysis of DEGs identified in ATC. **A:** Bar charts of the up- and downregulated DEGs in ATC, PTC, and THCA. Blue represents the upregulated DEGs, while orange represents the downregulated DEGs. ATC had more upregulated DEGs than downregulated DEGs, while PTC and THCA showed the opposite results. **B:** Bubble chart of the top 8 KEGG pathways of DEGs identified in THCA, PTC and ATC. The pathways were sorted according to p values ($p<0.001$) for enriched DEGs. Compared with THCA and PTC samples, more pathways were enriched in ATC samples. Larger bubbles indicate that the pathway contained more DEGs, and a redder bubble indicated greater significance of the pathway. **C-D:** Scatter diagrams drawn after PCA; ATC samples were significantly separated from PTC and normal samples in GSE29265 **(C)** and GSE33630 **(D)**. Red represents ATC samples, green represents PTC samples, and blue represents normal samples. Red points were separate from green and blue points. **(E)** Scatter diagram drawn after PCA; ATC samples were significantly separated from PTC samples in GSE65144.
PTC: papillary thyroid cancer; ATC: anaplastic thyroid cancer; THCA: thyroid carcinoma; DEGs: differentially expressed genes; PCA: principal component analysis.

## Results

*Identification of DEGs through a meta-analysis*

According to our inclusion criteria, six GEO datasets from NCBI (GSE65144, GSE33630, GSE29265, GSE6004, GSE3678, and GSE3467) were collected (Figure 1A). We analysed 131 thyroid cancer samples and 98 normal thyroid tissue samples (Table 1). According to the Z score (threshold of an absolute value>6) and p value (value<$10^{-12}$), we obtained 2048 DEGs in the comparison of THCA and normal tissues (supplementary Figure 1A). In addition, we divided the 131 thyroid cancer samples into two groups, including 32 ATC and 99 PTC samples. Compared with normal tissues, 1489 DEGs were identified in ATC samples from 3 datasets, including GSE65144, GSE33630, and GSE29265 (supplementary Figure 1B). Meanwhile, 2445 DEGs were obtained in PTC samples from 5 datasets (all except GSE65144) (supplementary Figure 1C). We identified 522 DEGs after a meta-analysis comparing GSE29265 and GSE33630 with the same cut-off value.

We further investigated the specific DEGs among THCA, ATC and PTC. Two hundred six shared genes were identified (Figure 1B). In addition, specific genes in ATC samples accounted for a larger proportion (885 genes, 59.4%), suggesting that a more specific pathway was involved in ATC development. By drawing the bar charts, we observed more upregulated DEGs than downregulated DEGs in ATC while PTC and THCA showed the opposite results (Figure 2A). We combined 1489 DEGs in ATC and 522 DEGs we identified from GSE29265 and GSE33630. We finally obtained 222 hub genes. The heatmap showed good consistency between the changes in different samples and changes in colour levels observed in GSE33630 (Figure 1C), GSE29265 (Figure 1D) and GSE65144 (Figure 1E).

*KEGG pathway enrichment analysis*

To further investigate the functions of the DEGs, we classified DEGs into functional categories based on KEGG pathways to further investigate their functions. At a cut-off of p<0.001, the results showed that 'p53 signalling pathway' (KEGG: 04115) was significantly enriched in all THCA, PTC, and ATC samples, and it was also the most significant pathway enriched in PTC and THCA. 'MicroRNAs in cancer' (KEGG: 05206) was also significantly enriched in all THCA, PTC and ATC samples. Compared with THCA and PTC, the greatest enrich-

**Table 2.** Outcome of the STRING analysis

| *GO term* | *Description* | *Observed gene count* | *Background gene count* | *False discovery rate* |
|---|---|---|---|---|
| Biological Process (GO) | | | | |
| GO: 0007049 | cell cycle | 58 | 1263 | 1.55E-28 |
| GO: 0000278 | mitotic cell cycle | 43 | 628 | 1.49E-26 |
| GO: 0051301 | cell division | 37 | 483 | 4.57E-24 |
| GO: 1903047 | mitotic cell cycle process | 37 | 564 | 5.74E-22 |
| GO: 0022402 | cell cycle process | 43 | 890 | 3.50E-21 |
| Molecular Function (GO) | | | | |
| GO: 0008017 | microtubule binding | 13 | 253 | 2.20E-05 |
| GO: 0015631 | tubulin binding | 14 | 344 | 5.16E-05 |
| GO: 0008092 | cytoskeletal protein binding | 21 | 882 | 0.00019 |
| GO: 0035173 | histone kinase activity | 3 | 16 | 0.0274 |
| Cellular Component (GO) | | | | |
| GO: 0005819 | spindle | 30 | 322 | 1.33E-21 |
| GO: 0015630 | microtubule cytoskeleton | 42 | 1118 | 4.01E-17 |
| GO: 0000775 | chromosome, centromeric region | 19 | 189 | 5.05E-14 |
| GO: 0000922 | spindle pole | 17 | 150 | 2.39E-13 |
| GO: 0044430 | cytoskeletal part | 43 | 1547 | 2.39E-13 |
| KEGG Pathways | | | | |
| hsa04110 | cell cycle | 9 | 123 | 5.31E-05 |
| hsa04114 | oocyte meiosis | 7 | 116 | 0.0018 |
| hsa04914 | progesterone-mediated oocyte maturation | 6 | 94 | 0.0034 |
| hsa04115 | p53 signalling pathway | 5 | 68 | 0.0056 |
| hsa00670 | one carbon pool by folate | 3 | 20 | 0.0149 |

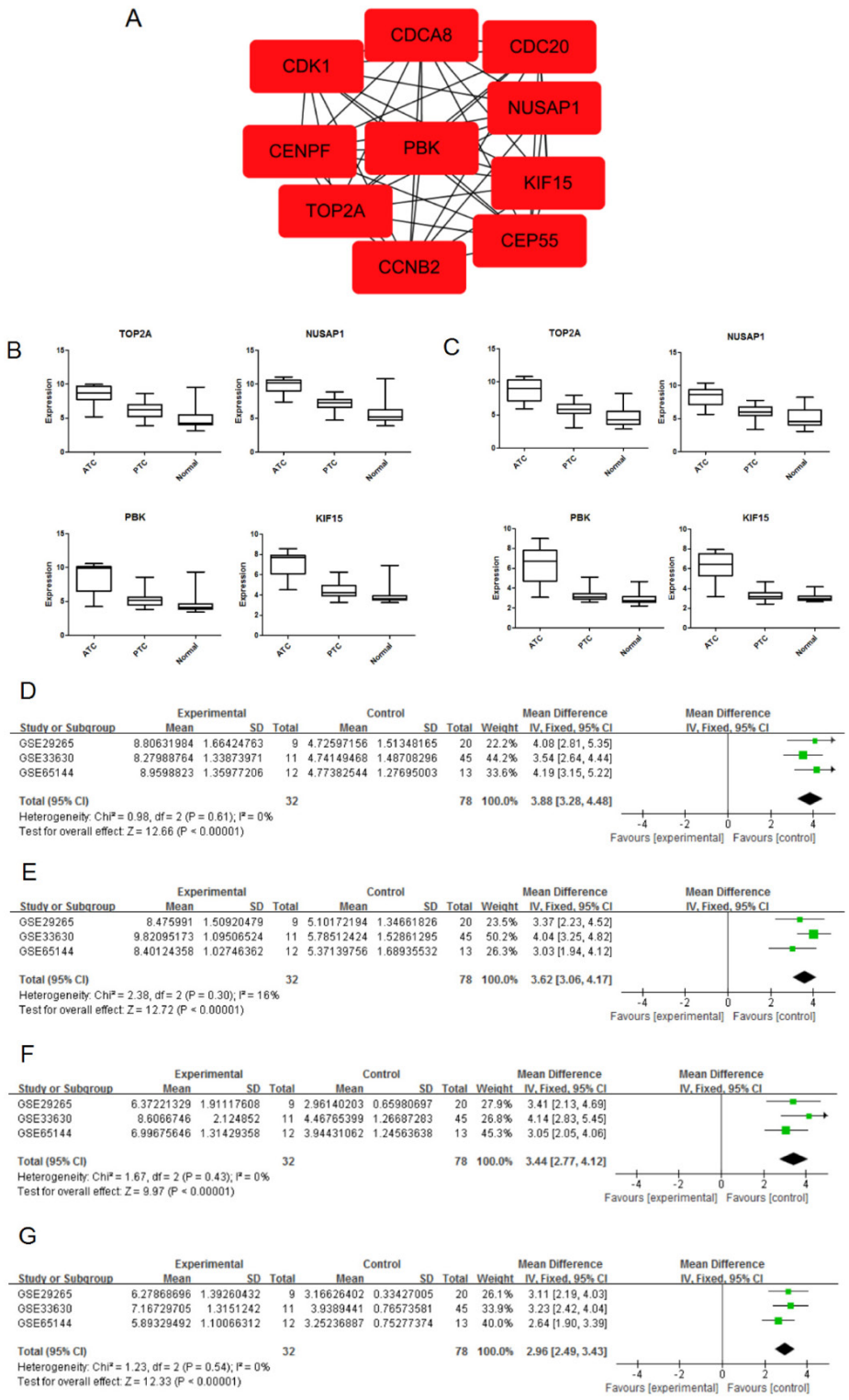
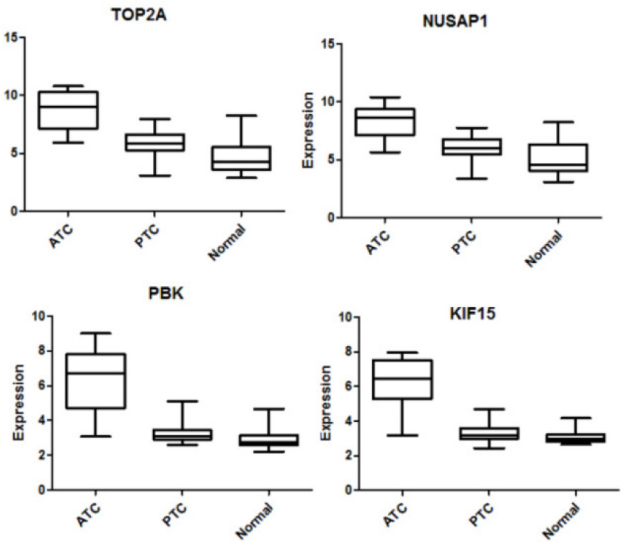



**Figure 3.** Expression of hub genes. **A:** Module obtained from the PPI network of DEGs using MCODE software. The colour of each gene that was closer to red indicates a higher degree, which was more effective. **B:** Box plots showing the expression of TOP2A, NUSAP1, PBK and KIF15 in ATC, PTC and normal tissues in GSE33630. The results revealed noticeably higher expression of TOP2A, NUSAP1, PBK, and KIF15 in the ATC samples than in PTC and normal samples. **C:** Forest chart of TOP2A expression in studies including ATC samples. The expression of the TOP2A gene is related to the poor prognosis of diverse tumours (HR=3.88, 95% CI=3.28-4.48, p<0.00001). **D:** The expression of the NUSAP1 gene is related to the poor prognosis of diverse tumours (HR=3.62, 95% CI=3.06-4.17, p<0.00001). **E:** The expression of the PBK gene is related to the poor prognosis of diverse tumours (HR=3.44, 95% CI=2.77-4.12, p<0.00001). **F:** The expression of the KIF15 gene is related to the poor prognosis of diverse tumours (HR=2.96, 95% CI=2.49-3.43, p<0.00001). **G:** The expression of HESS. TOP2A: DNA topoisomerase II alpha, NUSAP1: nucleolar and spindle-associated protein 1, PBK: PDZ binding kinase, KIF15: kinesin family member 15, CENPF: centromere protein F, CEP55: centrosomal protein 55, CDK1: cyclin-dependent kinase 1, CCNB2: cyclin B2, CDCA8: cell division cycle-associated 8, CDC20: cell division cycle 20, P: p value.

ment of pathways was observed in ATC, of which 'cell cycle' (KEGG: 04110) and 'pathways in cancer' (KEGG: 05200) were emphasized. The KEGG pathway enrichment analysis of ATC samples indicated that 'cell cycle' (KEGG: 04110) was the most significantly enriched pathway (p=2.32E-06, Figure. 2B). 'Pathways in cancer' (KEGG: 05200) in ATC included the largest number of DEGs (518 genes, Figure 2B).

We used outcome of PCA to verify the hub genes that we identified to determine whether they do contribute differentiating between ATC and PTC, normal samples. We drew scatter diagrams to visualize the outcome. Red represents ATC samples, green represents PTC samples, and blue represents normal samples. ATC samples were significantly separated from PTC and normal samples in GSE29265 (Figure 2C) and GSE33630 (Figure 2D), which means our hub gens contribute to distinguish ATC samples from PTC and normal samples. Figure 2E shows ATC samples were significantly separated from PTC samples in GSE65144.




**Supplementary Figure 2.** Box plots showing the expression of TOP2A, NUSAP1, PBK and KIF15 in ATC, PTC and normal tissues in GSE29265. Noticeably higher expression of TOP2A, NUSAP1, PBK and KIF15 was observed in the ATC samples than in PTC and normal samples.

*Hub gene and module screen of the PPI network*

We took the intersection of the DEGs identified in ATC and the DEGs identified in GSE29265 and GSE33630 and performed a STRING analysis using its multiple proteins analysis in Homo sapiens to identify the hub genes and perform a module analysis. According to the STRING database, 138 nodes and 1534 edges were included with PPI enrichment p values<1.0e-16 (Table 2).

After importing the STRING analysis to Cytoscape and performing the module analysis, we identified only one cluster (score: 47.922), with which we performed a cytohub analysis by calculating the degree. The top 10 genes (DNA topoi-

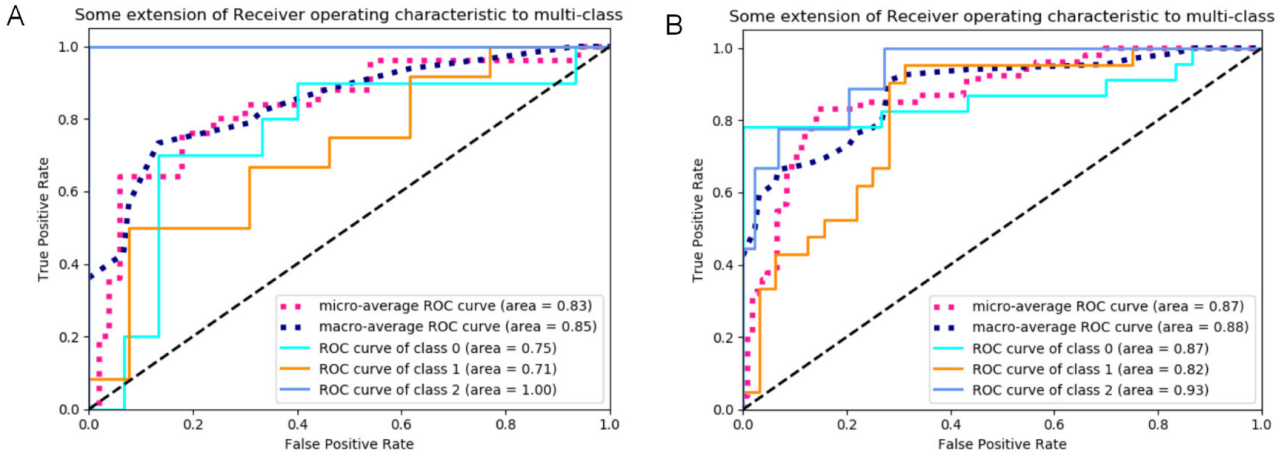



**Figure 4.** Individual ROC curves for each class and micro- and macro-averaged ROC curves were shown in the same figure. **A:** ROC curves of expressions of hub genes in GSE29265. For the individual classes of expression of hub genes in GSE29265, ROC AUC values were 0.75, 0.72, and 1 for normal, PTC, and ATC, respectively. The micro- and macro-averaged ROC values were 0.83 and 0.85. **B:** ROC curves of expressions of hub genes in GSE33630. ROC AUC values were 0.87, 0.82, and 0.93 for normal, PTC, and ATC, respectively. The micro- and macro-averaged ROC values were 0.87 and 0.88. Pink points represent micro- average ROC curve while blue points represent macro- average ROC curve. Azury lines represents ROC curve of class 0, orange represents that of class 1, and dark blue represents that of class 2. The area is closer to 1 represents hub genes are more meaningful. We can clearly conclude that in both GSE29265 and GSE33630, hub genes can distinguish ATC samples from PTC and normal samples.
ROC: Receiver Operating Characteristic. Class 0: normal samples versus ATC and PTC samples; class 1: PTC samples versus ATC and normal samples; Class 2: represents ATC samples versus normal and PTC samples.

somerase II alpha (TOP2A), nucleolar and spindle-associated protein 1 (NUSAP1), PDZ binding kinase (PBK), kinesin family member 15 (KIF15), centromere protein F (CENPF), centrosomal protein 55 (CEP55), cyclin-dependent kinase 1 (CDK1), cyclin B2 (CCNB2), cell division cycle-associated 8 (CDCA8) and cell division cycle 20 (CDC20) were identified as hub genes (Figure 3A). We chose the top 4 genes (TOP2A, NUSAP1, PBK, and KIF15) to draw box plots. Box plots show the result of these four genes in GSE33630 (Figure 3B) and GSE29265 (supplementary Figure 2). Noticeably higher expression levels of TOP2A, NUSAP1, PBK and KIF15 were observed in ATC tissues than in PTC tissues or normal tissues. By drawing a forest chart of TOP2A expression in studies including ATC samples (Figure 3C), our included studies showed good consistency. We used a fixed model because the number of included samples was not excessive and our studies displayed good homogeneity. The upper and lower bounds of 95% CIs for RRs were all greater than 1 and p values were far less than 0.00001, indicating that our study is valid. In addition, other forest charts of NUSAP1, PBK and KIF15 expression are explicitly shown in Figure 3D, 3E and 3F, respectively.

ROC curves typically feature true positive rate on the Y axis, and false positive rate on the X axis. This means that the top left corner of the plot is the ideal point whose false positive rate is zero and true positive rate is one. The resulting ROC curves are shown in Figure 4. Class 0 represents normal samples versus ATC and PTC samples, class 1 represents PTC samples versus ATC and normal samples, and Class 2 represents ATC samples versus normal and PTC samples. The area is closer to 1 implying hub genes are more meaningful. For the individual classes of expression of hub genes in GSE29265, ROC AUC values were 0.75, 0.72, and 1 for normal, PTC, and ATC, respectively. The micro- and macro-averaged ROC values were 0.83 and 0.85. In GSE33630, ROC AUC values were 0.87, 0.82, and 0.93 for normal, PTC, and ATC, respectively. The micro- and macro-averaged ROC values were 0.87 and 0.88. We can clearly conclude that in both GSE29265 and GSE33630, hub genes can distinguish ATC samples from PTC and normal samples.

## Discussion

Thyroid cancer includes a variety of pathological entities, mainly originating from follicles and parafollicular C cells in the thyroid gland [29]. By far the most common primary malignancy of the thyroid gland is the differentiated thyroid cancer [30].

ATC accounts for only 1 to 2 percent of all thyroid cancers, but it is one of the deadliest tumours in humans [31]. The tumorigenic process of the disease has not been completely elucidated [32]. We performed an integrative analysis of the 6 microarray datasets selected by dividing them into THCA, ATC and PTC to explain this phenomenon. We included 131 thyroid cancer samples and 98 normal thyroid tissue samples, including 32 ATC samples and 99 PTC samples. By performing a meta-analysis of the gene expression profiles using R statistical software, we identified a list of DEGs (up- or downregulated). These DEGs divided tumour samples from normal samples. Moreover, we performed a meta-analysis comparing GSE29265 and GSE33630 to identify DEGs that distinguished ATC from PTC samples. Then, we performed a pathway enrichment analysis; we speculated that ATC has more specific genes than PTC and THCA. Meanwhile, the cell cycle is a very important contributor to the poor prognosis of ATC and may be the key pathway that corresponds to the high mobility of ATC. After taking the intersection of the DEGs in ATC and the DEGs identified in GSE29265 and GSE33630 to perform the STRING analysis, we found the module with the most genes connected to the cell cycle. After the cytoHubba analysis in the Cytoscape app that calculated the degree, *TOP2A, NUSAP1, PBK, KIF, CENPF, CEP55, CDK1, CCNB2, CDC48* and *CDC20* were identified as hub genes.

In some human malignancies, such as non-small-cell lung cancer [33], hepatocellular carcinoma [34] and head and neck cancer [35], cell cycle-related proteins are overexpressed and contribute to the poor prognosis of patients. Seven hub genes (*CDK1, CCNB2, BUB1B, CDC20, RRM2, CHEK1* and *CDC45*) were selected as biomarkers of ATC, and the enrichment analysis showed that these genes were primarily accumulated in the 'cell cycle' in the study by Ding and colleagues [36]. *TOP2A, CDK1, CCNB1, VEGFA, BIRC5, MAPK1, CCNA2, MAD2L1, CDC20,* and *BUB1* were identified as hub genes of the PPI network in the study by Pan and colleagues of ATC [37].

Our result was very important because our hub genes not only distinguish between ATC samples and normal samples but also between ATC samples and PTC samples. By identifying biomarkers of ATC, our study contributes to guiding gene-targeted therapy for ATC, which is important for prolonging the survival of patients with ATC.

## Conflict of interests

The authors declare no conflict of interests.

## References


1. Fidler MM, Soerjomataram I, Bray F. A global view on cancer incidence and national levels of the human development index. Int J Cancer 2016;139:2436-46.
2. Cabanillas ME, McFadden DG, Durante C. Thyroid cancer. Lancet 2016; 388: 2783-95.
3. Saini S, Tulla K, Maker AV, Burman KD, Prabhakar BS. Therapeutic advances in anaplastic thyroid cancer: a current perspective. Mol Cancer 2018;17:154.
4. Incidence and Types of Thyroid Cancer: Papillary, Follicular, Medullary, and Anaplastic Thyroid Cancers. Available at: https:// www.endocrineweb.com/guides/thyroid-cancer/incidence-types-thyroid-cancer. htm. 2018.
5. Ain KB. Anaplastic thyroid carcinoma: a therapeutic challenge. Semin Surg Oncol 1999; 16:64-9.
6. Smallridge RC, Copland JA. Anaplastic thyroid carcinoma: pathogenesis and emerging therapies. Clin Oncol (R Coll Radiol) 2010;22:486-97.
7. Perri F, Lorenzo GD, Scarpati GD, Buonerba C. Anaplastic thyroid carcinoma: A comprehensive review of current and future therapeutic options. World J Clin Oncol 2011;2:150-7.
8. Deshpande HA, Roman S, Sosa JA. New targeted therapies and other advances in the management of anaplastic thyroid cancer. Curr Opin Oncol 2013;25:44-9.
9. Lowe NM, Loughran S, Slevin NJ, Yap BK. Anaplastic thyroid cancer: the addition of systemic chemotherapy to radiotherapy led to an observed improvement in survival--a single centre experience and review of the literature. Sci World J 2014;2014:674583.
10. Wei X, Cai S, Boohaker RJ et al. KAT5 promotes invasion and metastasis through C-MYC stabilization in ATC. Endocr Relat Cancer 2019;26:141-51.
11. Sugitani I, Kasai N, Fujimoto Y, Yanagisawa A. Prognostic factors and therapeutic strategy for anaplastic carcinoma of the thyroid. World J Surg 2001;25:617-22.
12. O'Neill JP, O'Neill B, Condron C, Walsh M, Bouchier-Hayes D. Anaplastic (undifferentiated) thyroid cancer: improved insight and therapeutic strategy into a highly aggressive disease. J Laryngol Otol 2005;119:585-91.
13. Molinaro E, Romei C, Biagini A et al. Anaplastic thyroid carcinoma: from clinicopathology to genetics and advanced therapies. Nat Rev Endocrinol 2017;13:644-60.
14. Enomoto K, Sato F, Tamagawa S et al. A novel therapeutic approach for anaplastic thyroid cancer through inhibition of LAT1. Sci Rep 2019;9:14616.
15. Hu S, Liao Y, Chen L. Identification of Key Pathways and Genes in Anaplastic Thyroid Carcinoma via Integrated Bioinformatics Analysis. Med Sci Monit 2018;24:6438-48.
16. Nagayama Y, Shigematsu K, Namba H, Zeki K, Yamashita S, Niwa M. Inhibition of angiogenesis and tumorigenesis, and induction of dormancy by p53 in a p53-null thyroid carcinoma cell line in vivo. Anticancer Res 2000;20:2723-8.
17. Moretti F, Farsetti A, Soddu S et al. p53 re-expression inhibits proliferation and restores differentiation of human thyroid anaplastic carcinoma cells. Oncogene 1997;14:729-40.
18. Fagin JA, Tang SH, Zeki K, Di Lauro R, Fusco A, Gonsky R. Reexpression of thyroid peroxidase in a derivative of an undifferentiated thyroid carcinoma cell line by introduction of wild-type p53. Cancer Res 1996;56:765-71.
19. Feng H, Cheng X, Kuang J et al. Apatinib-induced protective autophagy and apoptosis through the AKT-mTOR pathway in anaplastic thyroid cancer. Cell Death Dis 2018;9:1030.
20. Cao X, Dang L, Zheng X et al. Targeting Super-Enhancer-Driven Oncogenic Transcription by CDK7 Inhibition in Anaplastic Thyroid Carcinoma. Thyroid 2019;29:809-23.
21. Tiedje V, Stuschke M, Weber F, Dralle H, Moss L, Führer D. Anaplastic thyroid carcinoma: review of treatment protocols. Endocr Relat Cancer 2018;25:R153-61.
22. Yoshida A, Sugino K, Sugitani I, Miyauchi A. Anaplastic thyroid carcinomas incidentally found on postoperative pathological examination. World J Surg 2014;38:2311-6.
23. Ren D, Yang C, Liu N et al. Gene expression profile analysis of U251 glioma cells with SHRNA-mediated SOX9 knockdown. JBUON 2018;23:1136-48.
24. Ramasamy A, Mondry A, Holmes CC, Altman DG. Key issues in conducting a meta-analysis of gene expression microarray datasets. PLoS Med 2008;5:e184.
25. Zhao H, Li H. Network-based meta-analysis in the identification of biomarkers for papillary thyroid cancer. Gene 2018;661:160-8.
26. David CC, Jacobs DJ. Principal component analysis: a method for determining the essential dynamics of proteins. Methods Mol Biol 2014;1084:193-226.
27. Mi JX, Zhang YN, Lai Z, Li W, Zhou L, Zhong F. Principal Component Analysis based on Nuclear norm Minimization. Neural Netw 2019;118:1-16.
28. Szklarczyk D, Gable AL, Lyon D et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 2019; 47: D607-13.
29. Tsiambas E, Ragos N, Georgakopoulos G et al. E-C cadherin/a-Catenin deregulated co-expression in thyroid carcinoma based on tissue microarray digital image analysis. JBUON 2016;21:450-5.
30. Gkountouvas A, Nikas M, Chatjimarkou F, Thomas D, Georgiadis P, Kaldrimidis P. Thyroid cancer in Greece. A tertiary center experience. JBUON 2010;15:674-8.
31. Sugitani I, Onoda N, Ito KI, Suzuki S. Management of Anaplastic Thyroid Carcinoma: the Fruits from the ATC Research Consortium of Japan. J Nippon Med Sch 2018;85:18-27.
32. Ravi N, Yang M, Mylona N, Wennerberg J, Paulsson K. Global RNA Expression and DNA Methylation Patterns in Primary Anaplastic Thyroid Cancer. Cancers (Basel) 2020;12:680.
33. Li W, Zheng G, Xia J et al. Cell cycle-related and ex-

pression-elevated protein in tumor overexpression is associated with proliferation behaviors and poor prognosis in non-small-cell lung cancer. Cancer Sci 2018;109:1012-3.

34. Liu Z, Li J, Chen J et al. MCM family in HCC: MCM6 indicates adverse tumor features and poor outcomes and promotes S/G2 cell cycle progression. BMC Cancer 2018;18:200.
35. Wang R, Ma Z, Feng L et al. LncRNA MIR31HG targets HIF1A and P21 to facilitate head and neck cancer cell proliferation and tumorigenesis by promoting cell-cycle progression. Mol Cancer 2018;17:162.
36. Ding YG, Ren YL, Xu YS et al. Identification of key candidate genes and pathways in anaplastic thyroid cancer by bioinformatics analysis [published online ahead of print, 2020 Feb 17]. Am J Otolaryngol 2020; 102434.
37. Pan Z, Li L, Fang Q et al. Integrated Bioinformatics Analysis of Master Regulators in Anaplastic Thyroid Carcinoma. Biomed Res Int 2019;2019:9734576.