Thyroid carcinoma poses a heavy disease burden on the global health. The potential influence of some differentially expressed genes on the progression of thyroid carcinoma was still incomplete. In this study, we integrated transcriptome data with clinical data to investigate the relationship between them in thyroid cancer patients. First, the gene expression profile (GSE129562) from Gene Expression Omnibus was used to identify differentially expressed genes. Secondly, gene ontology function and Kyoto encyclopedia of genes and genomes pathway enrichment analyses of the differentially expressed genes were performed. Thirdly, protein-protein interactions of the differentially expressed genes were constructed. Between T1aN1b or T3N1b thyroid carcinoma and its paired thyroid normal tissue, a total of 729 differentially expressed genes were identified through our analysis, of which 405 genes were up-regulated and 324 genes were down-regulated. Of those differentially expressed genes, 138 differentially expressed genes were identified as immune-related genes. Their functions can be classified as antigen processing and presentation, antimicrobials, B cell receptor signaling pathway, chemokine receptors and so on. Gene ontology enrichment analysis illustrated that these differentially expressed genes were mostly enriched into extracellular matrix structural constituent, integrin binding and so on. Kyoto encyclopedia of genes and genomes enrichment analysis illustrated that these differentially expressed genes were mostly enriched into extracellular matrix-receptor interaction, focal adhesion and so on. The top 4 differentially expressed genes with the highest degrees in the protein-protein network are syndecan 1, tyrosine-protein kinase Fyn, decorin and thrombospondin-1. The mutations in the gene solute carrier family 34 member 2 and tenascin C are arguably the most common. Our results demonstrate the potential influence of some differentially expressed genes on the progression of thyroid cancer, provides a comprehensive bioinformatics analysis of the pathogenesis, which may contribute to future investigation into the molecular mechanisms and biomarkers.

Key words: Thyroid carcinoma, differentially expressed genes, gene ontology, Kyoto encyclopedia of genes and genomes enrichment analysis

Thyroid carcinoma is the most common thyroid malignant tumor, accounting for about 1 % of the malignant tumors in the whole body, including papillary carcinoma, follicular carcinoma, undifferentiated carcinoma and myeloid carcinoma[1,2]. Papillary carcinoma with low malignancy and good prognosis is the most common. Except for myeloid carcinoma, most thyroid cancers originate from follicular epithelial cells. The incidence rate is related to region, race and gender[3]. Women are more likely to suffer from the disease and the proportion of men and women is 1:(2–4). It can occur at any age, but it is more common in young adults. The vast majority of thyroid cancer occurs in one thyroid gland lobe, often as a single tumor[4].

Iodine is an essential trace element in human body. Iodine deficiency leads to decreased thyroid hormone synthesis and increased Thyroid Stimulating Hormone (TSH) level, which stimulates hyperplasia and hypertrophy of thyroid follicles, resulting in goiter, changes in thyroid hormones and increased incidence of thyroid cancer[5]. There are still conflicting opinions. A high iodine diet may increase the incidence of papillary thyroid cancer[6,7]. X-ray irradiation of the thyroid of experimental mice can induce thyroid cancer in animals and the nucleus is deformed, and the synthesis of Thyroxine (T4) is greatly reduced, leading to canceration. On the other hand, the thyroid gland is damaged and cannot produce endocrinins and the resulting TSH secretion can also promote thyroid cell canceration[8]. Increased serum
TSH level, induction of nodular goiter, mutagen and TSH stimulation can induce thyroid follicular carcinoma, and clinical studies have shown that TSH suppression therapy in the treatment of differentiated thyroid cancer surgery play an important role in the process, but if TSH stimulation is the cause of thyroid carcinoma remains to be proved\textsuperscript{[9,10]}. There are no obvious symptoms and signs in the early stage, but small thyroid masses are usually found by thyroid palpation and neck ultrasound during physical examination\textsuperscript{[11]}. Typical clinical manifestations are found in thyroid mass, hard and fixed texture, uneven surface is the common manifestation of all types of cancer\textsuperscript{[12]}. The glands have little movement up and down during swallowing. Undifferentiated carcinoma may prevent these symptoms in a short period of time and in addition to the obvious growth of the mass, it is accompanied by the characteristic of invading surrounding tissues\textsuperscript{[13]}. In the late stage, hoarseness, dysphagia and sympathetic nerve compression can cause Horner syndrome, pain in ear, occiput and shoulder and metastasis of local lymph nodes and distant organs. Cervical lymph node metastasis occurs earlier in undifferentiated carcinoma. Myeloid cancer can cause diarrhea, heart palpitations and flushing due to the production of calcitonin and serotonin by the tumor itself\textsuperscript{[14]}. Surgical treatment for thyroid cancer is the main treatment method, including thyroid surgery and neck dissection\textsuperscript{[15]}. Surgical resection should be performed whenever possible, regardless of pathological type, as long as surgery is indicated. For well differentiated papillary carcinoma or follicular carcinoma, the local recurrence after surgery can also be treated again. The scope of thyroid resection is related to the pathological type and stage of the tumor\textsuperscript{[16]}. The smallest scope is glandular and isthmus resection, and the largest is total thyroidectomy. Patients undergoing subtotal or total thyroidectomy should take thyroxine tablets for life to prevent hypothyroidism and inhibit TSH\textsuperscript{[17]}. Both papillary adenocarcinoma and follicular adenocarcinoma have TSH receptors and TSH affects the growth of thyroid cancer through its receptor. Dry thyroid tablets or levothyroxine are generally used and the plasma T4 and TSH levels should be measured regularly to adjust the drug dose, so that the thyroid hormone in the body can be maintained at a level slightly higher than normal but lower than hyperthyroidism\textsuperscript{[18]}. In the present study, the differentially expression of critical genes play a key role in the mechanism of common development of the thyroid carcinoma and will affect therapy as well as the efficacy of medicine. The relationship between Differentially Expressed Genes (DEGs) and the progression of thyroid carcinoma were still demanded to be explained. The sharing of transcriptome data and the development of new bioinformatics analysis tools has enabled us to integrate transcriptome data with clinical data to investigate the relationship between them. This can help us to understand the development of thyroid carcinoma from both perspectives and provide a new perspective for the prevention and treatment of the disease.

**MATERIALS AND METHODS**

**Data:**

The gene expression profiles (GSE129562), which are composed of 16 samples (3 T1aN0 and 5 T1aN1b or T3N1b thyroid carcinoma) and its paired thyroid normal tissue, were downloaded from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) and exploited as discovery datasets to identify DEGs.

**Identification of DEGs:**

The Limma package of R was used to identify the DEGs\textsuperscript{[19]}. The adjusted p-values (adj p-value) were calculated instead to avoid the appearance of false-positive results. Genes with |log2 Fold Change (FC)| larger than 1 and adj p-value<0.01 were taken as DEGs between thyroid carcinoma and its paired thyroid normal tissue samples. The relevant immune genes were searched in the Immunology Database and Analysis Portal (ImmPort) (https://www.immport.org/resources) to find potential immunotherapy targets.

**Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis:**

GO function and KEGG pathway enrichment analyses of the DEGs were performed using clusterProfiler and pathview packages of R, which are designed for automating the process of biological-term classification and the enrichment analysis of gene clusters\textsuperscript{[20]}.

**Protein-Protein Interaction (PPI) network construction:**

PPIs of the DEGs were constructed based on the
Search Tool for the Retrieval of Interacting Genes (STRING; http://string.embl.de/) with the confidence score $\geq 0.9^{[21]}$. Subsequently, the PPI network was visualized by means of Cytoscape software (version 3.5.1). Furthermore, the plug-in of Molecular Complex Detection (MCODE)$^{[22]}$ in Cytoscape software was applied to explore the significant modules in PPI network with default settings.

RESULTS AND DISCUSSION

The gene expression profiles (GSE129562) were used to identify DEGs, which are composed of 16 samples (3 T1aN0 and 5 T1aN1b or T3N1b thyroid carcinoma) and its paired thyroid normal tissue.

Between T1aN0 thyroid carcinoma and its paired thyroid normal tissues, a total of 9 DEGs were identified through our analysis, of which 8 genes were up-regulated and 1 gene was down-regulated (fig. 1A).

Between T1aN1b or T3N1b thyroid carcinoma and its paired thyroid normal tissue, a total of 729 DEGs were identified through our analysis, of which 405 genes were up-regulated and 324 genes were down-regulated (fig. 1B and fig. 2). Of those 729 DEGs, 138 DEGs were identified as immune-related genes (Table 1). Their functions can be classified as antigen processing and presentation, antimicrobials, B Cell Receptor (BCR) signaling pathway, chemokine receptors, chemokines, cytokine receptors, cytokines, interleukins, interleukins receptor, natural killer cell cytotoxicity and so on.

Functional enrichment analysis of DEGs was explained here. Between T1aN0 thyroid carcinoma and its paired thyroid normal tissue, GO enrichment analysis illustrated that these DEGs were enriched in several terms (fig. 3A), including Extracellular Matrix (ECM) structural constituent, integrin binding, proteoglycan binding, ECM binding, ECM structural constituent conferring tensile strength, growth factor binding, fibronectin binding, glycosaminoglycan binding, sulfur compound binding, virus receptor activity, exogenous protein binding, cytokine activity, platelet-derived growth factor binding, chemokine activity, laminin binding, low-density lipoprotein particle receptor activity, collagen binding, peptidase regulator activity, receptor ligand activity, signaling receptor activator activity, heparin binding, lipoprotein particle receptor activity, heparan sulfate proteoglycan binding, enzyme inhibitor activity, lipid transporter activity, lipoprotein particle binding, protein-lipid complex binding, retinol dehydrogenase activity, protease binding, chemokine receptor binding, serine-type peptidase activity, oxidoreductase activity acting on the aldehyde or oxo group of donors, serine hydrolase activity, cytokine binding, cargo receptor activity, transmembrane receptor protein tyrosine kinase activity, transmembrane receptor protein kinase activity, serine-type endopeptidase activity, peptidase inhibitor activity, Nicotinamide Adenine Dinucleotide (NAD$^+$) nucleosidase activity, Nicotinamide Adenine Dinucleotide Phosphate (NAD(P)$^+$) nucleosidase activity, NAD$^+$ nucleotidase, cyclic Adenosine Diphosphate (ADP)-ribose generating, G protein-coupled receptor binding, cytokine receptor binding, endopeptidase regulator activity, transmembrane-ephrin receptor activity, ephrin receptor activity, Wnt-protein binding and endopeptidase inhibitor activity. KEGG enrichment analysis illustrated that these DEGs were enriched in several pathways (fig. 3B). The enriched pathways were ECM-receptor interaction, focal adhesion, malaria, cytokine-cytokine receptor interaction, protein digestion and absorption, human papillomavirus infection, small cell lung cancer, cholesterol metabolism, p53 signaling pathway, proteoglycans in cancer, retinol metabolism, Phosphatidylinositol-3-Kinase (PI3K)-Protein Kinase B (Akt) signaling pathway, cell adhesion molecules, viral myocarditis, complement and coagulation cascades, Advanced Glycation End Products-Receptor for AGE (AGE-RAGE) signaling pathway in diabetic complications and mineral absorption.

STRING was used to construct the PPI network with high confidence 0.7 and disconnected nodes were hided. The most significant modules in PPI network were identified by using Cytoscape software with default settings. The regulatory network of DEGs between T1aN1b or T3N1b thyroid carcinoma and its paired thyroid normal tissue is complex with 704 nodes and 679 edges (fig. 4A and fig. 4B). The average degree of the network is 1.93 and the average local clustering coefficient is 0.294. The gene Syndecan (SDC) 1 has the largest degree followed by Tyrosine-Protein Kinase Fyn (FYN), Decorin (DCN) and so on. A significant module was identified by MCODE with 13 nodes and 78 edges (Table 2). The module consists of 13 DEGs, including SDC4, Collagen Type I Alpha (COL1A2) 2 chain, Glypican 3 (GPC3), SDC2, Beta-1,3-Glucuronyltransferase 1 (B3GAT1), DCN, Versican...
(VCAN), COL1A1, COL5A2, COL3A1, COL4A1, Prolyl 4-Hydroxylase Subunit Alpha 2 (P4HA2), COL5A1. Of the 13 DEGs, COL5A1 is the seeded genes and has the largest degree. The total of 729 DEGs obtained from the previous step was retrieved in The Cancer Genome Atlas (TCGA) database. Finally, only 23 relevant cases were identified with survival data (fig. 5A and fig. 5B). The mutations of Solute Carrier Family 34 Member 2 (SLC34A2) and Tenascin C (TNC) account for the most and the following is ETS-related Transcription Factor 4 (ELF4), COL1A1, KIT Proto-Oncogene, Receptor Tyrosine Kinase (KIT) and so on. So, mutations in the gene SLC34A2 and TNC are arguably the most common.

In the present study, the DEGs between T1aN0 thyroid carcinoma and its paired thyroid normal tissue, T1aN1b or T3N1b thyroid carcinoma and its paired thyroid normal tissue were explored. This study is meaningful since transcriptome data was integrated to investigate the potential pathogenesis of DEGs between controls and thyroid carcinoma. This study provides a reference for understanding the pathogenesis value of DEGs and formulating reasonable clinical diagnosis and treatment.

Between T1aN0 thyroid carcinoma and its paired thyroid normal tissue, a total of 9 DEGs were identified through our analysis, of which 8 genes were up-regulated and 1 gene was down-regulated (fig. 1A).

Between T1aN1b or T3N1b thyroid carcinoma and its paired thyroid normal tissue, a total of 729 DEGs were identified through our analysis, of which 405 genes were up-regulated and 324 genes were down-regulated (fig. 1B and fig. 2). Of those 729 DEGs, 138 DEGs were identified as immune-related genes (Table 1). Their functions can be classified as antigen processing and presentation, antimicrobials, BCR signaling pathway, chemokine receptors and so on. Between T1aN0 thyroid carcinoma and its paired thyroid normal tissue, GO enrichment analysis illustrated that these DEGs were mostly enriched into ECM structural constituent, integrin binding and so on (fig. 3A). KEGG enrichment analysis illustrated that these DEGs were mostly enriched into ECM-receptor interaction, focal adhesion and so on (fig. 3B).

![Volcano plots of DEGs between T1aN0 thyroid carcinoma and its paired thyroid normal tissue, (A) T1aN1b or T3N1b thyroid carcinoma and its paired thyroid normal tissue and (B) In GSE129562, ( ) Up-regulated genes; ( ) Down-regulated genes and ( ) No change](image-url)
Fig. 2: Heatmap plots of DEGs between T1aN1b or T3N1b thyroid carcinoma and its paired thyroid normal tissue in GSE129562

Fig. 3: (A) The enriched GO terms of DEGs between T1aN1b or T3N1b thyroid carcinoma and its paired thyroid normal tissue in GSE129562 and (B) The enriched KEGG pathways of DEGs between T1aN1b or T3N1b thyroid carcinoma and its paired thyroid normal tissue in GSE129562
Fig. 4: (A) PPI network of DEGs between T1aN1b or T3N1b thyroid carcinoma and its paired thyroid normal tissue in GSE129562 and (B) The top 30 DEGs with the highest degree in the PPI network.

Table 1: Immune-related DEGs between T1aN1b or T3N1b thyroid carcinoma and its paired thyroid normal tissue in GSE129562

<table>
<thead>
<tr>
<th>Category</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antigen processing and</td>
<td>CD1A, CTSB, CTSS, HLA-DRB3, HLA-DRB4, HLA-G, HSPA5, ICAM1, LGMN, THBS1</td>
</tr>
<tr>
<td>presentation</td>
<td></td>
</tr>
<tr>
<td>Antimicrobials</td>
<td>CXCL14, CXCL16, SLPI, CXCL12, CCL13, LCN2, S100A10, S100A11, S100A1, S100A16, PLAU, CTSS, NOX4, FABP4, CRABP1, APOD, BCL3, GDF15, CCL20, ZYX, ISG15, LYZ, NOD1, SLC11A1, TNFRSF10B, CCL28, SYTL1, TRIM22, ABCC4, GBP2, CCL15, CCL13, CCL28, CCL21, CCL20, CXCL12, OLR1</td>
</tr>
<tr>
<td>BCR signaling pathway</td>
<td>LYN</td>
</tr>
<tr>
<td>Chemokines</td>
<td>CCL13, CCL15, CCL20, CCL21, CCL28, CX3CL1, CXCL12, CXCL14, CXCL16, EDN3, PLAU, TNC</td>
</tr>
</tbody>
</table>

Fig. 5: (A) Distribution of most frequently mutated genes in TCGA database and (B) Overall survival plot of the most frequently mutated genes in TCGA database.
The top 4 DEGs with the highest degrees in the protein-protein network are SDC1, FYN, DCN and Thrombospondin-1 (THBS1).

The gene SDC1 encodes a transmembrane (type I) heparan sulfate proteoglycan and is a member of the SDC proteoglycan family. The SDCs mediate cell binding, cell signaling and cytoskeletal organization and SDC receptors are required for internalization of the Human Immunodeficiency Virus type 1 (HIV-1) Transactivator of Transcription (Tat) protein. The SDC1 protein functions as an integral membrane protein and participates in cell proliferation, cell migration and cell-matrix interactions via its receptor for ECM proteins. Altered SDC1 expression has been detected in several different tumor types. While several transcript variants may exist for this gene, the full-length natures of only two have been described to date. These two represent the major variants of this gene and encode the same protein. This gene FYN is a member of the protein-tyrosine kinase oncogene family. It encodes a membrane-associated tyrosine kinase that has been implicated in the control of cell growth. The protein associates with the p85 subunit of phosphatidylinositol 3-kinase and interacts with the FYN-binding protein. Alternatively spliced transcript variants encoding distinct isoforms exist. The protein encoded by the gene DCN is a member of the small leucine-rich proteoglycan family of proteins. Alternative splicing results in multiple transcript variants, where at least one of this encodes a preproprotein, that is proteolytically processed to generate the mature protein. This protein plays a role in collagen fibril assembly. Binding of this protein to multiple cell surface receptors mediates its role in tumor suppression, including a stimulatory effect on autophagy and inflammation and an inhibitory effect on angiogenesis and tumorigenesis. This gene and the related gene biglycan are thought to be the result of gene duplication. Mutations in this gene are associated with congenital stromal corneal dystrophy in human patients. The gene THBS1 encodes a subunit of a disulfide-linked homotrimeric protein. This protein is an adhesive glycoprotein that mediates cell-to-cell and cell-to-matrix interactions. This protein can bind to fibrinogen, fibronectin, laminin, type V collagen and integrin’s alpha-V/beta-1. This protein has been shown to play roles in platelet aggregation, angiogenesis and tumorigenesis.
The most significant module was identified by MCODE with 13 nodes and 78 edges (Table 2). The module consists of 13 DEGs, including SDC4, COL1A2, GPC3, SDC2, B3GAT1, DCN, VCAN, COL1A1, COL5A2, COL3A1, COL4A1, P4HA2 and COL5A1. Of the 13 DEGs, COL5A1 is the seeded genes and has the largest degree. The gene COL5A1 encodes an alpha chain for one of the low abundance fibrillar collagens[27]. Fibrillar collagen molecules are trimers that can be composed of one or more types of alpha chains. Type V collagen is found in tissues containing type I collagen and appears to regulate the assembly of heterotypic fibers composed of both type I and type V collagen. This gene product is closely related to type XI collagen and it is possible that the collagen chains of types V and XI constitute a single collagen type with tissue-specific chain combinations. The encoded procollagen protein occurs commonly as the heterotrimer, pro-alpha 1 (V)-pro-alpha 1 (V)-pro-alpha 2 (V). Mutations in this gene are associated with Ehlers-Danlos syndrome, types I and II. Alternative splicing of this gene results in multiple transcript variants. The total of 729 DEGs obtained from the previous step was retrieved in the TCGA database. Finally, only 23 relevant cases were identified with survival data (fig. 5). The mutations in the gene SLC34A2 and TNC are arguably the most common. The protein encoded by the gene SLC34A2 is a pH-sensitive sodium-dependent phosphate transporter[28]. Phosphate uptake is increased at lower pH. Defects in this gene are a cause of pulmonary alveolar microlithiasis. Three transcript variants encoding two different isoforms have been found for this gene. The gene TNC encodes an ECM protein with a spatially and temporally restricted tissue distribution[29]. This protein is homohexameric with disulfide-linked subunits and contains multiple Epidermal Growth Factor (EGF)-like and fibronectin type-III domains. It is implicated in guidance of migrating neurons as well as axons during development, synaptic plasticity and neuronal regeneration.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Node status</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDC4</td>
<td>Clustered</td>
<td>6</td>
</tr>
<tr>
<td>COL1A2</td>
<td>Clustered</td>
<td>6.61111</td>
</tr>
<tr>
<td>GPC3</td>
<td>Clustered</td>
<td>6</td>
</tr>
<tr>
<td>SDC2</td>
<td>Clustered</td>
<td>6</td>
</tr>
</tbody>
</table>

Some limitations should be acknowledged. First, only one dataset was included in the analysis, without considering the impact of population heterogeneity in different countries on the results. Second, the lack of verifiable data sets in this study limits the extrapolation of research results. Third, this study is only for the reanalysis of existing data and lacks the support and verification of experimental data. In conclusion, our results provide a comprehensive bioinformatics analysis between T1aN0 thyroid carcinoma and its paired thyroid normal tissue, T1aN1b or T3N1b thyroid carcinoma and its paired thyroid normal tissue, which could contribute to the understanding of the development of thyroid carcinoma, prevention and treatment of the disease.

Author’s contributions:
Chuan Li and Guorong Liao contributed equally to this work.

Conflict of interests:
The authors declared no conflict of interest.

REFERENCES