Introduction

Over the past 4 decades, the incidence of thyroid cancer has been increasing steadily in United States, with 14.42 cases per 100,000 persons per year in 2010–2013 (1). The most common histologic types were papillary thyroid cancer (PTC) (84%) and follicular thyroid cancer (11%). Thyroid cancer mortality rate rose by 1.1% annually from 0.40 per 100,000 person-years in 1994–1997 and to 0.46 per 100,000 person-years in 2010–2013 (1). Therefore, characterization of the molecular pathogenesis of thyroid cancer is critical to identifying key biomarkers and effective therapeutic targets for the disease.

Numerous studies have been conducted to characterize the genomic and transcriptomic landscapes of thyroid
characterize the expression profiles in 57 thyroid cancer and paired normal tissues (9). If genes showed false discovery rates (FDR) smaller than 0.05 and absolute log2fold-changes greater than 2, they were considered as DEGs. Next, principal component analysis (PCA) was conducted to evaluate whether DEGs could distinguish thyroid cancer from normal tissues.

**GO enrichment analysis**

To analyze the functional enrichment of DEGs and genes in co-expression modules, GO (6)-enrichment analysis was performed for all the DEGs and genes in co-expression modules with the online tool of GO (http://geneontology.org/). If the cutoff of FDR adjusted P value was smaller than 0.05, the enrichment of GO terms was considered to be significant.

**Construction of co-expression network in thyroid cancer**

Soft-thresholding power values were screened out in the construction of co-expression modules by the WGCNA algorithm (5). The gradient method was applied to evaluate the scale-free fit index and the mean connectivity degree of different co-expression modules with power values ranging from 1 to 20. The optimal power value was determined when the scale-free fit index was above 0.8, and then the construction of co-expression modules was performed by the WGCNA algorithm in R3.2.0. The minimum number of genes was set as 10 to ensure the high reliability of the results. The corresponding genetic information of co-expression modules was also extracted.

**Characterization of module-trait relationships in thyroid cancer**

Module-trait associations were analyzed using the correlation between the module eigengene and clinical traits. Gene ontology (GO) enrichment analysis was performed on the genes and the hub genes were identified in the modules of interest (6). These findings may be of importance in evaluating the malignant potential and clinical outcomes of thyroid cancer patients.
summing the connection strengths with other module genes and dividing this number by the maximum intramodular connectivity. The genes with maximum intramodular connectivity were regarded as intramodular hub genes (11).

The interested modules and hub genes were visualized by Integrative Visual Analysis Tool for Biological Networks and Pathways (VisANT) (12) software.

Results

DEGs analysis

The 57 thyroid cancer samples comprise 48 papillary, 6 follicular and 3 tall cell thyroid cancer samples. Twenty nine cancer samples had BRAFV600E mutations, while 28 samples were wild-type. Six and 51 cancer samples were RAS-mutant or wild-type respectively. Raw count data of thyroid cancer and normal samples were obtained from TCGA. DEGs were determined between 57 pairs of thyroid cancer and normal tissues with DEseq2 package in R. Overall, DEseq2 found 750 up-regulated and 296 down-regulated genes (Figure 1A). ARHGAP36, Tmprss6, Dmbx1, LOC400794, Gabrb2, Zcchc12, Klk6, Klk10, Cst2 and Grm4 were the top-ranking up-regulated genes, while Chga, Slc6a15, Ihh, Tspan19, Pmp2, Kcna1, Pcdh11x, Tff3, Gpr142 and C6orf176 were the top-ranking down-regulated genes in thyroid cancer (Table 1). In order to confirm the DEGs in thyroid cancer, we obtained RNA-seq data of 8 pairs of thyroid cancer and normal thyroid tissues from GEO. 77 genes were determined as DEGs by DESeq2, 64.81% (35/54) up-regulated and 52.17% (12/23) down-regulated genes were overlapped with the DEGs obtained from 57 pairs of thyroid cancer and normal counterparts (Figure S1). Next, PCA was conducted to assess whether DEGs could distinguish thyroid cancer from normal tissues. Cancer and normal tissues were aggregated to the lower and upper sides respectively, demonstrating that DEGs were able to classify the samples into two distinct subsets (Figure 1B).

GO term enrichment analysis

The enrichment of GO terms was analyzed for 750 up-regulated and 296 down-regulated genes on the home page of GO. GO term enrichment analysis showed down-regulated genes were significantly enriched in 136 biological processes, up-regulated genes in 281 biological processes (FDR adjusted P value <0.05). The main GO biological process terms for down-regulated genes showed a wide variety of functional processes ranging from cellular developmental process, cell differentiation, single-multicellular organismal process, cell development and regulation of system process. While the primary GO terms for up-regulated genes were related to various cellular processes, such as negative regulation of endopeptidase activity, cell-cell adhesion, epidermis...
development, extracellular matrix organization, regulation of protein phosphorylation, regulation of signal transduction, regulation of cell proliferation and a variety of metabolic processes.

Table 1 The top 10 down-regulated and up-regulated genes in thyroid cancer

<table>
<thead>
<tr>
<th>Gene</th>
<th>Fold change</th>
<th>P value</th>
<th>FDR-adjusted P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Down-regulated</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHGA</td>
<td>4.80</td>
<td>1.55×10^{-15}</td>
<td>1.48×10^{-14}</td>
</tr>
<tr>
<td>SLC6A15</td>
<td>4.61</td>
<td>5.82×10^{-41}</td>
<td>4.16×10^{-39}</td>
</tr>
<tr>
<td>IHH</td>
<td>4.54</td>
<td>1.08×10^{-34}</td>
<td>5.03×10^{-33}</td>
</tr>
<tr>
<td>TSPAN19</td>
<td>4.42</td>
<td>4.05×10^{-51}</td>
<td>5.56×10^{-49}</td>
</tr>
<tr>
<td>PMP2</td>
<td>4.18</td>
<td>2.46×10^{-38}</td>
<td>1.46×10^{-36}</td>
</tr>
<tr>
<td>KCNA1</td>
<td>4.05</td>
<td>1.17×10^{-27}</td>
<td>3.24×10^{-26}</td>
</tr>
<tr>
<td>PCDH11X</td>
<td>3.91</td>
<td>2.43×10^{-31}</td>
<td>8.81×10^{-30}</td>
</tr>
<tr>
<td>TFF3</td>
<td>3.88</td>
<td>1.24×10^{-34}</td>
<td>5.71×10^{-33}</td>
</tr>
<tr>
<td>GPR142</td>
<td>3.86</td>
<td>4.25×10^{-45}</td>
<td>4.04×10^{-43}</td>
</tr>
<tr>
<td>C6orf176</td>
<td>3.79</td>
<td>4.10×10^{-24}</td>
<td>8.52×10^{-23}</td>
</tr>
<tr>
<td><strong>Up-regulated</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARHGAP36</td>
<td>-7.53</td>
<td>8.54×10^{-75}</td>
<td>5.22×10^{-72}</td>
</tr>
<tr>
<td>TMPRSS6</td>
<td>-7.22</td>
<td>1.15×10^{-88}</td>
<td>2.42×10^{-85}</td>
</tr>
<tr>
<td>DMBX1</td>
<td>-7.17</td>
<td>2.47×10^{-58}</td>
<td>4.72×10^{-56}</td>
</tr>
<tr>
<td>LOC400794</td>
<td>-7.05</td>
<td>2.04×10^{-90}</td>
<td>4.83×10^{-87}</td>
</tr>
<tr>
<td>GABR2</td>
<td>-6.86</td>
<td>6.29×10^{-121}</td>
<td>1.19×10^{-116}</td>
</tr>
<tr>
<td>ZCCHC12</td>
<td>-6.72</td>
<td>1.34×10^{-87}</td>
<td>2.31×10^{-84}</td>
</tr>
<tr>
<td>KLK6</td>
<td>-6.58</td>
<td>4.42×10^{-43}</td>
<td>3.58×10^{-41}</td>
</tr>
<tr>
<td>KLK10</td>
<td>-6.57</td>
<td>6.60×10^{-94}</td>
<td>2.50×10^{-90}</td>
</tr>
<tr>
<td>CST2</td>
<td>-6.51</td>
<td>2.79×10^{-43}</td>
<td>2.31×10^{-41}</td>
</tr>
<tr>
<td>GRM4</td>
<td>-6.43</td>
<td>1.23×10^{-87}</td>
<td>2.31×10^{-84}</td>
</tr>
</tbody>
</table>

**Construction of co-expression modules in thyroid cancer**

Among the 568 thyroid cancer samples in TCGA, 501 had both RNA-sequencing expression and clinical traits data. Therefore, expression values of 1,046 genes in the 501 samples of thyroid cancer were used to construct the co-expression modules by the WGCNA package. One of the most important parameters was the soft-thresholding power value, which mainly affected the scale-free fit index and the average connectivity degree of co-expression modules. Firstly, the power value was screened out (Figure S2). When the power value was equal to seven, the scale-free fit index was larger than 0.8 and the mean connectivity degree was higher. Therefore, the soft-thresholding power value used to construct co-expression modules and the WGCNA identified five distinct gene co-expression modules which are shown in different colors (Figure 2). These numbers of genes in the five modules differed largely, with 608, 293, 70, 57 and 19 genes for the turquoise, grey, blue, brown and yellow modules respectively (Figure 2).
Module-trait association analysis in thyroid cancer

The clinical traits data were obtained from the TCGA database. Module-trait associations were analyzed with the correlation between module eigengene and clinical traits. As showed in the Figure 3, the turquoise module was negatively correlated with patients’ age, cancer stage and multifocality (P value <0.05 for all cases, Figures 2 and 3). In contrast, the blue module was significantly positively associated with patients’ age, cancer stage and multifocality (P value <0.05 for all cases, Figures 2 and 3). The brown module was negatively correlated with tumor size (P value =0.04, Figures 2 and 3). Then, we plotted scatterplots of GS for patients’ age, cancer stage and multifocality vs. MM in the blue, turquoise and brown modules. The GS for patients’ age, cancer stage and multifocality was positively correlated with MM in the blue and turquoise modules (P value <0.05 for all cases, Figure 4). However, the GS for tumor size showed non-significant correlation with MM in the brown module (P=0.81, Figure S3).

Functional enrichment analysis of genes in the interested modules

GO enrichment analysis was performed on the genes in the turquoise, blue and brown modules. There was large difference in the biological processes in which the three modules were enriched. Genes in turquoise module were significantly over-represented in 171 GO terms (FDR adjusted P value <0.05), such as cell migration (GO:0016477), regulation of cell proliferation (GO:0042127), regulation of cell differentiation (GO:0045595), regulation of cellular protein metabolic process (GO:0032268), signal transduction (GO:0007165). While genes in the turquoise module were significantly
under-represented in 42 GO terms, such as gene expression (GO:0010467), DNA repair (GO:0006281), DNA metabolic process (GO:0006259) and cellular response to DNA damage stimulus (GO:0006974). However, no significant enrichment of GO terms was found for the genes in the blue and brown modules.

Figure 4 The correlations between GS of clinical traits and MM in the blue and turquoise modules of thyroid cancer. Scatterplots of GS for patients’ age vs. MM in the blue (A) and turquoise (B) modules. Scatterplots of GS for cancer stage vs. MM in the blue (C) and turquoise (D) modules. Scatterplots of GS for multifocality vs. MM in the blue (E) and turquoise (F) modules. GS and MM refer to gene significance and module membership respectively. GS, gene significance; MM, module membership.
Visualization of the interested modules and hub genes

The modules were constructed using the VisANT software. The genes with largest intramodular connectivity were considered as intramodular hub genes. The hub genes in the turquoise and blue modules were bold with red in Figure 5. SERPINA1 and MRO were the hub genes in the turquoise and blue modules respectively (Figure 5).

Discussion

RNA-seq analyses have previously been used to predict therapeutic biomarkers and to identify gene expression patterns associated with clinical traits in thyroid cancer (13,14). This study presented a large number of DEGs between 57 pairs of thyroid cancer and normal tissues. By comparing the DEGs to the tumor associated gene and tumor suppressor databases (15), five known oncogenes were up-regulated, including ELF3, HMG2A2, LCN2, MET and RUNX2; by contrast, 10 TSGs were found down-regulated, including EGR2, FOXA2, GPC3, IGFPL1, LRP1B, MT1G, NR4A3, PROX1, TMEFF2 and WNT11. Additionally, some new DEGs were discovered, such as ARHGA36, TMPRSS6, DMBX1, CHGA and SLC6A15, which have not been reported previously and their functions remain unknown in thyroid cancer.

WGCNA is a method frequently used to explore the co-expression modules of highly correlated genes. Genes in the same module were considered to be related with each other in function. Therefore, the analysis allowing for identification of biologically-relevant modules and hub genes that may eventually serve as biomarkers for detection or treatment (5). In this study, a total of five co-expression modules were constructed by the 1,046 genes from the 57 pairs of thyroid cancer and normal samples by the WGCNA method. We identified the blue and turquoise co-expression modules that relate to clinical traits (patient’s age, cancer stage and multifocality, Figures 2–4). The result of functional enrichment analysis showed that the turquoise module was found to be mainly enriched in cellular processes associated with cell migration, regulation of cell proliferation, regulation of cell differentiation, and regulation of cellular protein metabolic process. While genes in the blue module were not significantly enriched in any GO terms. Therefore, we speculate that turquoise module is the most important module in the cancer stage and multifocality of thyroid cancer.

The turquoise and blue co-expression modules were constructed by the VisANT software and the hub genes SERPINA1 and MRO were identified (Figure 5). The SERPINA1, also known as α1-Antitrypsin, is a protease inhibitor that can act on a variety of targets such as serine proteases. It has been demonstrated that SERPINA1 expression can be stimulated by E2 in MCF-7 cells, and a high expression of this protein inhibits colony formation (16). SERPINA1 is up-regulated in colorectal cancer (17), glioma (18), cutaneous squamous cell carcinoma (19) and thyroid cancer (20). SERPINA1 has been proposed as a biomarker for various cancers, such as papillary thyroid carcinoma (20), lung cancer (21) and breast carcinoma (22-24). The high expression of SERPINA1 indicates a poor prognosis.

![Figure 5](https://example.com/figure5.png)
of high-grade glioma (18) but a favorable clinical outcome in breast cancer patients (25). The hub gene of the blue module is MRO, which is down-regulated in lung cancer cell lines (26). The hub genes in the turquoise and blue modules are either oncogenes or anti-oncogenes. Therefore, we speculate that the hub genes are critical in the cancer stage and multifocality of thyroid cancer.

Conclusions
Taken together, the turquoise module was regarded as the most critical module in the cancer stage and multifocality of thyroid cancer. More importantly, the hub gene SERPINA1 may function as a potential biomarker for thyroid cancer, which also needs much further research.

Acknowledgements
None.

Footnote
Conflicts of Interest: The authors have no conflicts of interest to declare.

References
20. Vierlinger K, Mansfeld MH, Koperek O, et al. Identification of SERPINA1 as single marker for papillary thyroid carcinoma through microarray meta analysis and


**Figure S1** The overlap of DEGs between 57 (TCGA) and 8 (GEO) pairs of thyroid cancer and normal thyroid tissues. (A) The overlap of down-regulated genes between the two datasets; (B) the overlap of up-regulated genes between the two datasets. DEGs, differentially expressed genes.

**Figure S2** Analysis of network topology for various soft-thresholding powers. The left panel shows the scale-free fit index (y-axis) as a function of the soft-thresholding power (x-axis). The right panel displays the mean connectivity (degree, y-axis) as a function of the soft-thresholding power (x-axis).

**Figure S3** A scatterplot of GS for tumor size vs. MM in the brown module. GS, gene significance; MM, module membership.