Cell. Mol. Biol. (ISSN: 1165-158X)

Journal Homepage: www.cellmolbiol.org

Cellular and Molecular Biology

Original Article

Naive and regulatory B-cell transcription patterns guide the increased risk of papillary thyroid carcinoma in obesity

Zihui Yu[1], Ziying Xu[1], Shang Li[2,3], Ziyan Tian[1], Jing Yuan[1,*]

[1] *Department of Bacteriology, Capital Institute of Pediatrics, Beijing, China*
[2] *Department of Orthopedics, General Hospital of Chinese PLA, Beijing, China*
[3] *National Clinical Research Center for Orthopedics, Sports Medicine & Rehabilitation, Beijing, China*

Article Info

Article history:

Received: November 17, 2023
Accepted: February 26, 2024
Published: April 30, 2024

Use your device to scan and read the article online

Abstract

A growing number of studies suggest a positive association between obesity and the high incidence of papillary thyroid cancer (PTC), suggesting that the abnormal levels of adipokines associated with obesity may be a risk factor for these aggressive thyroid cancers, but the underlying regulatory mechanisms are not yet clear. We downloaded bulk RNA sequence data for subcutaneous adipose tissue (SAT) in obesity and healthy population and tumor tissues of PTC from GEO database. Through analysis of Differential Expression Genes (DEGs), Gene Set Variation Analysis (GSVA) and Weighted Correlation Network Analysis (WGCNA), we identified co-expressed genes between obesity and PTC, and their pathways were mainly enriched in the regulation of B-cells. Furthermore, through TCGA-THCA (thyroid carcinoma) cohorts analysis, we identified B-cell regulatory-related genes LEF1, TNFRSF13C, SHLD2 and SHLD3 as independent prognostic markers of PTC. Next, we explored the transcriptional regulation mechanism of the increased risk of PTC in obesity through analysis of DNA methylation CpGs data and single-cell RNA sequences (scRNA-seq) from GEO database. PTC-induced hypomethylation of the promoter region may be involved in the transcriptional regulation of these genes, while these genes were further identified in naive and regulatory B-cells of both diseases. Notably, both of the gene expressions in naive and regulatory B-cells showed high similarity in both diseases. Our data reveals the high frequency of PTC in obese populations may be explained by the comparable transcriptional patterns of naive and regulatory B-cells, and offers novel insights for the analysis of critical genes and underlying biological mechanisms for obesity and PTC.

Keywords: Obesity, Papillary thyroid carcinoma, B-cells, Transcriptional regulation, Co-expressed genes, DNA methylation.

1. Introduction

The increasing demographic trend of obesity is posing a major challenge to the world's health systems and becoming a novel medical condition within developing countries. The prevalence of obesity increased dramatically over the past five decades, and the pandemic of COVID-19 further exacerbates weight gain [1]. According to the latest data from NCD Risk Factor Collaboration, 39% of the global adult population was classified as overweight in 2016 [2]. It is predicted that over 2 billion children and adolescents will suffer from obesity by 2025, and China will rank in the top population of obese children [3]. Excessive deposit of body fat could increase the risk of cardiovascular diseases [4], metabolic disorders [5], and even tumorigenesis [6,7]. Abnormal adipose accumulation impairs the metabolism of lipoproteins and triglyceride and also induces an inflammatory microenvironment, resulting in a higher risk of cancer occurrence and aggression [8]. Lipids are not only exploited as building blocks for cancer cell membranes and energy sources but also severed as cancer cell protectants by scavenging overwhelming aggregation of reactive oxygen species (ROS) [9]. Emerging evidence has suggested the underlying connections between obesity and numerous malignant diseases [10].

Over the past two decades, the global incidence of thyroid carcinoma in children and adolescents aged 0–19 years has kept climbing [11,12]. Most clinical studies indicated the positive connections between obesity and a higher incidence of papillary thyroid cancer (PTC), suggesting it is becoming the most common endocrine-related malignancy disease [13-16]. Multiple obesity-related complications contribute to the development of thyroid cancer, including diabetes, metabolic syndrome, and insulin resistance. Overweight thyroid carcinoma patients are likely to have lymph node invasion and multifocal lesions, even with an increased risk of recurrence after tumor-resection surgery [16-18]. The inflammation state caused by hypertrophic adipocytes is considered the major trigger of endocrine-related cancer. Perpetuated vicious feedback between pro-inflammatory adipokines and immune cells secretome promotes the development of thyroid cancer [19]. Newly research indicated that hypoadiponectinemia

* Corresponding author.
E-mail address: yuanjing6216@163.com (J. Yuan).
Doi: http://dx.doi.org/10.14715/cmb/2024.70.4.19

within obesity seems to be a predisposing factor for thyroid cancer [20]. Although there is no evidence of direct impact when recombinant adiponectin was exerted on thyroid carcinoma cell lines in vitro, several clinical studies have reported lower circulating adiponectin levels in patients with thyroid cancer [6]. As a preventable and curable disease, enormous progress has been made in the development of anti-obesity medications and weight-reducing management patterns [21].

Pathological state after obesity allied with radiation and other risk factors together interfere with the transcriptomes of gland cells and induce thyroid cancer. Genomic analysis is a promising approach to distinguish the characterization between radiation-induced and sporadic thyroid carcinoma [22]. DNA methylation is often involved in important biological functions, affecting embryonic development, cell differentiation, disease and even tumor progression [23,24]. In human beings, hypermethylation in the promoter region often inhibits gene expression, while hypermethylation in the gene body (including 5'UTR, exon, 3'UTR, and intron) region often promotes gene expression [25]. Transcriptome sequencing can depict the expression state of transcripts in each biological process, quantifying and integrating the variations by bioinformatics methods [26,27].

We combined transcriptome analysis of subcutaneous adipose tissue (SAT) from normal-weight people, obesity, and cancerous tissue from PTC in order to investigate the reasons why obese people are more susceptible to PTC and to identify the transcriptional regulatory mechanisms between obesity and PTC. Through the analysis of the DNA methylation data and single-cell RNA sequences (scRNA-seq), the similar regulatory patterns that exist between them are explained in order to serve as a foundation for future therapeutically pertinent research.

2. Materials and methods

2.1. Data processing

Initially, the whole transcriptome data of SAT of 26 healthy and 35 obese people (GSE205668) and the transcriptome of tumor tissue samples from 16 patients with PTC (GSE165724) were obtained from the GEO database. CpGs sites file of thyroid malignant nodules and normal adjacent thyroid tissue (GSE107738) were obtained from the GEO database. Single-cell data of PTC (GSE191288), while single-cell data for the SAT in obesity (GSE163830) (Figure 1).

2.2. Analysis of inter-sample correlation and differentially expressed genes

After normalizing the expression matrix using the vst function from the R package DEseq2 (V.1.36.0), we measured the correlation between samples. Next, principal component analysis (PCA) by using prcomp function after the dist function to get the Pearson distance between samples. Analyze the differentially expressed genes in the gene counts matrix using the DESeq function in DEseq2 (V.1.36.0). P-value 0.05 & (log2FoldChange >= 2 | log2FoldChange <= -2) is the threshold for determining DEGs, whereas P-value 0.05 & (log2FoldChange >= -2 & log2FoldChange <= 2) is the threshold for identifying common genes. The R package EnhancedVolcano (V.1.14.0) is used to draw the volcano map.

2.3. Gene set variation analysis (GSVA)

To measure counts per million (CPM) and carry out log2 processing, use the cpm function of the R package edgeR (V.3.38.4). GSVA was carried out on these genes using the R package GSVA (V.1.44.5), and the reference gene sets were chosen from the C5 (ontology gene sets) produced from Homo sapiens in the MSigDB database. Use the gsva function from the GSVA package, specifying method="gsva" and kcdf="Gaussian" for the analysis parameters.

2.4. Weighted correlation network analysis (WGCNA)

The $\log_2$ (cpm+1) matrix of genes served as the input file for the WGCNA analysis, which was carried out by using the R package WGCNA (V.1.71). The pickSoftThreshold function was used to calculate the power value, and the blockwiseModules function to measure the Weight co-expression network. The clustering between samples is seen using the plotDendroAndColors function. The labeled Heatmap algorithm presents the relationship between gene modules and diseases. The correlation between each gene Module was depicted using the plotEigengeneNetworks function.

2.5. Gene ontology and KEGG pathway enrichment analyses

The enrichGO function and enrichKEGG function from the R package clusterProfiler (V.4.4.4), respectively, are used for the analysis of GO terms and KEGG pathways, with pvalueCutoff set to 0.05. The cnetplot function is used for GO terms-genes network visualization. For presentation, the R packages ggplot2 (V.3.3.6) and GOPlot (V.1.0.2) were also utilized.

2.6. Identification and functional analysis of hub genes

Using the R package ggstatsplot (V.0.9.5), an investigation of the correlations between gene expression levels was conducted. The cnetplot function is used for GO terms-genes network visualization.

2.7. Construction and verification of prognosis signature associated

We downloaded the RNA-seq data from TCGA-THCA and removed the samples without follow-up information. Then, ENSEMBL IDs were converted to gene symbols

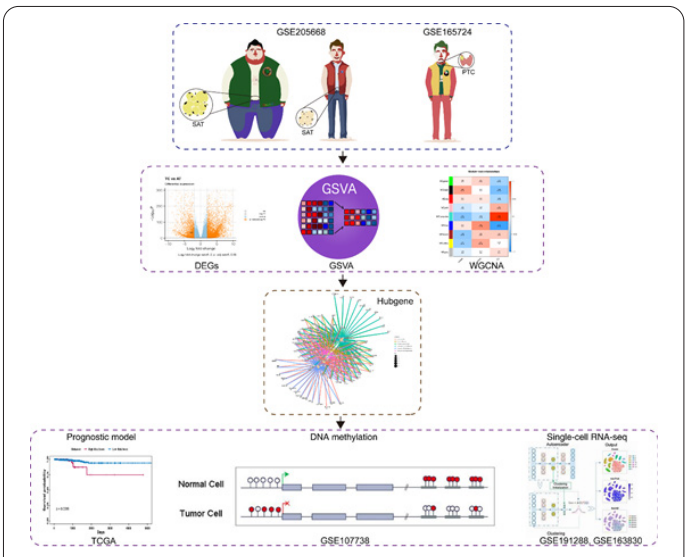

Fig. 1. The workflow of research.

and the count value was converted to $\log_2$ (cpm+1) in the retained data. Afterwards, B-cells regulatory-related genes were selected in later analyses. Univariate and multivariate Cox regression analyses and minimum absolute contraction and selection operators (LASSO) were used to construct prognostic immune-related models to predict overall survival in THCA cases. LASSO-Cox regression was performed to avoid overfitting, while those closely related genes were deleted, and important genes were extracted from the genes screened by univariate Cox regression. In addition, the risk score was calculated by multiplying gene expression by the linear combined regression coefficient obtained by multivariate Cox regression. All cases were classified as a high or low-risk group based on a median risk score [28].

2.8. DNA methylation analysis

Differentially methylated regions (DMRs) were called using DSS (V.2.30.1) with default parameters [29]. The DMRs detection was based on the DMLs results ($P < 0.01$). Finally, regions with difference in methylation > 0.1 were identified as DMRs. The Bioconductor package TxDb.Hsapiens.UCSC.hg19.knownGene was used to be reference. We annotated DMRs using Bioconductor package ChIPseeker [30] (tssRegion=c(-1500,500), flankDistance=5000).

2.9. Single-cell gene expression quantification and cluster classification

The QC process was performed using Seurat (V.4.3.0). We removed the low-quality cells with less than 200 UMIs or with more than 10% mitochondrion-derived UMI counts. Batch effects among the patients were eliminated using the IntegrateData function in Seurat. The top 30 principal components, along with the top 2,000 variable genes, were used in this process. The ScaleData function was then used to regress the influence of UMI counts and mitochondria-derived UMI counts percentages. Major cell types were then identified using Seurat's FindClusters function and visualized using 2D uniform manifold Approximation and Projection (UMAP). The FindAllMarkers function was used to list the markers of each cell cluster [31]. The FeaturePlot, VlnPlot, and DotPlot function was used to show gene expression.

3. Result

3.1. The SAT in obesity and PTC had common immune regulation genes through the analysis of differentially expressed genes (DEGs)

To examine the transcriptional regulation of obesity and PTC, we analyzed the PTC and SAT gene expression data that were obtained. The samples were consistently based on the transcriptome data, but when we investigated the PCA of the transcriptome, we could see that the SAT and thyroid samples had significant differences in terms of transcriptome composition (Figure S1A). The healthy and obese people's SAT was more consistent (Figure S1B). We first examined co-expressed genes, and through our analysis, we were able to see that the genes that were highly expressed in PTC in comparison to SAT (including healthy and obese people) were mainly enriched in regions like transporter activity and immunoglobulin receptor binding (Figure 2A), while the low expression GO pathway was characterized by activity signaling receptor embryonic skeletal system-related pathways (Figure 2B). By analyzing the KEGG metabolic pathway, it is possible to see that the pathways that are upregulated include Transcriptional misregulation in cancer, Cell Adhesion Molecules, Intestinal Immune Network for IgA Production (Figure S1C), and the pathways that are downregulated include IL-17 signaling pathway, Fat Digestion and Absorption, Cholesterol Metabolism, and other pathways (Figure S1D).

To further view changes in SAT in obese people, we analyzed the DEGs in SAT of healthy and obese people, and it was seen from GO and KEGG-enriched pathways that obese SAT were mainly highly expressed relative to healthy tissues as adaptive-based built domains, differentially involved in proliferation effector process, the mononuclear lymphocyte cell proliferation and other pathways (Figure 2C), and down-regulated for carboxylic fatty lipid aerobic catabolic, aerobic ATP electron transport, peptide insulin hormone stimulus (Figure 2D); The majority of the KEGG pathways are connected to regular cell metabolism, and none of them have any practical significance (Figure S1E-1F). We used the intersection of the co-expressed genes in SAT (including healthy and obese people) and PTC and DEGs in normal and obese people's SAT (Figure 3A). Further analysis of these co-expressed genes revealed that the GO keywords mostly comprise multiple immune-related pathways, including those for T cell regu-

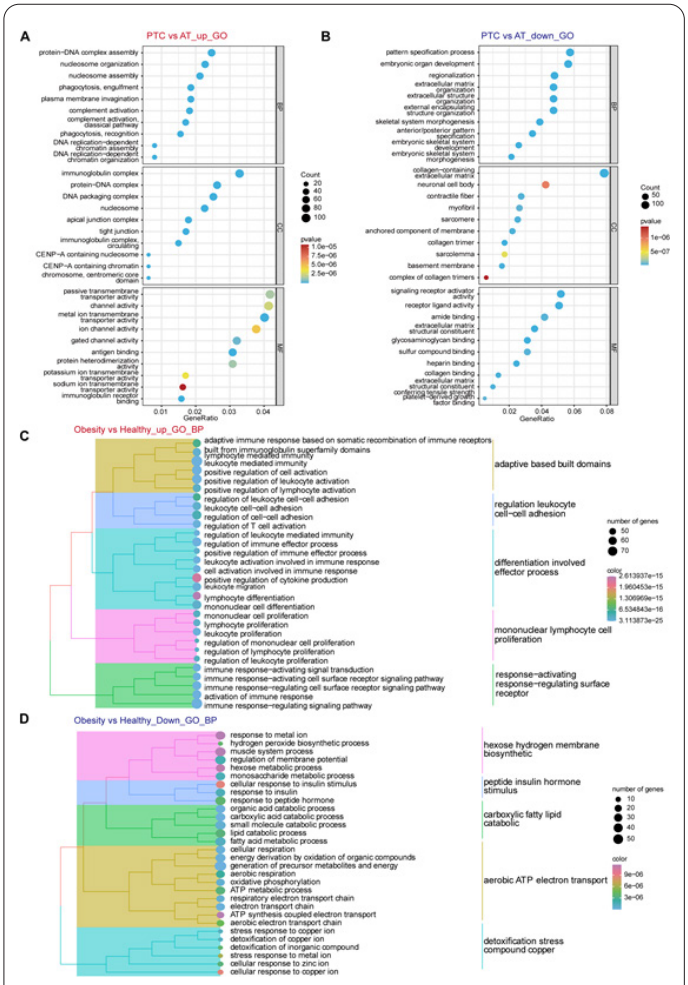

Fig. 2. Differentially expressed genes in SAT and PTC. (A) The GO terms of the up-regulated pathways in PTC compared to SAT. (B) The GO terms of the down-regulated pathways in PTC compared to SAT. (C) The GO terms of the up-regulated pathways in Obesity compared to Healthy. (D) The GO terms of the down-regulated pathways in Obesity compared to Healthy.

lation, T cell differentiation, immune receptor activity, etc (Figure 3B). When we check the KEGG enrichment of these genes, we can find that the intersected co-expressed genes are primarily enriched in the pathways for Diabetic cardiomyopathy cell cytotoxicity, Biosynthesis amino acids cytokine, Glycine Platelet activation serine (Figure 3C). In summary, our results indicated that obese people and PTC cause similar inflammatory reactions.

Further, we looked at the similarities and differences in transcriptome information of these data through GSVA. Obesity and PTC share a large number of the same genes and pathways. When we investigated the pathways that were highly expressed in healthy people, we found that they were primarily involved in lipid metabolism. In contrast, the pathways that were shared by both diseases included those that positively regulated interleukin 17 production, B-cell proliferation, and a few pathways involved in apoptosis (Figure 3D). The several pathways that exercise common cell physiological functions were unnecessary (Figure S2). It was suggested that the lipid metabolism ability of SAT in obesity was decreased, and inflammatory and apoptotic changes tended to PTC.

3.2. The SAT in both healthy and obese people and PTC have unique transcription patterns through the WGCNA analysis

To further explore the co-expressed genes of SAT in obese people and PTC, we used WGCNA analysis to cluster all genes in weighted clusters. The three sets of data were clustered using the Pearson correlation coefficient, and the adjacency matrix and topological overlap matrix were constructed using a scale-free network (Figure S3A). We created a sample clustering tree (Figure S3B) after eliminating the outliers. Nine modules were finally found using the average hierarchical clustering and dynamic tree cropping (Figure 4A). The genes of the black module are significantly connected with the obese group, the genes of the blue and yellow modules are highly correlated with the healthy group, and the genes of the turquoise module are substantially correlated with the PTC group. The distinguishing genes were grouped together. The results showed that the eight modules could be clustered into three clusters, distinguishing the modules belonging to the three groups (Figure 4B). We enriched the GO terms for each module gene (Figure 4C).

Meanwhile, further GO terms analysis of the gene in the four modules closely related to the three groups revealed that the blue module was significantly associated with cartilage development, ribosome biogenesis, connective tissue development, extracellular matrix organization, extracellular structure organization, etc. (Figure 5A);

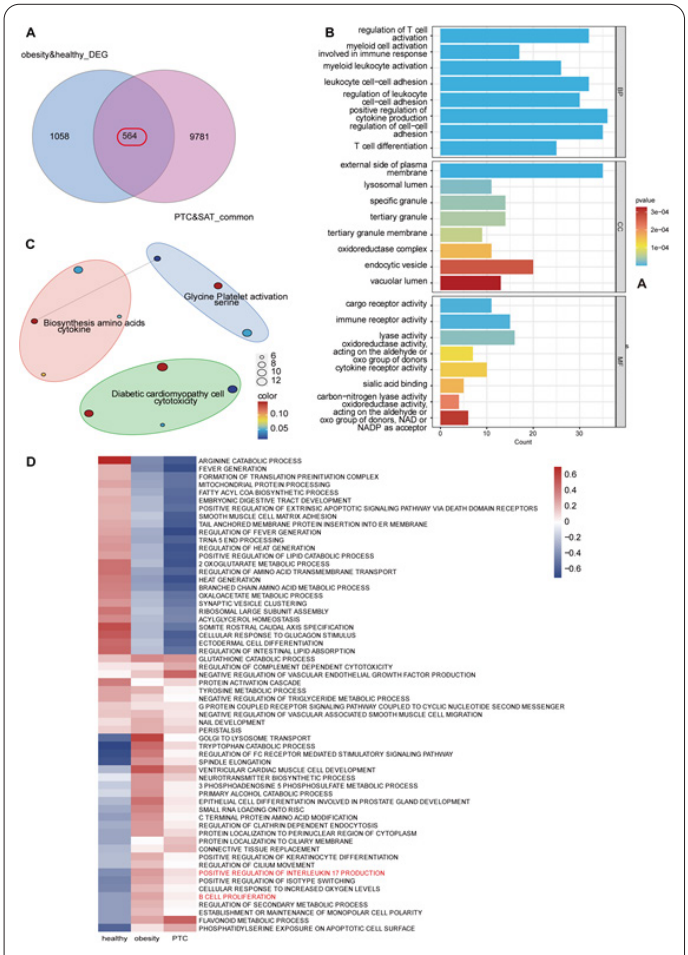

Fig. 3. Co-expression of obese SAT and PTC in transcriptome. (A) The Venn diagram showed that 564 overlapping genes in SAT and PTC co-expressed genes and differentially expressed genes in normal and obese SAT. (B) The GO terms of 564 overlapping genes. (C) The KEGG pathways of 564 overlapping genes. (D) GSVA of healthy people, obesity, and PTC.

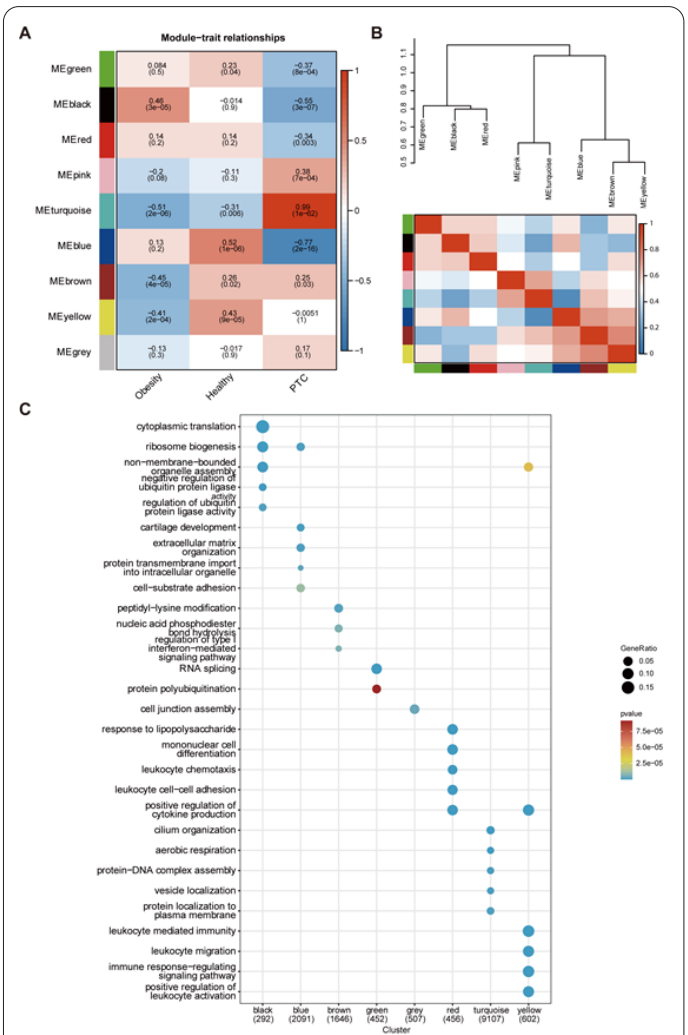

Fig. 4. WGCNA in SAT of healthy and obesity and PTC. (A) Module–trait associations. Each row corresponds to a module, and each column corresponds to a trait. Each cell contains the corresponding correlation and *P* value. The table is color-coded by correlation according to the color legend. (B) Eigengene dendrogram and eigengene adjacency plot. (C) Gene Ontology analysis.

whereas the yellow module was associated with immune diseases such as leukocyte mediated immunity, leukocyte migration, leukocyte cell-cell adhesion, immune response regulating signaling pathway, immune response−regulating cell surface receptor signaling pathway (Figure 5B). The black module was associated with cytoplasmic translation, ribosome biogenesis, ribosomal large subunit biogenesis, ribonucleoprotein complex biogenesis, and ribosome assembly (Figure 5C). The pathways involved in cellular secretion, including cilium organization, protein localization to the plasma membrane, and magnesium ion transmembrane transport, are all closely associated with the turquoise module (Figure 5D). These results indicate that the transcriptome of the obese group is linked to ribosomal changes, whereas the healthy group's transcriptome is more sensitive to cellular outcomes and immune regulation, and the transcriptome of the PTC group shows a correlation with immune cell transport and secretion.

3.3. B-cells regulatory-related genes can serve as the prognostic multi-gene signature for PTC

In order to determine the ultimate co-expression of the SAT in obese people and PTC, we conducted an intersection analysis of the GSVA genes, co-expressed genes obtained from DEGs analysis, and genes in the WGCNA modules, and used these intersections for further analysis (Figure 6A). By checking the GO of the co-expression genes, we focused on the top 6 GO pathways, such as B-cell activation, B-cell proliferation, lymphocyte, and mononuclear cell proliferation, regulation of B-cell activation, and regulation of B-cell proliferation (Figure 6B), which are all involved in the control of immune and inflammatory responses. We further examined the expression of immune cells in both diseases. The relative abundance of immune cells was examined in obesity and PTC using the MCP-counter algorithm. The MCP-counter examines eight immune cell types (T cells, $CD8^{+}$T cells, cytotoxic lymphocytes, B lineage cells, NK cells, monocytic lineage cells, and neutrophils) and 2 stromal cell populations (endothelial cells and fibroblasts). We found that their B lineage and monocytic lineage had higher enrichment abundance (Figure 6C). We hypothesize that the high prevalence of PTC in the obese population may be due to similar B-cell changes.

In order to further clarify the impact of B-cell regulatory-related genes on the prognosis of PTC, 58 genes enriched in top B-cell regulatory-related pathways in co-expression genes were used for subsequent analysis.

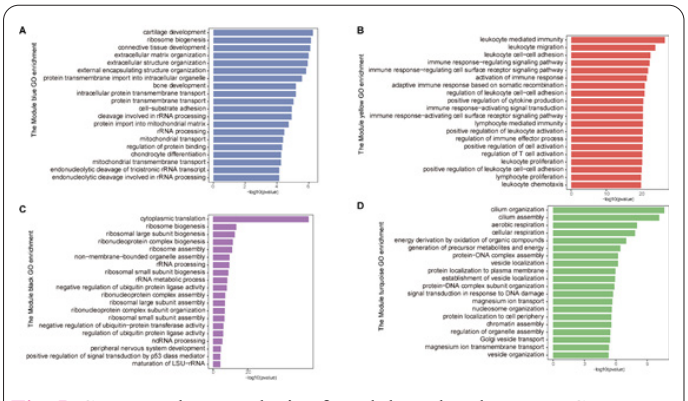

Fig. 5. Gene ontology analysis of modules related to gene. Gene ontology analysis of the genes involved in the blue module (A), the yellow module (B), the black module (C) and the turquoise module (D).

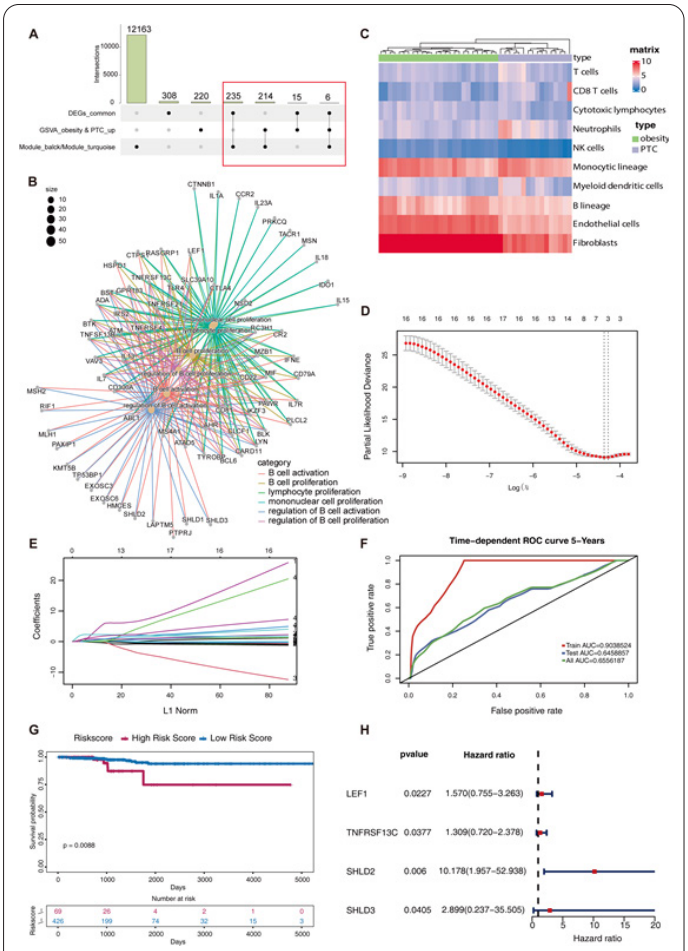

Fig. 6. B-cell regulatory-related genes affected the prognosis of PTC. (A) The upset diagram showed that seven algorithms have screened out 234+214+15+6 overlapping hub genes in obesity and PTC. (B) The GO biological process analyses overlap genes from (A). (C) MCP-counter algorithm shows the heatmap of immune cell relative. (D) The log value of the independent variable lambda (the abscissa represents the confidence interval of each lambda, and the ordinate represents errors in cross-validation). (E) The changing trajectory of each independent variable (the abscissa represents the corrected lambda, and the ordinate represents the coefficient of the independent variable). (F) The 5-year ROC curve in train, test and all cohorts is based on four-gene signature stratification. (G) The KM (Kaplan-Meier) survival curve of the four-gene signature-based stratification in TCGA training cohorts. (H) Forest plot of univariate Cox analysis.

In TCGA, the transcriptome data of thyroid carcinoma (THCA) patients were randomly scored as train and test cohorts, with 248 and 247 samples respectively. Based on the training cohorts, this study chose 4 genes, Lymphoid Enhancer Binding Factor 1 (LEF1), TNF Receptor Superfamily Member 13C (TNFRSF13C), Shieldin Complex Subunit 2 (SHLD2) and Shieldin Complex Subunit 3 (SHLD3) for constructing the prognosis signature via univariate and multivariate Cox regression analysis as well as LASSO (Figure 6D-6E). Thereafter, the risk score values were calculated according to the following formula: Risk score = [LEF1 expression* (0.6311182)] + [TNFRSF13C expression * (0.3624955)] + [SHLD2 expression* (2.0839631)] + [SHLD3 expression* (0.3430136)]. Based on the median risk score, all cases were classified as high- or low-risk groups. Also, the area under the curves (AUCs) of gene signature for 5-year survival in train, test and all cohorts were 0.90, 0.64, and 0.65, respectively (Figure 6F). Notably, high-risk patients had remarkably reduced

OS relative to low-risk patients in all cohorts (Figure 6G). For better exploring the significance of the B-cells regulatory related gene signature in independently predicting prognosis, univariate analysis was conducted, which revealed that the expression of LEF1, TNFRSF13C, SHLD2 and SHLD3 might serve as the independent factor for predicting the prognosis for TCGA-THCA cohorts (Figure 6H). Taken together, B-cell regulatory-related genes can affect the prognosis of thyroid cancer.

3.4. The hypomethylation of the promoter region affects the B-cell regulation of PTC

DNA methylation has been widely considered to regulate gene expression and affect the occurrence and development of diseases. To further clarify the reason that B-cell regulatory-related genes might regulate PTC progression, the DNA methylation cpg sites files of thyroid malignant nodules and normal adjacent thyroid tissue were downloaded from GEO database to analyze DMRs. 11726 DMRs were identified, and after annotation to remove DMRs located in the distal intergenic region, 6599 DMRs were finally obtained. The malignant nodules had 2701 hypermethylated DMRs and 3898 hypomethylated DMRs compared with normal adjacent thyroid tissue. The global levels of DNA methylation in malignant nodules DMRs were higher than those in normal adjacent thyroid tissue (Figure 7A). Through boxplot of DMRs methylation levels in two groups, it was found that in promoter, exon, intron and 3'UTR genomic regions, the DNA methylation levels of malignant nodules were higher than those of normal adjacent thyroid tissue (Figure 7B). The hypermethylated DMRs in genebody region (including 5 'UTR, exon, 3' UTR, intron) enrich in GO terms including cell-substrate adhesion (Figure 7C). Unfortunately, the hypermethylated DMRs in promoter region did not enrich in GO terms. The KEGG pathways of the hypermethylated DMRs in both regions are related to axon guidance and parathyroid hormone synthesis, secretion and action (Figure S4A). Notably, the GO terms of regulation of T cell and leukocytes were shown by hypomethylated DMRs in promoter region (Figure 7D). The hypomethylated DMRs in genebody region show proximity to genes associated with cell junction assembly and synapse organization (Figure 7E). The KEGG pathways of the hypomethylated DMRs in both regions are related to hormone secretion (Figure S4B). In order to further clarify whether genes that function regulation of T cell and leukocyte and hypomethylated in promoter region were involved in the progression of PTC, we performed Protein-Protein Interaction Networks (PPI) analysis of these genes together with LEF1, TNFRSF13C, SHLD2 and SHLD3. ACTB, SPN and SHLD1 were found to have strong interactions with LEF1, TNFRSF13C, SHLD2 and SHLD3 (Figure 7F). Moreover, IGV analysis showed that the promoter regions of ACTB, SPN and SHLD1 were hypomethylated (Figure 7G). Taken together, we speculated that PTC would cause changes in DNA methylation levels, resulting in DNA methylation in the promoter region of regulation of T cell and leukocyte related genes, thus promoting the expression of LEF1, TNFRSF13C, SHLD2 and SHLD3.

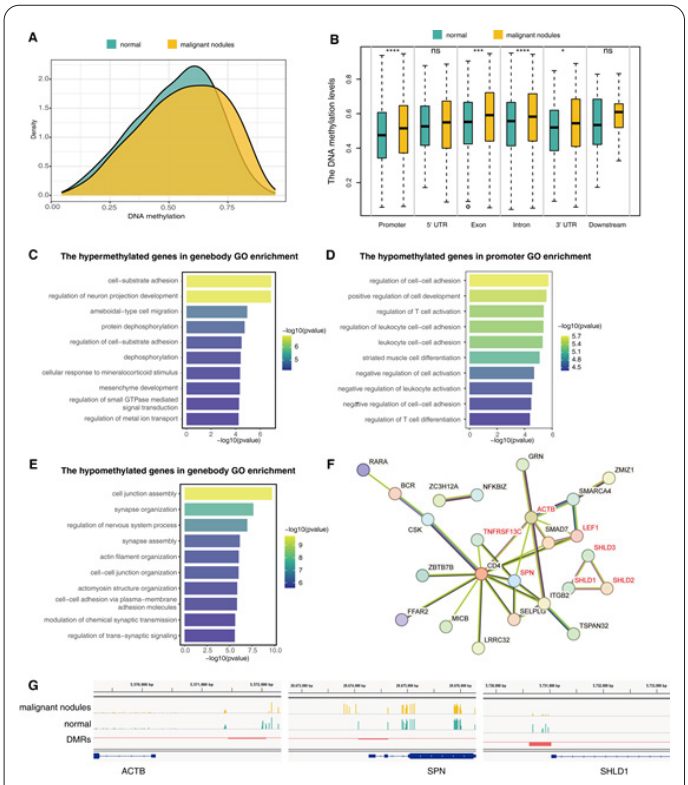

Fig. 7. The DNA methylation patterns of thyroid malignant nodules and normal adjacent thyroid tissue. (A) Distribution of DNA methylation levels in thyroid malignant nodules and normal adjacent thyroid tissue. (B) The DNA methylation levels of DMRs in thyroid malignant nodules and normal adjacent thyroid tissue. The DNA methylation levels of DMRs at indicated genomic regions in thyroid malignant nodules and normal adjacent thyroid tissue. *, $P < 0.05$, ***, $P < 0.001$, ****, $P < 0.0001$ (Student's t-test). (C) The top enriched GO terms of hypermethylated genes of genebody region in thyroid malignant nodules compared to normal adjacent thyroid tissue. (D) The top enriched GO terms of hypomethylated genes of promoter region in thyroid malignant nodules compared to normal adjacent thyroid tissue. (E) The top enriched GO terms of hypomethylated genes of genebody region in thyroid malignant nodules compared to normal adjacent thyroid tissue. (F) PPI network diagram. (G) The DNA methylation pattern in genomic loci of ACTB, SPN and SHLD1. DMRs, differentially methylated regions.

3.5. Naive and regulatory B-cells expressing LEF1, TNFRSF13C, SHLD2 and SHLD3 participate in the transcriptional regulation of the SAT in obesity and PTC

We next used scRNA-seq of the SAT in obese individuals and cancerous tissue in PTC to explain the mechanism of B-cell subtypes regulation of the increased risk of PTC in obesity. First, we identified a total of 10 cell subpopulations in the SAT of three obese individuals (Figure 8A and Figure S5A). Prognosis-related genes LEF1, TNFRSF13C, SHLD2 and SHLD3 were all identified to be highly expressed in B-cells (Figure 8B). Plasma cells1, Plasma cells2, Naive B-cells, and Regulatory B-cells were the four subpopulations of B-cells identified by the SAT in obese individuals (Figure 8C and Figure S5B) (marker genes were BIRC3, CD27, IGHD, and CD248). The expressions of LEF1, TNFRSF13C, SHLD2 and SHLD3 were more common in naïve and regulatory B-cells (Figure 8D). To determine whether similar B-cell alterations were also present in PTC, scRNA-seq was analyzed in six PTC patients and one paraneoplastic tissue. A total of 11 cell subpopulations were identified (Figure 8E and Figure S5C), and similarly, LEF1, TNFRSF13C, SHLD2 and SHLD3 were all highly expressed in B-cells (Figure 8F). Subsequently,

B-cells of PTC were divided into six subpopulations, such as Precursor B-cells, Plasma cells1, Plasma cells2, Plasma cells3, Naive B-cells, and Regulatory B-cells (Figure 8G and Figure S5D) (marker genes HIST1H4C, IGHM, CD27, IGHA1, FCER2, and CD69 respectively). As with the SAT in obese people, the expression levels of LEF1, TNFRSF13C, SHLD2 and SHLD3 were also highly expressed in naive and regulatory B-cells (Figure 8H). By using Pearson correlation analysis to evaluate the relation between the SAT in obese individuals and PTC for naive and regulatory B-cells gene expression, we found that the SAT in obese individuals and PTC for naive B-cells had a Pearson correlation coefficient of 0.92 (Figure 8I), while the Pearson correlation coefficient for regulatory B-cells was 0.84. (Figure 8J). In summary, the high risk of PTC in obesity may be explained by the comparable naive and regulatory B-cell transcriptional patterns.

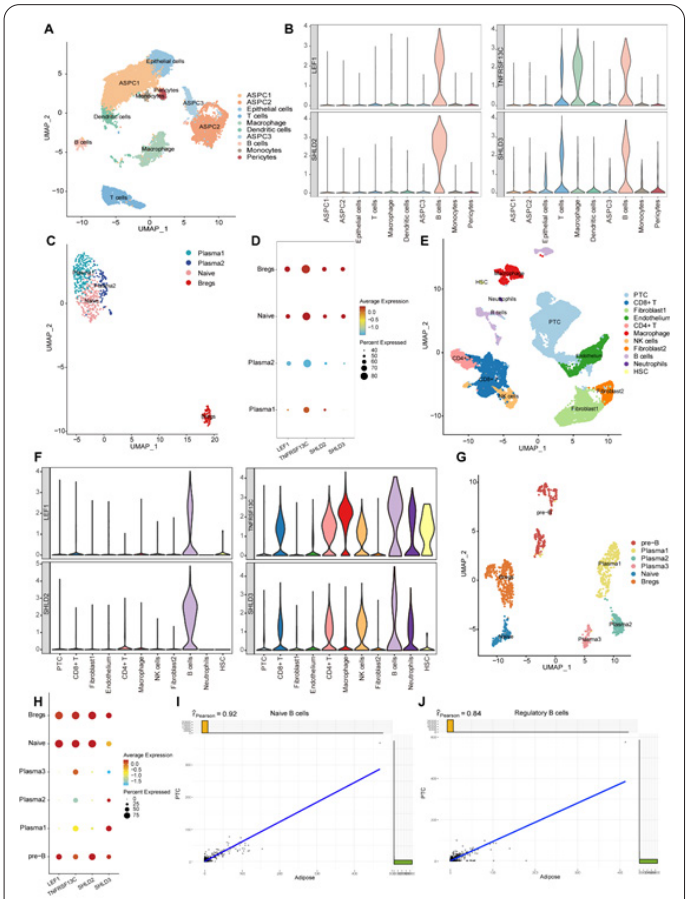

Fig. 8. The Single-cell atlas of the SAT in obesity and PTC. (A) The UMAP diagram shows 10 different cell types of the SAT in obesity. (B) Violin plots showing the expression levels of LEF1, TNFRSF13C, SHLD2 and SHLD3 in each cluster of the SAT in obesity. The Y-axis is the log scale normalized read count. (C) The UMAP diagram shows 4 different cell types of B-cells in the SAT in obesity. (D) Dot plots depicting the expression of LEF1, TNFRSF13C, SHLD2 and SHLD3 in different cell types of B-cells in the SAT in obesity. (E) The UMAP diagram shows 11 different cell types of the PTC. (F) Violin plots showing the expression levels of LEF1, TNFRSF13C, SHLD2 and SHLD3 in each cluster of the PTC. The Y-axis is the log scale normalized read count. (G) The UMAP diagram shows 6 different cell types of B-cells in the PTC. (H) Dot plots depicting the expression of LEF1, TNFRSF13C, SHLD2 and SHLD3 in different cell types of B-cells in the PTC. (I) Linear correlation analysis of naive B-cells from the SAT in obesity and PTC.(J) Linear correlation analysis of regulatory B-cells from the SAT in obesity and PTC.

4. Discussion

The results of this research, which were based on the work of numerous previous clinical research studies, demonstrated a strong correlation between obesity and papillary thyroid carcinoma [32]. Therefore, it is crucial to identify common therapeutic targets for both, and transcriptome and DNA methylation investigations can reveal the genes responsible for both illness onset and therapy [33], so this work was carried out.

We downloaded the expression profile data of 16 patients with PTC as well as people who were healthy and obese from GEO. It was found that obese tissues and PTC shared a large number of genes, which indicates that there is a large amount of commonality in gene expression between the two diseases, and when statistical analysis of these co-expression pathways was conducted, we found that several fat metabolism transport and other pathways were jointly under-expressed, while the pathways related to immunological inflammation were over-expressed. These pathways were shown to co-express significantly. Apolipoprotein E (*APOE*) affects the m6A modification in PTC and initiates immunological signaling pathways, according to the literature [16]. Then, we conducted a follow-up analysis of DEG intersections between obese and normal SAT and found that more immune-related pathways were enriched in GO terms, such as T cell differentiation regulation-related pathways. A substantial literature also suggests that obesity itself causes immune system dysregulation and tends to increase the risk of infection, especially tumors [17, 18]. This is in line with the results as well, and the KEGG pathway contains comparable disease-related pathways. To further improve the reliability of the results, WGCNA was performed on the SAT and PTC tissue data, and the results revealed four major modules that were significantly linked to immunological diseases, as well as the formation of a few intracellular structures and ribosome assembly translation. This indicates that both diseases may impact translation and may potentially be connected to intracellular glycometabolism. As a result, we further intersected and checked the genes from the three previous analysis methods for extensive analysis.

Combining the GSVA genes, co-expressed genes obtained from DEGs analysis, and genes in the WGCNA modules, we discovered that these shared genes are present in pathways connected to B-cell activation, proliferation, and regulation. This finding suggests that these two diseases result in a greater number of immune system changes. Analysis based on the TCGA-THCA cohorts confirms that B-cell regulatory-related genes affected the prognosis of thyroid cancer patients, while the expression of LEF1, TNFRSF13C, SHLD2 and SHLD3 might serve as the independent factor for predicting the prognosis. Although the global DNA methylation levels of PTC are still uncertain, abnormal DNA methylation changes affect the grade and stage of PTC [34]. In our study, PTC causes hypomethylation in promoter region of regulation of T cell and leukocyte-related genes, in which ACTB, SPN and SHLD1 interact with LEF1, TNFRSF13C, SHLD2 and SHLD3. Notably, beta-actin ACTB plays a key role in cytoskeleton maintenance and is stable expression in thyroid cancer [35]. It has been reported that upregulated expression of LEF1 could promote the proliferation of PTC [36]. TNFRSF13C can serve as a prognostic biomarker for co-

lon adenocarcinoma [37], it can also serve for thyroid cancer in our study. As the shieldin complex, SHLD1, SHLD2 and SHLD3 control DNA repair pathway choice by counteracting DNA end-resection and participating in immunoglobulin class switching [38]. We speculate that these genes are involved in the regulation of B-cells in PTC and exhibit similar changes in SAT in obesity. To confirm this suspicion we downloaded single-cell expression data for both diseases. Firstly, in the obese population, we found that LEF1, TNFRSF13C, SHLD2 and SHLD3 were highly expressed in B-cells, especially in naive and regulatory B. Following that, we investigated the single-cell expression profiles of PTC patients and related paraneoplastic tissues and discovered that LEF1, TNFRSF13C, SHLD2 and SHLD3 were similarly highly expressed in B-cells from PTC patients. These cells were also mostly identified in naive and regulatory B-cells. Both diseases were shown to have a strong Pearson correlation coefficient between naive and regulatory B-cells. Therefore, the expression patterns of naive and regulatory B-cells could explain the important relationship between PTC and obesity.

Our study has some limitations when compared to other studies; for instance, disease development might be regulated by multiple histological levels, and we have only preliminary investigated the co-regulatory mechanisms of obesity and PTC at the transcriptome and DNA methylation level. For instance, research on RNA methylation has been extensively used to explain how various diseases develop [39]. More research is required on these degrees of modifications. In summary, our study highlights potential transcriptomic regulatory pathways of obesity and PTC. The core genes, molecular mechanisms, and prospective therapeutic targets that control obesity and PTC were investigated through a thorough mapping of the transcription pattern. It provides a new viewpoint and supporting data to reduce the elevated incidence of papillary PTC in people brought on by obesity.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Data availability

The datasets presented in this study can be found in online repositories (GEO database). The names of the repository/repositories and accession number(s) can be listed below.

Repository/Repositories	Accession Number
Gene Expression Omnibus	GSE205668
Gene Expression Omnibus	GSE165724
DNA Methylation Signatures	GSE107738
Single Cell Expression	GSE191288
Single Cell Expression	GSE163830

Competing interests

The authors declare no potential conflicts of interest.

Funding

This study was supported by the Postdoctoral Research Fund of Chaoyang District, Beijing, China in 2022. Natural Science Foundation for Key Programs of China Grants (82130065), National Natural Science Foundation of China (82002330), FENG foundation (FFBR 202103).

Authors' contributions

JY designed the study. ZY performed data analysis. ZY and ZX prepared the figures and tables. ZY, ZX, and SL wrote the manuscript and approved the final draft. ZY, ZX, and ZT participated in data interpretation and analysis. ZY, ZX, and SL were involved in proofreading and deep editing and approved the final manuscript. JY devised the main conceptual idea and supervised the project, performed proofreading and deep editing of the manuscript, and approved the final draft. All authors contributed to the article and approved the submitted version.

Acknowledgments

We are grateful to the researchers who built the public database and the researchers who shared GSE205668, GSE165724, GSE191288, GSE163830 and GSE107738 data sets, which made our study possible through their generous contributions.

References

1. Loos R, Yeo G (2022) The genetics of obesity: from discovery to biology. Nat Rev Genet 23:120-133. doi: 10.1038/s41576-021-00414-z
2. (2017) Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128.9 million children, adolescents, and adults. Lancet 390:2627-2642. doi: 10.1016/S0140-6736(17)32129-3
3. Jebeile H, Kelly AS, O'Malley G, Baur LA (2022) Obesity in children and adolescents: epidemiology, causes, assessment, and management. Lancet Diabetes Endo 10:351-365. doi: 10.1016/S2213-8587(22)00047-X
4. Davis LL, Nolan MZ (2021) The Influence of Obesity on Care of Adults with Cardiovascular Disease. Nurs Clin N Am 56:511-525. doi: 10.1016/j.cnur.2021.07.002
5. Rohm TV, Meier DT, Olefsky JM, Donath MY (2022) Inflammation in obesity, diabetes, and related disorders. Immunity 55:31-55. doi: 10.1016/j.immuni.2021.12.013
6. Franchini F, Palatucci G, Colao A, Ungaro P, Macchia PE, Nettore IC (2022) Obesity and Thyroid Cancer Risk: An Update. Int J Env Res Pub He 19:1116. doi: 10.3390/ijerph19031116
7. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M (2008) Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. Lancet 371:569-578. doi: 10.1016/S0140-6736(08)60269-X
8. Iyengar NM, Gucalp A, Dannenberg AJ, Hudis CA (2016) Obesity and Cancer Mechanisms: Tumor Microenvironment and Inflammation. J Clin Oncol 34:4270-4276. doi: 10.1200/JCO.2016.67.4283
9. Butler LM, Perone Y, Dehairs J, Lupien LE, de Laat V, Talebi A et al (2020) Lipids and cancer: Emerging roles in pathogenesis, diagnosis and therapeutic intervention. Adv Drug Deliver Rev 159:245-293. doi: 10.1016/j.addr.2020.07.013
10. Hoy AJ, Nagarajan SR, Butler LM (2021) Tumour fatty acid metabolism in the context of therapy resistance and obesity. Nat Rev Cancer 21:753-766. doi: 10.1038/s41568-021-00388-4
11. Guleria P, Srinivasan R, Rana C, Agarwal S (2022) Molecular Landscape of Pediatric Thyroid Cancer: A Review. Diagnostics 12:3136. doi: 10.3390/diagnostics12123136

12. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A et al (2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. Ca-Cancer J Clin 71:209-249. doi: 10.3322/caac.21660
13. Matrone A, Ferrari F, Santini F, Elisei R (2020) Obesity as a risk factor for thyroid cancer. Curr Opin Endocrinol 27:358-363. doi: 10.1097/MED.0000000000000556
14. Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K (2016) Body Fatness and Cancer--Viewpoint of the IARC Working Group. New Engl J Med 375:794-798. doi: 10.1056/NEJMsr1606602
15. Matrone A, Basolo A, Santini F, Elisei R (2022) Understanding the effect of obesity on papillary thyroid cancer: is there a need for tailored diagnostic and therapeutic management? Expert Rev Endocrino 17:475-484. doi: 10.1080/17446651.2022.2131529
16. Wu WX, Feng JW, Ye J, Qi GF, Hong LZ, Hu J et al (2022) Influence of Obesity Parameters on Different Regional Patterns of Lymph Node Metastasis in Papillary Thyroid Cancer. Int J Endocrinol 2022:3797955. doi: 10.1155/2022/3797955
17. Cui N, Sun Q, Chen L (2021) A meta-analysis of the influence of body mass index on the clinicopathologic progression of papillary thyroid carcinoma. Medicine 100:e26882. doi: 10.1097/MD.0000000000026882
18. O'Neill RJ, Abd ES, Kerin MJ, Lowery AJ (2021) Association of BMI with Clinicopathological Features of Papillary Thyroid Cancer: A Systematic Review and Meta-Analysis. World J Surg 45:2805-2815. doi: 10.1007/s00268-021-06193-2
19. Le Moli R, Vella V, Tumino D, Piticchio T, Naselli A, Belfiore A et al (2022) Inflammasome activation as a link between obesity and thyroid disorders: Implications for an integrated clinical management. Front Endocrinol 13:959276. doi: 10.3389/fendo.2022.959276
20. Zhou Y, Yang Y, Zhou T, Li B, Wang Z (2021) Adiponectin and Thyroid Cancer: Insight into the Association between Adiponectin and Obesity. Aging Dis 12:597-613. doi: 10.14336/AD.2020.0919
21. Muller TD, Bluher M, Tschop MH, DiMarchi RD (2022) Anti-obesity drug discovery: advances and challenges. Nat Rev Drug Discov 21:201-223. doi: 10.1038/s41573-021-00337-8
22. Ory C, Ugolin N, Levalois C, Lacroix L, Caillou B, Bidart JM et al (2011) Gene expression signature discriminates sporadic from post-radiotherapy-induced thyroid tumors. Endocr-Relat Cancer 18:193-206. doi: 10.1677/ERC-10-0205
23. Guo H, Zhu P, Yan L, Li R, Hu B, Lian Y et al (2014) The DNA methylation landscape of human early embryos. Nature 511:606-610. doi: 10.1038/nature13544
24. Xu Z, Yu Z, Chen M, Zhang M, Chen R, Yu H et al (2022) Mechanisms of estrogen deficiency-induced osteoporosis based on transcriptome and DNA methylation. Front Cell Dev Biol 10:1011725. doi: 10.3389/fcell.2022.1011725
25. Aquino, E. M. et al. (2018) Current Understanding of DNA Methylation and Age-related Disease. OBM Genetics 2:1-1. doi: 10.21926/obm.genet.1802016.
26. Li Z, Jiang N, Lim EH, Chin W, Yeoh AE (2021) Role of transcriptome sequencing in clinical diagnosis of B-cell acute lymphoblastic leukemia. Leukemia 35:2135-2137. doi: 10.1038/s41375-021-01185-6
27. Cao S, Lin Y, Yang Z (2019) Vaccinia Virus Transcriptome Analysis by RNA Sequencing. Methods Mol Biol 2023:157-170. doi: 10.1007/978-1-4939-9593-6_10
28. Hu B, Yang XB, Sang XT (2020) Development and Verification of the Hypoxia-Related and Immune-Associated Prognosis Signature for Hepatocellular Carcinoma. J Hepatocell Carcino 7:315-330. doi: 10.2147/JHC.S272109
29. Park Y, Wu H (2016) Differential methylation analysis for BS-seq data under general experimental design. Bioinformatics 32:1446-1453. doi: 10.1093/bioinformatics/btw026
30. Yu G, Wang LG, He QY (2015) ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics 31:2382-2383. doi: 10.1093/bioinformatics/btv145
31. Hu J, Chen Z, Bao L, Zhou L, Hou Y, Liu L et al (2020) Single-Cell Transcriptome Analysis Reveals Intratumoral Heterogeneity in ccRCC, which Results in Different Clinical Outcomes. Mol Ther 28:1658-1672. doi: 10.1016/j.ymthe.2020.04.023
32. Kitahara CM, Neta G, Pfeiffer RM, Kwon D, Xu L, Freedman ND et al (2012) Common obesity-related genetic variants and papillary thyroid cancer risk. Cancer Epidem Biomar 21:2268-2271. doi: 10.1158/1055-9965.EPI-12-0790
33. Hong M, Tao S, Zhang L, Diao LT, Huang X, Huang S et al (2020) RNA sequencing: new technologies and applications in cancer research. J Hematol Oncol 13:166. doi: 10.1186/s13045-020-01005-x
34. Zafon C, Gil J, Perez-Gonzalez B, Jorda M (2019) DNA methylation in thyroid cancer. Endocr-Relat Cancer 26:R415-R439. doi: 10.1530/ERC-19-0093
35. Weber R, Bertoni AP, Bessestil LW, Brasil BM, Brum LS, Furlanetto TW (2014) Validation of reference genes for normalization gene expression in reverse transcription quantitative PCR in human normal thyroid and goiter tissue. Biomed Res Int 2014:198582. doi: 10.1155/2014/198582
36. Jiang C, He J, Xu S, Wang Q, Cheng J (2022) NR4A1 promotes LEF1 expression in the pathogenesis of papillary thyroid cancer. Cell Death Discov 8:46. doi: 10.1038/s41420-022-00843-7
37. Huang W, Su D, Liao X, Yang T, Lu Y, Zhang Z (2023) Prognostic costimulatory molecule-related signature risk model correlates with immunotherapy response in colon cancer. Sci Rep-Uk 13:789. doi: 10.1038/s41598-023-27826-7
38. Mirman Z, Lottersberger F, Takai H, Kibe T, Gong Y, Takai K et al (2018) 53BP1-RIF1-shieldin counteracts DSB resection through CST- and Polalpha-dependent fill-in. Nature 560:112-116. doi: 10.1038/s41586-018-0324-7
39. Huang J, Sun W, Wang Z, Lv C, Zhang T, Zhang D et al (2022) FTO suppresses glycolysis and growth of papillary thyroid cancer via decreasing stability of APOE mRNA in an N6-methyladenosine-dependent manner. J Exp Clin Canc Res 41:42. doi: 10.1186/s13046-022-02254-z