


PRIMARY RESEARCH

Open Access



Construction of the coexpression network involved in the pathogenesis of thyroid eye disease via bioinformatics analysis

Jinxing Hu^{1,2}, Shan Zhou^{1,2*}  and Weiyong Guo^{1,2}

Abstract

Background: Thyroid eye disease (TED) is the most common orbital pathology that occurs in up to 50% of patients with Graves' disease. Herein, we aimed at discovering the possible hub genes and pathways involved in TED based on bioinformatical approaches.

Results: The GSE105149 and GSE58331 datasets were downloaded from the Gene Expression Omnibus (GEO) database and merged for identifying TED-associated modules by weighted gene coexpression network analysis (WGCNA) and local maximal quasi-clique merger (ImQCM) analysis. EdgeR was run to screen differentially expressed genes (DEGs). Transcription factor (TF), microRNA (miR) and drug prediction analyses were performed using ToppGene suite. Function enrichment analysis was used to investigate the biological function of genes. Protein–protein interaction (PPI) analysis was performed based on the intersection between the list of genes obtained by WGCNA, ImQCM and DEGs, and hub genes were identified using the MCODE plugin. Based on the overlap of 497 genes retrieved from the different approaches, a robust TED coexpression network was constructed and 11 genes (ATP6V1A, PTGES3, PSMD12, PSMA4, METAP2, DNAJA1, PSMA1, UBQLN1, CCT2, VBP1 and NAA50) were identified as hub genes. Key TFs regulating genes in the TED-associated coexpression network, including NFRKB, ZNF711, ZNF407 and MORC2, and miRs including hsa-miR-144, hsa-miR-3662, hsa-miR-12136 and hsa-miR-3646, were identified. Genes in the coexpression network were enriched in the biological processes including proteasomal protein catabolic process and proteasome-mediated ubiquitin-dependent protein catabolic process and the pathways of endocytosis and ubiquitin-mediated proteolysis. Drugs perturbing genes in the coexpression network were also predicted and included enzyme inhibitors, chlorodiphenyl and finasteride.

Conclusions: For the first time, TED-associated coexpression network was constructed and key genes and their functions, as well as TFs, miRs and drugs, were predicted. The results of the present work may be relevant in the treatment and diagnosis of TED and may boost molecular studies regarding TED.

Keywords: Gene coexpression networks, Thyroid eye disease, Autoimmune inflammatory disease, Differentially expressed genes, Protein–protein interaction

Background

Thyroid eye disease (TED), also known as thyroid-associated ophthalmopathy (TAO), Graves' ophthalmopathy or thyroid orbitopathy, is one of the common autoimmune diseases behind the eye. This disease is commonly seen with Graves' disease and Hashimoto's thyroiditis, but can also occur in healthy people and patients with

*Correspondence: zhoushanningbo2020@126.com; chowshan_0512@163.com

¹ Department of Endocrinology, HwaMei Hospital, University of Chinese Academy of Sciences, 41 Northwest Street Zhejiang Province, Ningbo 315010, China

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

hyperthyroidism [1]. TED is caused by systemic autoimmune attacks of the orbit and other targeted tissues, such as thyroid, skin, pretibial soft tissues [2], periorbital connective tissue, extraocular muscles and orbital fat tissue. Its symptoms include mild eye irritation, severe disfigure and even permanent blindness [3]. Although the majority of TED cases exhibit only mild eye disease, 3–5% of cases exhibit vision loss mainly due to exposure corneal ulceration or compressive optic neuropathy [4]. TED is the result of a combination of genetic and environmental factors. Many of these factors, including race, gender, age, smoking history, thyroid hormone, radioactive stimulating hormone (RAI) therapy and thyroid-stimulating hormone receptor antibodies, have been confirmed by numerous studies [5]. A previous study suggested that the immune response is an important factor involved in TED [6]. However, the pathologic mechanism of TED is still poorly understood. Understanding the pathological mechanism of TED will lead to the discovery of targeted molecular approaches for TED and provide strong evidence for the future treatment of TED. Therefore, it is necessary to scrutinize the molecular mechanism of TED, especially the key genes and pathways related to TED.

Up to date, the transcriptome data of TED are very limited. In addition to data limitation, the methods used for identifying genes with biological relevance in TED are also limited to differential expression analysis-based approaches. Although studies reported by Rosenbaum et al. [6], Zhu and Yang [7] and Lee et al. [8] identified some genes related to TED, the unicity of their approach may not help capture some highly TED-related genes. Thus, integrating multiple approaches may help identify the credible key genes associated with TED.

Gene coexpression networks (GCNs) are key tools for identifying molecular mechanism of diseases [9]. Their main steps include calculating coexpression and selecting important thresholds to filter networks. GCNs can be used to screen candidate biomarkers and therapeutic targets [10, 11]. Although this approach cannot provide a causal relationship, the coexpression network can find regulatory genes in many phenotypes [12]. The weighted gene coexpression network analysis (WGCNA) is a systematic biological method used for detecting coexpressed genes that cannot be detected by differential analysis and has good applications in microarray data or deep sequencing. The problem of multiple tests inherent in microarray data analysis is alleviated in WGCNA [10] which studies biological networks based on pairwise correlations between variables, which is suitable for most high-dimensional data sets. Although principal component analysis (PCA) can also handle high-dimensional data, this method

provides very little information [13]. Thus, WGCNA is usually performed to identify modules containing a high correlation with disease characteristics. Generally, WGCNA is widely implemented in the identification of gene modules involved in diverse diseases. For example, in the study of lupus arthritis (LA), researchers identified highly coexpressed gene modules associated with LA through WGCNA analysis, revealed immune/inflammatory cells dominated by myeloid phenotype and identified cell composition and physiological pathways that are relevant in LA [14]. Researchers have also identified key modules in the psoriatic arthritis (PsA) dataset GSE61281 through WGCNA; through GO and KEGG analysis of key modules and other analysis tools, they determined that RHOH/TRAF1 has great effects on the pathogenesis of PsA [15]. WGCNA has been also widely applied in autoimmune diseases [16, 17] and cancer [18, 19] research. WGCNA was implemented by Li and colleagues in the research of rheumatoid arthritis and allowed the discovery of 4 key modules associated with RA, revealing the biological processes and pathways related to immunity and infection involved in RA [20]. The WGCNA also bridges the gap between individual genes and systemic oncology [21]. WGCNA is not only used for identifying coexpressed mRNAs, but also coexpressed microRNA and lncRNAs. For example, Giulietti et al. [22] discovered the expression of LINC00675 and LINC01133 lncRNAs as well as the occurrence and development of pancreatic cancer by using the coexpression network. Zhou et al. [23] found that hsa-mir-125b-5p, hsa-mir-145-5p, hsa-let-7c-5p, hsa-mir-218-5p and hsa-mir-125b-2-3p were the prognostic and pathological hub miRNAs involved in colon cancer by using WGCNA. However, WGCNA has not been applied for identifying coexpression networks involved in TED. In addition to the widely used WGCNA, there are also GCNs such as mutual rank, highest reciprocal ranks (HRR) and ImQCM [24, 25]. ImQCM is a new network mining method that can be used to identify densely connected modules [26]. The main feature of this method is the use of the local maximum pass to initialize the search, avoid excessive overlap between modules and improve the efficiency of calculation. In addition, this method includes the weight normalization process and can find modules with more balanced sizes. ImQCM has been used in cancer and neurodegenerative syndrome [25, 27]. In a latest report, ImQCM has been used to analyze rhabdomyosarcoma subtypes [28] and 41 coexpression modules were found, of which 17/41 showed obvious up-regulation or down-regulation in the fusion state. However, ImQCM has not been applied to screen coexpression modules associated with TED. Thus, applying

WGCNA and lmQCM to TED-related transcriptome data is promising in discovering new key genes and pathways involved in TED.

Therefore, in our present study, the WGCNA and lmQCM algorithms were employed to generate gene coexpression modules associated with TED. The key genes obtained from different approaches were used for TED coexpression network construction, followed by the identification of TFs, hub genes and miRs that may participate in TED pathogenesis. Potential therapeutics were also predicted based on coexpression network genes. The present work may promote the understanding of the pathological mechanism of TED.

Results

Analysis and functional enrichment analysis of the DEGs

Through differential expression analysis, we obtained 4,999 DEGs (2767 up-regulated, 2232 down-regulated) between TED and normal samples. The volcano plot and heatmap of DEGs are shown in Fig. 1C, D, respectively. As shown in Additional file 1: Table S1, the DEGs were enriched in the biological process (BP) terms of Golgi vesicle transport, positive regulation of establishment of protein localization, positive regulation of cellular protein localization and endoplasmic reticulum to Golgi vesicle. The predominant molecular function (MF) terms were endopeptidase activity, ubiquitin-like protein ligase binding and ubiquitin protein ligase binding. For the cellular component, the most enriched terms were mitochondrial matrix, nuclear speck and transport vesicle. The KEGG pathways of the DEGs were enriched in MAPK signaling pathway, endocytosis and protein processing in endoplasmic reticulum.

Construction of weighted gene coexpression networks and functional enrichment analysis

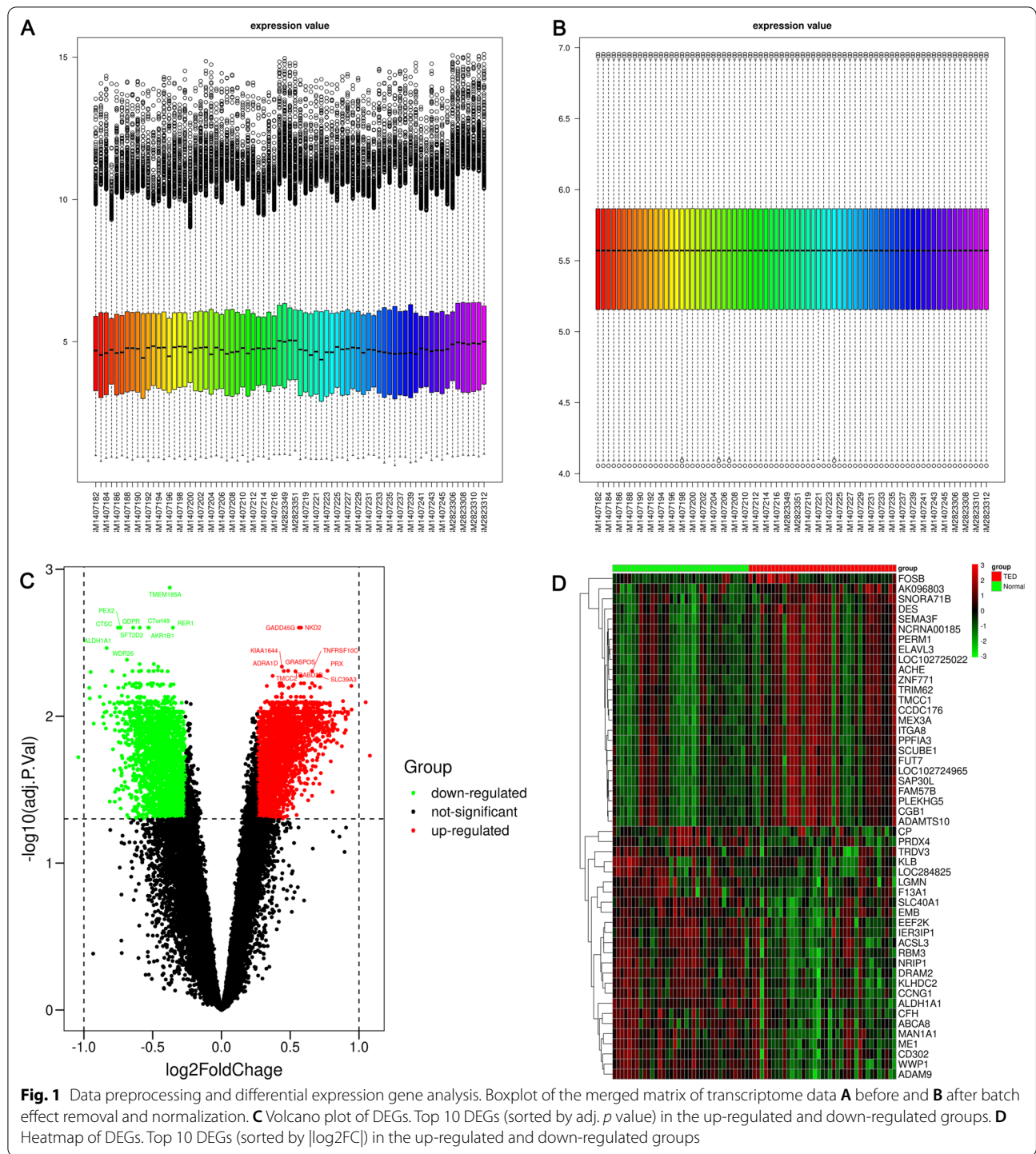
The sample dendrogram and trait heatmap of samples in the TED matrix are shown in Fig. 2A. In WGCNA analysis, TED coexpression network was built when the power value equaled 6 (Fig. 2B). The cluster dendrogram showing the dynamic cut and merging of merged modules is depicted in Fig. 2C. The heatmap of eigengene adjacency is reported in Fig. 2D, while the network heatmap of selected genes is reported in Fig. 2E. Similarly, another TED coexpression network was also constructed by lmQCM analysis. Finally, we got 11 WGCNA modules (Fig. 3A) and 13 lmQCM modules (Fig. 3C). Based on the correlation between MEs and TED trait in the modules among WGCNA modules or lmQCM modules, we selected the modules most highly correlated to TED as TED-specific modules. The blue module (correlation = 0.37, P value = 0.0009) containing 17,008 genes was the module most correlated with TED among

modules identified by WGCNA and was selected as the key TED-associated module for further analysis (Fig. 3A). The functional enrichment analysis of genes in the blue (Fig. 3B, Additional file 2: Table S2) indicated that these genes were enriched in the biological processes (BP) of positive regulation of cellular protein localization, regulation of dendrite development, macromolecule deacylation, memory, protein deacylation, protein deacetylation, cell differentiation in spinal cord and positive regulation of mRNA catabolic process. In the category of molecular function (MF), “transcription coregulator activity” and “ubiquitin-like protein transferase activity” were the most significantly enriched terms, whereas in the category of cellular component (CC), “nuclear speck,” “neuron to neuron synapse” and “postsynaptic specialization” were the most enriched terms (Fig. 3B, Additional file 2: Table S2). The pathways of the genes in the blue module were neuroactive ligand–receptor interaction, endocytosis, MAPK signaling pathway, ubiquitin-mediated proteolysis, Cushing syndrome, hippo signaling pathway, sphingolipid signaling pathway, cholinergic synapse and aldosterone synthesis and secretion (Fig. 3B, Additional file 2: Table S2). The detailed information can be found in Additional file 2: Table S2.

Among the lmQCM modules (Fig. 3C), the module most correlated with TED was the red module (correlation = 0.39, P value = 0.0006) which contained 1873 genes. We found that genes in the red module obtained from lmQCM analysis (Fig. 3D, Additional file 3: Table S3) were mostly enriched in the BP terms of RNA splicing, proteasomal protein catabolic process, proteasomal-mediated ubiquitin-dependent protein catabolic process and Golgi vesicle. The most enriched MF terms of genes in the red module were “ubiquitin-like protein transferase activity,” “ATPase activity,” “cadherin binding” and “translation regulator activity” while the most enriched CC terms were “nuclear speck,” “cell-substrate junction,” “focal adhesion,” “chromosomal region” and “spliceosomal complex” (Fig. 3D, Additional file 3: Table S3). For the KEGG pathway analysis, amyotrophic lateral sclerosis, endocytosis, Salmonella infection, protein processing in endoplasmic reticulum, diabetic cardiomyopathy, ubiquitin-mediated proteolysis and spinocerebellar ataxia were the most enriched pathways (Fig. 3D, Additional file 3: Table S3). The more detailed information about the modules identified by lmQCM is summarized in Additional file 3: Table S3.

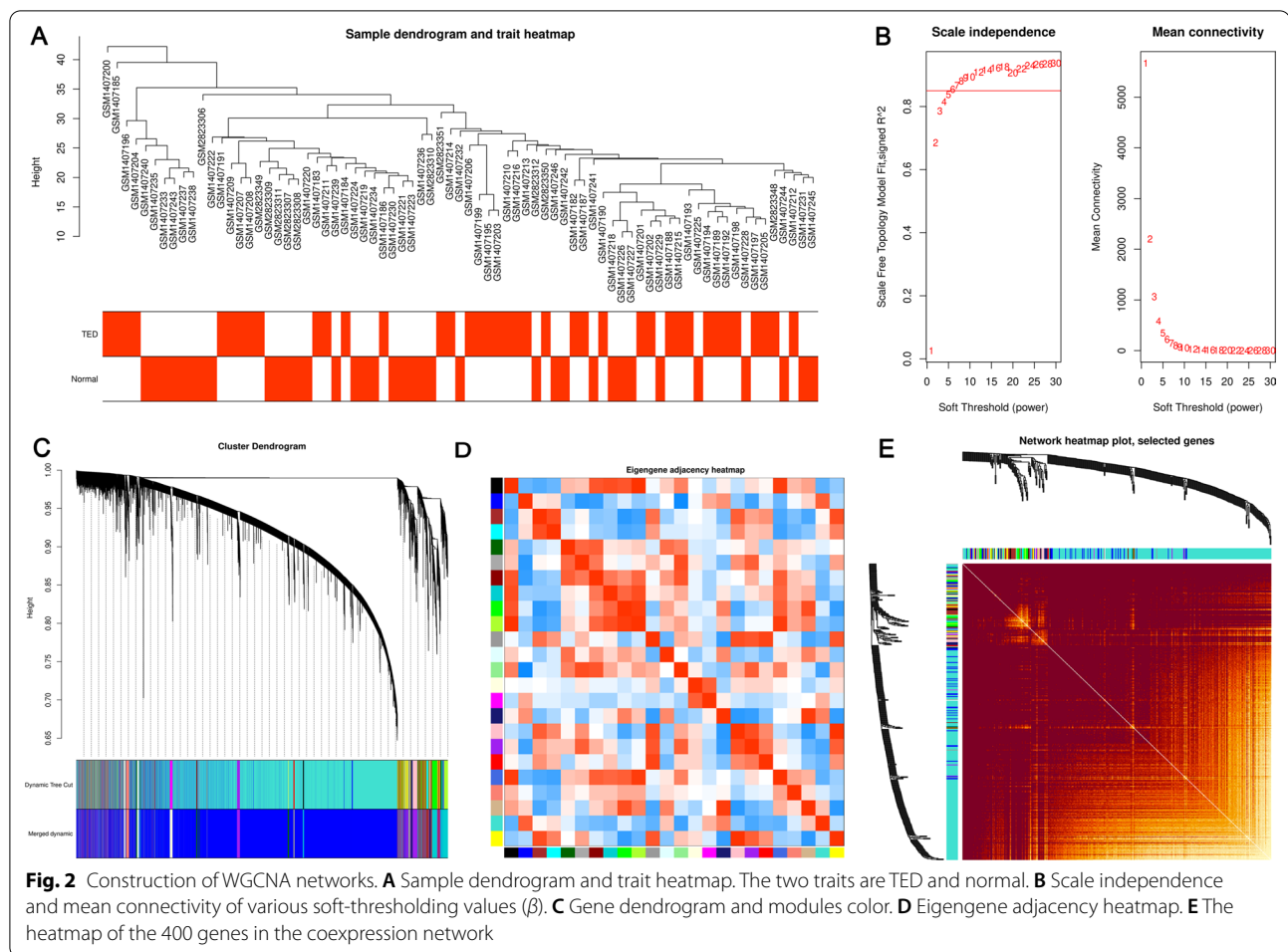
Construction of a robust TED coexpression network and identification of hub genes

A total of 497 overlap genes were obtained by intersecting the DEGs and genes in the blue (WGCNA) and red (lmQCM) modules associated with TED (Fig. 4A). The



enrichment analysis of the TED-specific genes (Fig. 4B, Additional file 4: Table S4) demonstrated that the genes were enriched in “proteasomal protein catabolic process,” proteasome-mediated ubiquitin-dependent protein catabolic process, protein polyubiquitination, Golgi vesicle transport, post-translational protein modification,

regulation of mRNA metabolic process, macroautophagy and endosome organization in the category of BP. In the MF category, ubiquitin-like transferase activity, ubiquitin transferase activity, ubiquitin-like protein ligase binding and ubiquitin-like protein ligase activity were the most enriched terms (Fig. 4B, Additional file 4: Table S4). In



the CC category, nuclear speck, nuclear envelope and ubiquitin ligase complex were the most enriched terms (Fig. 4B, Additional file 4: Table S4). For KEGG pathways, endocytosis, ubiquitin-mediated proteolysis, amyotrophic lateral sclerosis, spliceosome, spinocerebellar ataxia, autophagy in animal, mRNA surveillance pathway, AMPK signaling pathway, RNA degradation and proteasome were the most enriched pathways of the 497 overlap genes (Fig. 4B, Additional file 4: Table S4). Then, the 497 overlap genes were used as input for PPI analysis. PPI network was constructed in STRING to explore the interactions between genes involved in TED. The PPI network generated from the overlap genes included 443 nodes and 1650 edges, and the average number of neighborhoods was 7.478 (Fig. 5). For clearer visualization, we kept nodes with degree greater than ten (Fig. 6). MCODE analysis allowed the identification of 2 predominant clusters of hub genes. The 11 hub genes in cluster 1 with the highest score were ATP6V1A, PTGES3, PSMD12, PSMA4, METAP2, DNAJA1, PSMA1, UBQLN1, CCT2, VBP1 and NAA50, among which PTGES3, PSMD12,

PSMA4, PSMA1, UBQLN1, CCT2 and VBP1 had node degree higher than ten.

Prediction of transcription factors (TFs), miRs and drugs targeting genes in the robust TED coexpression network

We uploaded TED-specific genes to ToppGene to investigate the potential TFs, miRs and drugs perturbing the biological function and pathways involved in TED pathogenesis. Detailed information of the TFs is summarized in Additional file 5: Table S5. Based on the overlap genes, we retrieved 221 TED-associated TFs at a p value cutoff of 0.01. As shown in Additional file 5: Table S5, NFKB had the highest number (115 genes) of regulatory targets among genes in the coexpression network, followed by ZNF711 (104 genes), ZNF407 (97 genes), MORC2 (91 genes), NFE2L1 (89 genes), UBPI (87 genes) and HOXA2 (86 genes). In total, 476 genes in the TED-associated coexpression network were targeted by the predicted TFs. To further investigate the regulation of genes in the TED coexpression network, ToppGene was used for the prediction of 4581 miRs targeting these genes

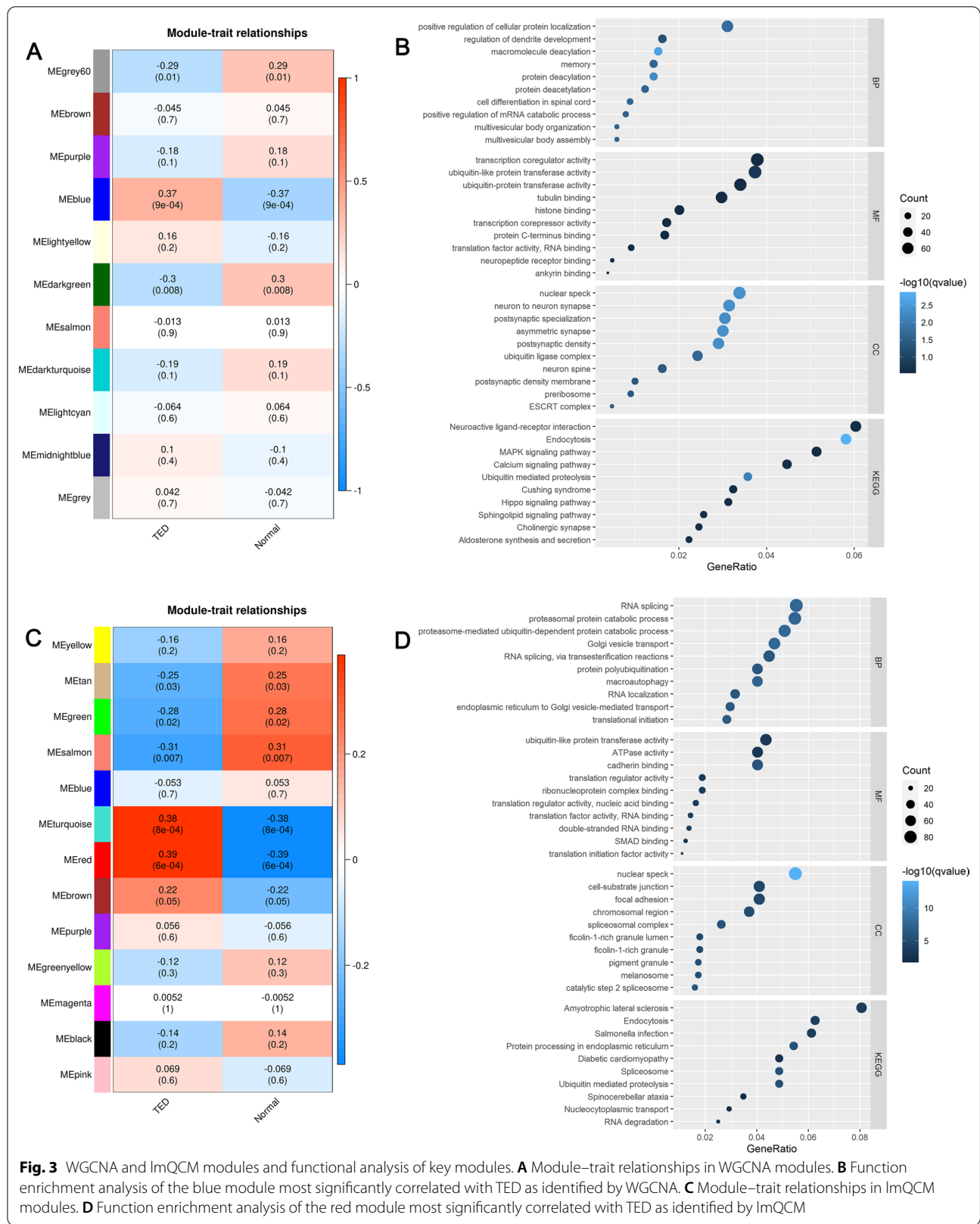
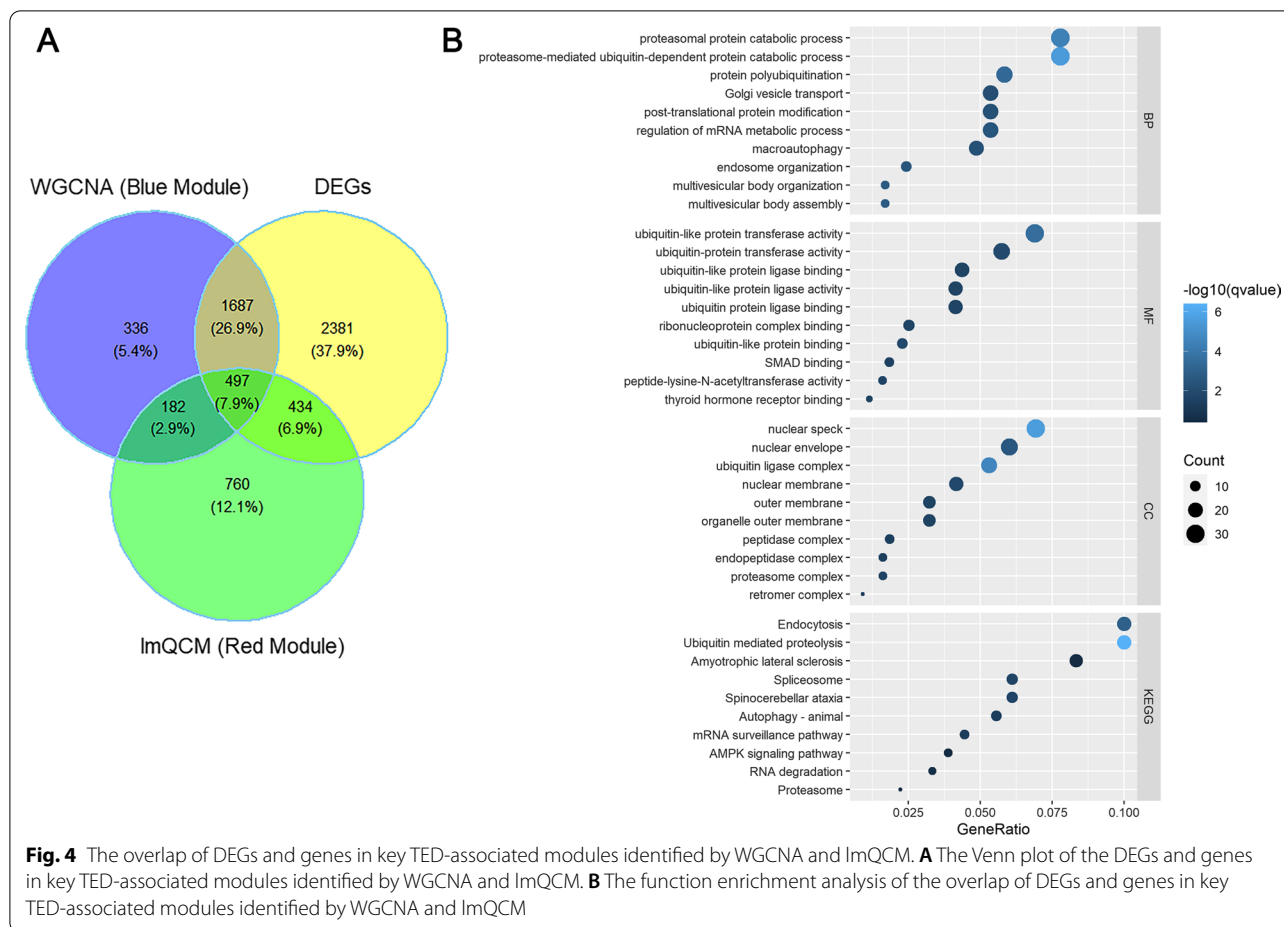


Fig. 3 WGCNA and ImQCM modules and functional analysis of key modules. **A** Module–trait relationships in WGCNA modules. **B** Function enrichment analysis of the blue module most significantly correlated with TED as identified by WGCNA. **C** Module–trait relationships in ImQCM modules. **D** Function enrichment analysis of the red module most significantly correlated with TED as identified by ImQCM

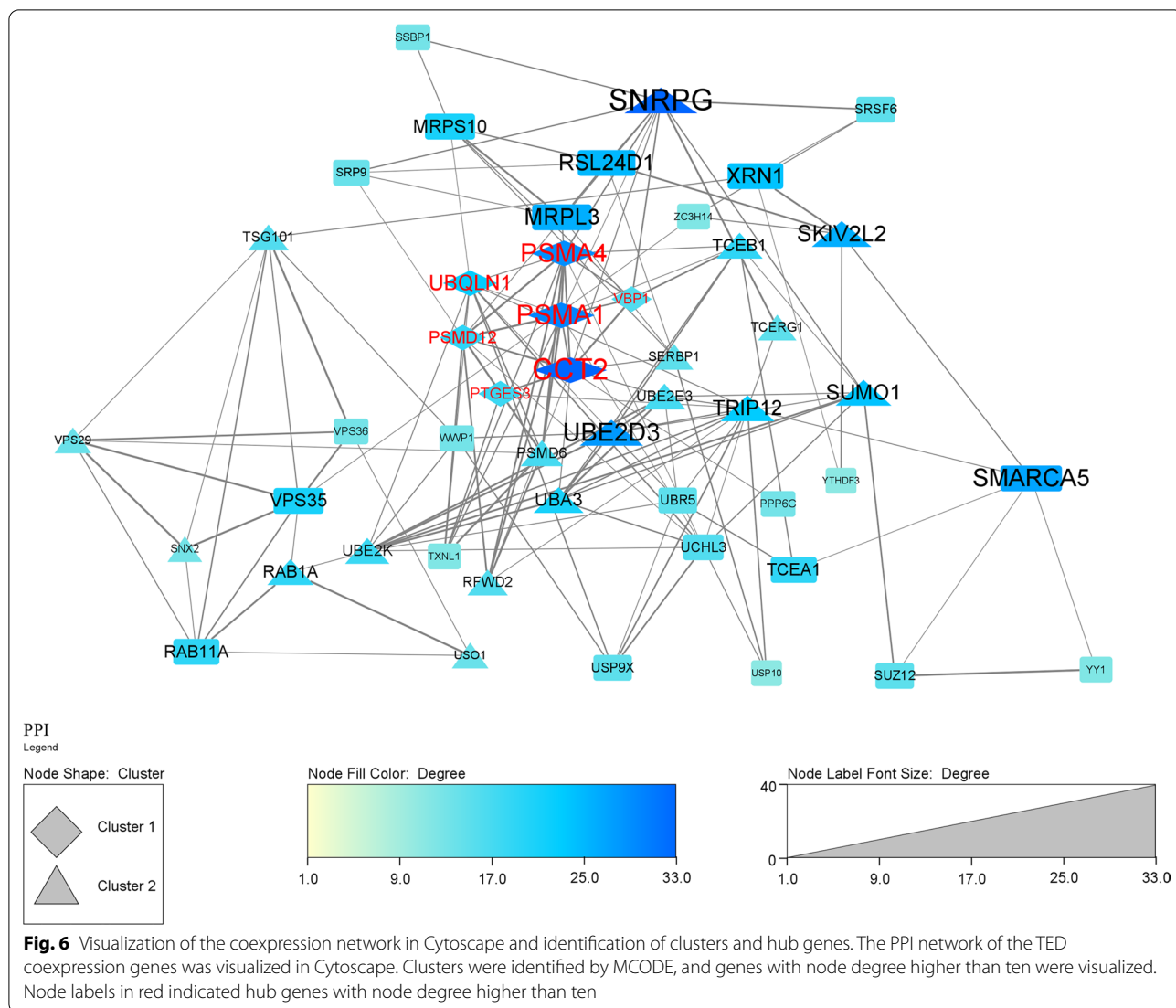


from different databases. As shown in Additional file 5: Table S5, we found that hsa-miR-144 had the highest number of targets (134 genes) among genes in the coexpression network, followed by hsa-miR-3662 (130 target genes), hsa-miR-12136 (126 target genes), hsa-miR-3646 (122 target genes), hsa-miR-607 (121 target genes), hsa-miR-548x-3p (120 target genes), hsa-miR-548aj-3p (120 target genes) and hsa-miR-411* (116 target genes). Moreover, for therapeutic purpose, we predicted the drugs targeting the genes in the TED-associated coexpression network. As shown in Additional file 5: Table S5, 138 drugs were predicted to significantly target genes in the coexpression networks. Among these drugs, enzyme inhibitors had the highest number of target genes (112 genes), followed by the chlorodiphenyl (103 target genes), finasteride (90 target genes), nefazodone (89 target genes) and sodium arsenate (85 target genes).

Discussion

Thyroid eye disease (TED) is one of the most common autoimmune inflammatory diseases [29]. In about 3–5% of patients, TED is accompanied by loss of

vision and compressive optic neuropathy [4]. Exploring disease-related genes on the molecular level can contribute to the development of drugs for TED treatment. In this study, we used the differential expression analysis, WGCNA and ImQCM approaches to identify TED-associated modules and constructed a solid coexpression network for TED. Based on the overlap of important genes retrieved from the different approaches, 11 genes (ATP6V1A, PTGES3, PSMD12, PSMA4, METAP2, DNAJA1, PSMA1, UBQLN1, CCT2, VBP1 and NAA50) were identified as hub genes in the PPI network of the overlap coexpression genes. We also found that NFRKB, ZNF711, ZNF407 and MORC2 were key candidate TFs possibly regulating genes in the TED-associated coexpression network. Similarly, the genes in the coexpression network were mostly regulated by predicted miRs such as hsa-miR-144, hsa-miR-3662, hsa-miR-12136 and hsa-miR-3646. Functional enrichment analysis indicated that genes in the coexpression network were enriched in the biological processes related to proteasomal protein catabolic process, proteasome-mediated ubiquitin-dependent



sequencing. Many of these were associated with inflammation, cytokine signaling, adipogenesis, IGF-1 signaling and glycosaminoglycan binding. Gopinath et al. [3] demonstrated that cardiac calsequestrin gene CASQ2 was markedly up-regulated (~2.2-fold) in patients with the orbital disease and/or healthy people by microarray analysis. Zhu and Yang [7] indicated that POMC, IL-2, GNG3, CXCR4, TLR4, CSF1R, LPAR3 and CXCL8 were hub genes among many DEGs. It was also found that 71 DEGs were associated with TED in the comparative Toxicogenomics Database. Rosenbaum et al. [6] extracted RNA from the orbit of 83 volunteers and found that TED had fewer inflammatory marker characteristics than other ophthalmologic diseases through PCA. In the present study, by combining multiple bioinformatics approaches, we were able to construct, for the first time, a robust regulatory coexpression network. The regulatory

interaction genes identified here might shed light in the understanding of the pathogenesis of TED. There are various tools for weighted gene coexpression network mining, for example, quasi-clique merger (QCM), lmQCM, Markov clustering (ML) and WGCNA. Except for WGCNA, lmQCM is suitable for identifying small yet densely connected gene coexpression networks. A detailed description of each approach is provided in “Introduction” and “Methods” sections. Here, we performed lmQCM and WGCNA network and found that 11 and 13 modules were identified by WGCNA and lmQCM, respectively. The correlation between TED and both lmQCM modules and WGCNA modules was similar, ranging from - 0.31 to 0.39. However, we obtained a small number of genes in TED-specific modules through lmQCM compared with WGCNA. Moreover, 1488 genes of TED-specific modules were found in both WGCNA

modules and lmQCM modules, suggesting that the genes obtained by both methods were the most credibly related to TED and that both methods are highly efficient in module discovery. Pathway enrichment analysis results showed that the pathways regulated by the genes in the TED-specific modules identified by lmQCM and WGCNA showed overlap in functional terms, indicating that both WGCNA and lmQCM were able to extract important genes involved in TED.

In eukaryotes, the binding of transcription factors and histone modification can accurately regulate gene expression. TFs can control the transcription rate of genetic information from DNA to messenger RNA by binding to specific DNA sequences. A number of studies have reported that transcription factors can participate in the activation and shutdown of gene functions. In order to explore the transcription factors regulating the expression of genes involved in TED, we used TopGene to predict the TFs that regulate the expression of TED-specific genes. We found a total of 221 TFs that may be related to the expression of TED-associated coexpression genes. Among them, NFRKB, ZNF711, ZNF407 and MORC2 were the most prevalent, but these TFs and many of other identified TFs have not been reported in TED before. Thus, we can conclude that we identified new TED-related TFs which might be important in the research on the pathogenesis of TED. More relevant evidence needs to be provided in future research.

Ubiquitination is believed to be involved in the regulation of almost all life activities, including cell cycle, proliferation, apoptosis, gene expression, transcription regulation, inflammatory immunity, etc. Several studies have found that ubiquitination has an important regulatory role in autoimmune diseases [31–33]. Ubiquitination controls the development, activation and differentiation of T cells and maintains an effective adaptive immune response to pathogens and immune tolerance to self-tissues. Autoimmune diseases are not only directly affected by ubiquitination, but may also be regulated by ubiquitination-related enzymes. Wang and his colleagues found that mutations in the ubiquitin-conjugating enzyme E2L3 (UBE2L3) gene may be related to the high risk of Hashimoto's thyroiditis (HT) in Han Chinese, indicating that ubiquitination may be involved in thyroid-related autoimmune diseases [34]. TED is an autoimmune disease, and its occurrence and development may be closely related to ubiquitination. In addition, a previous study reported that small ubiquitin-like modifier 4 (SUMO4) is a common autoimmune disease susceptibility gene in the Japanese population [35]. We found that the genes in the regulatory coexpression network were enriched in the biological processes related to proteasomal protein catabolic process, proteasome-mediated

ubiquitin-dependent protein catabolic process and protein polyubiquitination and the pathway of ubiquitin-mediated proteolysis, indicating that ubiquitination plays a significant role in the pathogenesis of TED. The ubiquitin–proteasome pathway is responsible for the degradation of most intracellular proteins in eukaryotic cells [36]. Proteasome dysfunction is related to autoimmune diseases [37], neurodegeneration [38], heart [39] and other diseases. We found that genes in the regulatory network were also enriched in the proteasome pathway and the processes of proteasomal protein catabolic process, and proteasome-mediated ubiquitin-dependent protein catabolic process, which can explain a previous finding that the immune proteasome LMP2 can be used as a new target for the treatment of autoimmune hypothyroidism [40]. A previous study also demonstrated that TNF- α can promote mTOR-dependent proteasome-mediated PDCD4 degradation in orbital fibroblasts and that rapamycin can enhance TNF- α induction and IL6 secretion by inhibiting PDCD4 degradation in orbital fibroblasts [41]. Thus, the coexpression network constructed in this study may regulate the pathogenesis of TED via the ubiquitin–proteasome pathway (UPP). Although we only discussed some of the pathways, other pathways also deserve a more in-depth study.

In order to identify the interaction between proteins associated with the pathogenesis of TED and the related hub genes, we performed PPI analysis. We found 11 hub genes (ATP6V1A, PTGES3, PSMD12, PSMA4, METAP2, DNAJA1, PSMA1, UBQLN1, CCT2, VBP1 and NAA50) in the PPI network. Several studies reported that the PSMD12, PSMD14, PSMD5, PSMD6 were associated with the pathway of the proteasome [42–45]. PSMA1, PSMA4 and PSMB7 are the three subunits of the 20S proteasome [46]. The regulatory effects of PSMA1 and PSMD12 in the liver of thyroid hormone were previously reported [47], which is consistent with the down-regulation of PSMA1 and PSMD12 in TED as observed in the present work. The 11 hub genes may regulate TED by regulating the pathway of UPP, which is consistent with the findings in the functional enrichment analysis. At present, there is still little about the role of these hub genes in the regulation of thyroid diseases. Although we have discussed the regulatory function of PSMA1, PSMD12 in the thyroid, there is no report indicating that the 11 hub genes are directly related to the occurrence and progression of TED, and thus, more studies are needed to validate their regulatory role in TED.

The miRs are important regulators of gene expression. However, there are limited data on the involvement of miRs in the pathogenesis of TED. Up to date, only miR-130a, miR-146a and miR-155 and miR-1287-5p have been reported in TED [48–50]. In the present work, we

predicted 4581 miRs targeting genes in the coexpression network involved in TED. The most prevalent miRs were hsa-miR-144, hsa-miR-3662, hsa-miR-12136 and hsa-miR-3646. Thus, our study provided a new set of miRs with relevant potential functions in the pathogenesis of TED. Further studies based on these miRs might enlighten our understanding on the pathogenesis of TED and open new avenues for the development of therapeutics for this disease.

To date, there is no effective drug for the treatment of TED. Few studies have indicated that some drugs including teprotumumab, tocilizumab, rituximab and mycophenolate can improve the outcomes of TED [51]. Herein, we predicted a set of drugs that could target the predicted gene coexpression network. Enzyme inhibitors and chlorodiphenyl (54% chlorine), finasteride, nefazodone, sodium arsenate, beta-methylcholine, propiconazole, pentachlorophenol, glucose, thimerosal, thapsigargin, succimer were the drugs with the highest number of target genes. The predicted genes may be efficient in the treatment of TED, which needs further experimental screening and validation.

Our study presents some limitations. Although TED-specific genes were found through the combined use of DEGs, lmQCM and WGCNA, due to the small sample size we used, there may be some deviations. In addition, our results were only based on public data, and more in vivo and in vitro experiments are still needed to verify our findings. We found 11 hub genes regulating the pathway of ubiquitin and proteasome through PPI analysis, but the role of these genes in the occurrence and development of TED still needs in-depth validation. In addition, we tried to identify the association of genes in the TED coexpression network to other diseases in ToppGene, but the analysis output revealed no disease association (no output data), which implied that the coexpression network might be specific to TED; however, tremendous work is needed to confirm this observation. We hereby call on more colleagues to experimentally verify our results.

Conclusions

By harnessing different bioinformatics approaches, we constructed a robust gene coexpression network for TED 11 hub genes involved in this disease. Furthermore, we predicted the TFs, miRs and drugs that could target genes in the coexpression network. The functions of the coexpression network were also identified and indicated the important implication of protein ubiquitination in TED. The role of the identified hub genes in the formation and development of TED has not been reported so far. So, this is the first study to propose that these 11 hub genes may be related to the pathological process of TED,

but more evidence needs to be given in future research. In the future, we will conduct more experimental research to explore the central genes of TED, so as to lay the foundation for diagnosing the occurrence and development of TED and developing new therapeutic targets for the treatment of TED.

Methods

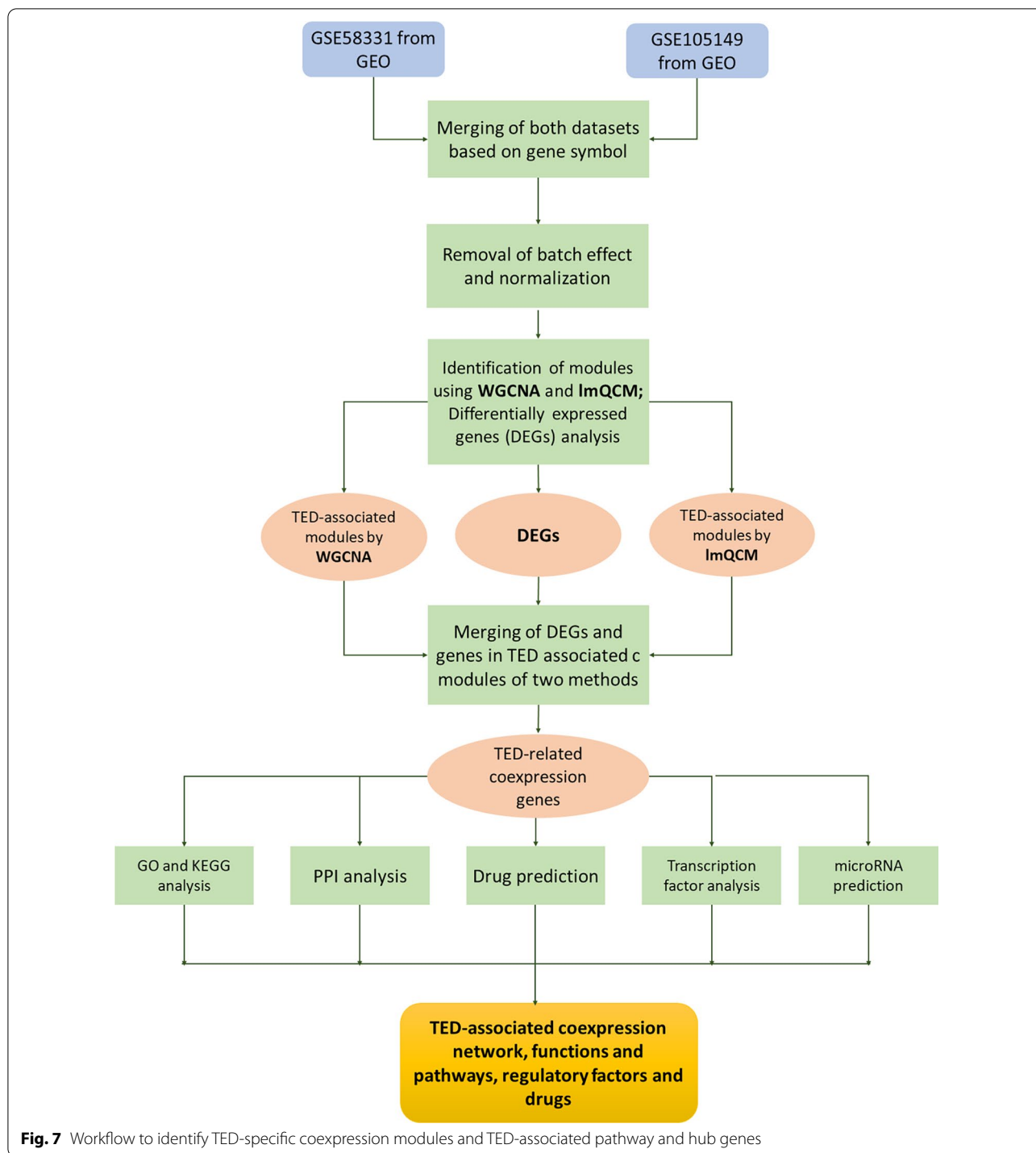
Data collection and preprocessing

The data used in this study were obtained from two datasets in the GEO database (<https://www.ncbi.nlm.nih.gov/>): GSE105149 (containing four TED samples and seven Normal samples) and GSE58331 (containing 35 TED samples and 29 Normal samples). The platform of the two datasets was GPL570. The array probes of each dataset were mapped to their respective gene IDs using corresponding array annotations. Probes matching multiple genes were represented by those with the highest average expression level. As shown in Fig. 7, the two datasets were merged into a matrix containing 22,880 genes and 75 samples. The batch effect of the merged matrix was removed by the R “sva” package (<http://www.bioconductor.org/packages/release/bioc/html/sva.html>) (Fig. 1A). The `normalizeBetweenArrays` function in the R “limma” package (<http://www.bioconductor.org/packages/release/bioc/html/limma.html>) was used to perform normalization of the data in the merged matrix (Fig. 1B).

Gene coexpression network construction and modules selection

WGCNA network was performed to identify TED-related modules based on the preprocessed matrix. Firstly, the similarity of gene coexpression between gene m and gene n was defined as $S_{mn} = |\text{cor}(m, n)|$. Secondly, a power function was used to correlate adjacency between the genes: $a_{mn} = \text{power}(S_{mn}, \beta) = |S_{mn}|^\beta$. Thirdly, the gradient method (the power value ranging from 1 to 30) was used to test scale independence and mean connectivity. When the value was above 0.80 [11], a proper power value was screened out to establish a scale-free network. Finally, the adjacency matrix was converted into a topological overlap matrix and the module was screened out by using hierarchical average linkage clustering analysis, which was a gene dendrogram with a minimal size of 60. Then, 400 genes were randomly selected for heatmap drawing.

Additionally, we performed the local maximized quasi-clique merger (lmQCM) network mining [26] based on the preprocessed matrix. First, the samples in the TED matrix were divided into two groups: TED group and normal group. Then, the Pearson correlation coefficient (PCC) of each pair of genes in the TED and normal groups was calculated separately. Finally, we obtained



coexpression networks associated with TED and normal groups. In the networks, nodes correspond to genes, and weight corresponds to PCC value. The lmQCM [26] was used to detect densely connected modules in the weighted network. The parameters for lmQCM were set as follows: $\gamma = 0.55$, $t = 1$, $\lambda = 1$, $\beta = 0.4$,

minimum cluster size = 10. The lmQCM is available both in web version (<https://apps.medgen.iupui.edu/rsc/tsunami/>) and R package in CRAN as “lmQCM.”

To identify the module associated with TED in weighted networks constructed by WGCNA and lmQCM, respectively, we did the following processing.

Based on the modules identified in the previous step, we used the first principal component of each module (also named module eigengene, ME) to represent the expression levels of genes in each module. The correlation between MEs and TED was evaluated using PCC, and the modules with the highest or lowest PPC were selected as the TED-specific modules for the coexpression networks from WGCNA or lmQCM.

Identification of DEGs

The differential expression analysis based on the merged matrix was performed by “limma” package in R with the threshold of $|\log_2FC| > 0.263$ and adj. p value < 0.05 . The expression levels of the top 10 up-regulated and down-regulated genes (ranked by $|\log_2FC|$) were used to draw the heatmap by R “pheatmap” package. The volcano plot was visualized by the R “ggpubr” (<https://cran.r-project.org/web/packages/ggpubr/index.html>) and “ggthemes” (<https://cran.r-project.org/web/packages/ggthemes/index.html>) packages. The top 10 significant genes in up-regulated and down-regulated genes (ranked by adj. p value) were highlighted with gene symbols.

Overlap of DEGs and genes in TED-specific modules

We merged the DEGs and genes in TED-specific modules and got TED-specific genes. The Venn diagram analysis was performed using the ggvenn library in R programming software to identify the intersection between lists of genes from DEGs, the blue module in WGCNA and the red module in lmQCM.

ToppGene Suite analysis for prediction of TFs, microRNAs and candidate drugs

The prediction of TFs, microRNAs and candidate drugs based on the TED-specific genes were explored by analysis in ToppGene Suite (<https://toppgene.cchmc.org/>). The p value cutoff was set to 0.01.

Protein–protein interaction (PPI) network construction

To explore the interaction between proteins coded by TED-specific genes, PPI analysis was performed. Herein, we uploaded the TED-specific genes to the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING, <https://string-db.org/cgi/>), which is a database of known and predicted protein–protein interaction. We selected the organism “Homo sapiens” and downloaded *.TSV and *.png files from the STRING database. The network was also imported into Cytoscape 3.4.0 software for visualization. The MCODE plug-in in Cytoscape was used for the identification of gene clusters. The nodes with no interaction with other proteins in the PPI network were removed, and only genes with node degree higher than ten were visualized in Cytoscape.

Functional enrichment analysis

The biological function of the genes was investigated by the R “clusterProfiler” package [52]. The GO terms and pathways were sorted by adjusted p values (adj. p value), and adj. p values < 0.05 were considered significant. The top 10 terms of GO biological process, cellular component and molecular function and the top 10 pathways of the KEGG pathway were selected and visualized by the R “ggplot2” package [53].

Abbreviations

TED: Thyroid eye disease; TAO: Thyroid-associated ophthalmopathy; WGCNA: Weighted gene coexpression network analysis; TF: Transcription factor; TOM: Topological overlap matrix.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40246-022-00412-0>.

Additional file 1: Table S1 Results of the functional enrichment analysis of genes in differentially expressed genes (DEGs) as the output of cluster profiler analysis.

Additional file 2: Table S2 Results of the functional enrichment analysis of genes in the blue module identified in WGCNA analysis as the output of cluster profiler analysis.

Additional file 3: Table S3 Results of the functional enrichment analysis of genes in the red module identified in lmQCM analysis as the output of cluster profiler analysis.

Additional file 4: Table S4 Results of the functional enrichment analysis of the 497 TED-associated coexpression genes as the output of cluster profiler analysis.

Additional file 5: Table S5 Detailed information of predicted transcription factors.

Acknowledgements

Not applicable.

Author contributions

JH contributed to writing manuscript and data analysis. WG performed the experiments and obtained data from databases. SZ designed the study and revised manuscript. All the authors read and approved the final version of the manuscript.

Funding

This work was supported by the Zhejiang Medical Science and Technology Research Foundation (2018KY697).

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Endocrinology, HwaMei Hospital, University of Chinese Academy of Sciences, 41 Northwest Street Zhejiang Province, Ningbo 315010, China. ²Ningbo Institute of Life and Health Industry, University of Chinese Academy of Sciences, Ningbo 315010, China.

Received: 5 September 2021 Accepted: 2 September 2022

Published online: 08 September 2022

References

- Hsu HJ, Hsu CK, Chen TS, Hsu CH. Thyroid eye disease. *QJM*. 2016;109(1):67–8.
- Kiljanski J, Nebes V, Stachura I, Kennerdell JS, Wall JR. Should Graves' disease be considered a collagen disorder of the thyroid, skeletal muscle and connective tissue? *Horm Metab Res*. 1995;27(12):528–32.
- Gopinath B, Wescombe L, Nguyen B, Wall JR. Can autoimmunity against calsequestrin explain the eye and eyelid muscle inflammation of thyroid eye disease? *Orbit*. 2009;28(4):256–61.
- Wiersinga WM, Bartalena L. Epidemiology and prevention of Graves' ophthalmopathy. *Thyroid*. 2002;12(10):855–60.
- Stan MN, Bahn RS. Risk factors for development or deterioration of Graves' ophthalmopathy. *Thyroid*. 2010;20(7):777–83.
- Rosenbaum JT, Choi D, Wong A, Wilson DJ, Grossniklaus HE, Harrington CA, et al. The role of the immune response in the pathogenesis of thyroid eye disease: a reassessment. *PLoS ONE*. 2015;10(9):e0137654.
- Zhu FF, Yang LZ. Bioinformatic analysis identifies potentially key differentially expressed genes and pathways in orbital adipose tissues of patients with thyroid eye disease. *Acta Endocrinol (Buchar)*. 2019;5(1):1–8.
- Lee BW, Kumar VB, Biswas P, Ko AC, Alameddine RM, Granet DB, et al. Transcriptome analysis of orbital adipose tissue in active thyroid eye disease using next generation RNA sequencing technology. *Open Ophthalmol J*. 2018;12:41–52.
- Kumari S, Nie J, Chen HS, Ma H, Stewart R, Li X, et al. Evaluation of gene association methods for coexpression network construction and biological knowledge discovery. *PLoS ONE*. 2012;7(11):e50411.
- Di Y, Chen D, Yu W, Yan L. Bladder cancer stage-associated hub genes revealed by WGCNA co-expression network analysis. *Hereditas*. 2019;156:7.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
- van Dam S, Vósa U, van der Graaf A, Franke L, de Magalhães JP. Gene co-expression analysis for functional classification and gene-disease predictions. *Brief Bioinform*. 2018;19(4):575–92.
- DiLeo MV, Strahan GD, den Bakker M, Hoekenga OA. Weighted correlation network analysis (WGCNA) applied to the tomato fruit metabolome. *PLoS ONE*. 2011;6(10):e26683.
- Hubbard EL, Catalina MD, Heuer S, Bachali P, Robl R, Geraci NS, et al. Analysis of gene expression from systemic lupus erythematosus synovium reveals myeloid cell-driven pathogenesis of lupus arthritis. *Sci Rep*. 2020;10(1):17361.
- Raza W, Wang J, Jousset A, Friman VP, Mei X, Wang S, et al. Bacterial community richness shifts the balance between volatile organic compound-mediated microbe-pathogen and microbe-plant interactions. *Proc Biol Sci*. 2020;287(1925):20200403.
- Tang Y, Zha L, Zeng X, Yu Z. Identification of biomarkers related to systemic sclerosis with or without pulmonary hypertension using co-expression analysis. *J Comput Biol J Comput Mol Cell Biol*. 2020;27(10):1519–31.
- Li H, Yang C, Zhang J, Zhong W, Zhu L, Chen Y. Identification of potential key mRNAs and lncRNAs for psoriasis by bioinformatic analysis using weighted gene co-expression network analysis. *Mol Genet Genomics* MGG. 2020;295(3):741–9.
- Guo R, Chu A, Gong Y. Identification of cancer stem cell-related biomarkers in intestinal-type and diffuse-type gastric cancer by stemness index and weighted correlation network analysis. *J Transl Med*. 2020;18(1):418.
- Yu X, Guo J, Zhou Q, Huang W, Xu C, Long X. A novel immune-related prognostic index for predicting breast cancer overall survival. *Breast Cancer*. 2020. <https://doi.org/10.1007/s12282-020-01175-z>.
- Li X, Yang Y, Sun G, Dai W, Jie X, Du Y, et al. Promising targets and drugs in rheumatoid arthritis: a module-based and cumulatively scoring approach. *Bone Jt Res*. 2020;9(8):501–14.
- Ghazalpour A, Doss S, Zhang B, Wang S, Plaisier C, Castellanos R, et al. Integrating genetic and network analysis to characterize genes related to mouse weight. *PLoS Genet*. 2006;2(8):e130.
- Giulietti M, Righetti A, Principato G, Piva F. LncRNA co-expression network analysis reveals novel biomarkers for pancreatic cancer. *Carcinogenesis*. 2018;39(8):1016–25.
- Zhou XG, Huang XL, Liang SY, Tang SM, Wu SK, Huang TT, et al. Identifying miRNA and gene modules of colon cancer associated with pathological stage by weighted gene co-expression network analysis. *Oncotargets Ther*. 2018;11:2815–30.
- Randhawa V, Pathania S. Advancing from protein interactomes and gene co-expression networks towards multi-omics-based composite networks: approaches for predicting and extracting biological knowledge. *Brief Funct Genomics*. 2020. <https://doi.org/10.1093/bfpg/ela015>.
- Zhang J, Huang K. Pan-cancer analysis of frequent DNA co-methylation patterns reveals consistent epigenetic landscape changes in multiple cancers. *BMC Genomics*. 2017;18(Suppl 1):1045.
- Zhang J, Huang K. Normalized ImQCM: an algorithm for detecting weak quasi-cliques in weighted graph with applications in gene co-expression module discovery in cancers. *Cancer Informatics*. 2014;13(Suppl 3):137–46.
- Xiang S, Huang Z, Wang T, Han Z, Yu CY, Ni D, et al. Condition-specific gene co-expression network mining identifies key pathways and regulators in the brain tissue of Alzheimer's disease patients. *BMC Med Genomics*. 2018;11(Suppl 6):115.
- Helm BR, Zhan X, Pandya PH, Murray ME, Pollok KE, Renbarger JL, Ferguson MJ, Han Z, Ni D, Zhang J, Huang K. Gene co-expression networks restructured gene fusion in rhabdomyosarcoma cancers. *Genes*. 2019;10(9):665.
- Weiler DL. Thyroid eye disease: a review. *Clin Exp Optom*. 2017;100(1):20–5.
- Grusha la O, Ismailova DS, Gankovskaia OA. Risk factors of corneal damage in patients with Thyroid Eye Disease. *Vestn Oftalmol*. 2010;126(6):35–8.
- Rasaei R, Sarodaya N, Kim KS, Ramakrishna S, Hong SH. Importance of deubiquitination in macrophage-mediated viral response and inflammation. *Int J Mol Sci*. 2020;21(21):8090.
- Hos NJ, Fischer J, Hos D, Hejazi Z, Calabrese C, Ganesan R, Murthy AMV, Rybniker J, Kumar S, Krönke M, Robinson N. TRIM21 is targeted for chaperone-mediated autophagy during Salmonella typhimurium infection. *J Immunol*. 2020;205(9):2456–67.
- Chitrakar A, Budda SA, Henderson JG, Axtell RC, Zenewicz LA. E3 ubiquitin ligase Von Hippel-Lindau protein promotes Th17 differentiation. *J Immunol*. 2020;205(4):1009–23.
- Wang Y, Zhu YF, Wang Q, Xu J, Yan N, Xu J, et al. The haplotype of UBE2L3 gene is associated with Hashimoto's thyroiditis in a Chinese Han population. *BMC Endocr Disord*. 2016;16:18.
- Tsurumaru M, Kawasaki E, Ida H, Migita K, Moriuchi A, Fukushima K, et al. Evidence for the role of small ubiquitin-like modifier 4 as a general autoimmunity locus in the Japanese population. *J Clin Endocrinol Metab*. 2006;91(8):3138–43.
- Tsimokha AS, Artamonova TO, Diakonov EE, Khodorkovskii MA, Tomilin AN. Post-translational modifications of extracellular proteasome. *Molecules (Basel, Switzerland)*. 2020;25(15):3504.
- Nagayama Y, Nakahara M, Shimamura M, Horie I, Arima K, Abiru N. Prophylactic and therapeutic efficacies of a selective inhibitor of the immunoproteasome for Hashimoto's thyroiditis, but not for Graves' hyperthyroidism, in mice. *Clin Exp Immunol*. 2012;168(3):268–73.
- Corrigan TS, Lotti Diaz LM, Border SE, Ratigan SC, Kasper KQ, Sojka D, et al. Design, synthesis, and in vitro evaluation of aza-peptide aldehydes and ketones as novel and selective protease inhibitors. *J Enzyme Inhib Med Chem*. 2020;35(1):1387–402.
- Dalal S, Shook PL, Singh M, Singh K. Cardioprotective potential of exogenous ubiquitin. *Cardiovasc Drugs Ther*. 2020. <https://doi.org/10.1007/s10557-020-07042-539>.
- Kimura HJ, Chen CY, Tzou SC, Rocchi R, Landek-Salgado MA, Suzuki K, et al. Immunoproteasome overexpression underlies the pathogenesis of

- thyroid oncocytes and primary hypothyroidism: studies in humans and mice. *PLoS ONE*. 2009;4(11): e7857.
41. Lee WM, Paik JS, Cho WK, Oh EH, Lee SB, Yang SW. Rapamycin enhances TNF- α -induced secretion of IL-6 and IL-8 through suppressing PDCD4 degradation in orbital fibroblasts. *Curr Eye Res*. 2013;38(6):699–706.
 42. Khalil R, Kenny C, Hill RS, Mochida GH, Nasir R, Partlow JN, et al. PSMD12 haploinsufficiency in a neurodevelopmental disorder with autistic features. *Am J Med Genet B Neuropsychiatr Genet*. 2018;177(8):736–45.
 43. Yokoyama S, Iwakami Y, Hang Z, Kin R, Zhou Y, Yasuta Y, et al. Targeting PSMD14 inhibits melanoma growth through SMAD3 stabilization. *Sci Rep*. 2020;10(1):19214.
 44. Levin A, Minis A, Lalazar G, Rodriguez J, Steller H. PSMD5 inactivation promotes 26S proteasome assembly DURING colorectal tumor progression. *Cancer Res*. 2018;78(13):3458–68.
 45. Imai F, Yoshizawa A, Fujimori-Tonou N, Kawakami K, Masai I. The ubiquitin proteasome system is required for cell proliferation of the lens epithelium and for differentiation of lens fiber cells in zebrafish. *Development (Cambridge, England)*. 2010;137(19):3257–68.
 46. Mintie CA, Singh CK, Ndiaye MA, Barrett-Wilt GA, Ahmad N. Identification of molecular targets of dietary grape-mediated chemoprevention of Ultraviolet B skin carcinogenesis: a comparative quantitative proteomics analysis. *J Proteome Res*. 2019;18(10):3741–51.
 47. Pihlajamäki J, Boes T, Kim EY, Dearie F, Kim BW, Schroeder J, et al. Thyroid hormone-related regulation of gene expression in human fatty liver. *J Clin Endocrinol Metab*. 2009;94(9):3521–9.
 48. Hammond CL, Roztocil E, Gonzalez MO, Feldon SE, Woeller CF. MicroRNA-130a is elevated in thyroid eye disease and increases lipid accumulation in fibroblasts through the suppression of AMPK. *Invest Ophthalmol Vis Sci*. 2021;62(1):29.
 49. Wang N, Hou SY, Qi X, Deng M, Cao JM, Tong BD, et al. LncRNA LPAL2/miR-1287-5p/EGFR axis modulates TED-derived orbital fibroblast activation through cell adhesion factors. *J Clin Endocrinol Metab*. 2021;106(8):e2866–86.
 50. Woeller CF, Roztocil E, Hammond C, Feldon SE. TSHR signaling stimulates proliferation through PI3K/Akt and induction of miR-146a and miR-155 in thyroid eye disease orbital fibroblasts. *Invest Ophthalmol Vis Sci*. 2019;60(13):4336–45.
 51. Khong JJ, McNab A. Medical treatment in thyroid eye disease in 2020. *Br J Ophthalmol*. 2021;105(3):299–305.
 52. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–7.
 53. Wickham H. *ggplot2: elegant graphics for data analysis*. New York: Springer-Verlag; 2016.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

