# Mining Prognostic Significance Genes in Human Thyroid Cancer Using Bioinformatics Analysis

**Hui Zhao**
Second Affiliated Hospital of Harbin Medical University
**Pengjie Li**
Second Affiliated Hospital of Harbin Medical University
**Junjian Li**
First Affiliated Hospital of Wenzhou Medical University
**Lian Duan**
First Affiliated Hospital of Wenzhou Medical University
**Yanzhu Jiao**
Second Affiliated Hospital of Harbin Medical University
**Shuang Li**
Second Affiliated Hospital of Harbin Medical University
**Yanni Bai**
Second Affiliated Hospital of Harbin Medical University
**Xiaqing Zhang**
Second Affiliated Hospital of Harbin Medical University
**Meihua Liang** ( 15146113366@163.com )
Second Affiliated Hospital of Harbin Medical University
**Xiaoyi Huang**
Third Affiliated Hospital of Harbin Medical University

**Research Article**

**Keywords:** thyroid cancer, reprogramming, differential expression, functional annotation, DNA methylation

**Posted Date:** January 5th, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-137529/v1

## Abstract

**Background** Thyroid carcinoma (THC) is very common, yet its pathogenesis and the key tumor marker genes remain unclear.

**Methods** Gene expression datasets from the Gene Expression Omnibus (GEO) and the Cancer Genome Atlas Project (TCGA) were used for gene differential expression analysis. Functional annotation analysis, Clinical prognosis analysis and Differential DNA methylation analysis were conducted on the differentially expressed genes (DEGs).

**Results** Compared with induced pluripotent stem cells (iPSCs), 237 differentially expressed THC intersection genes derived from GEO and TCGA were obtained, of which 153 genes were closely related to clinicopathological features and prognostic effects. Biological function analysis indicated that most of these DEGs were involved in the proteinaceous extracellular matrix, epithelial-to-mesenchymal transition (EMT), and PI3K-Akt signaling pathway, resulting in effects on tumor invasion and metastasis. Finally, the results of differential methylation levels demonstrated that the high expression of 4 genes (*CHI3L1*, *NFE2L3*, *S100A2*, and *LAMB3)* was strongly correlated with the development of thyroid cancer.

**Conclusions** Proteinaceous extracellular matrix, EMT, and PI3K-Akt signaling pathways were of great significance in the metastasis and invasion of THC. Genes such as *CHI3L1*, *NFE2L3*, *S100A2,* and *LAMB3* were susceptible to THC.


## Background

Thyroid carcinoma (THC), the disease with the highest incidence among endocrine tumors, ranking eighth amongst the world's most common cancers had continued to increase in morbidity and mortality over the past decade [1]. According to data stratified by gender, the incidence of thyroid cancer in women was much higher compared to men [2]. In terms of geographical distribution, the incidence of thyroid cancer in developed eastern regions was much higher than that in underdeveloped central and western regions, indicating that environmental factors had a significant effect on the incidence of thyroid disease [3, 4]. Particularly, exposure to ionizing radiation during childhood or adolescence [5], as well as exposure to low doses of radiation was found to increase the likelihood of the onset of thyroid nodules and THC [6]. Surgical resection of thyroid cancer and treatment with radioactive iodine were effective, however, examinations such as preoperative CT scans and fine-needle aspiration (FNA) generally did not provide a definitive diagnosis. Therefore, diagnostic semi-thyroidectomy was often observed in uncertain cases, accounting for 75% of benign thyroid masses; malignant tumors were approximately ten times more likely to undergo surgery than benign thyroid masses such as thyroid tumors and cysts [7]. It was of great significance to study the pathogenesis of THC by genomic technology in order to improve the level of preoperative diagnosis. Thyroid cancer may be associated with a variety genes, but currently, there have been only a few genes reported to play a crucial role in thyroid cancer, such as *MET*, *LGALS3*, *KRT19*, *DPP4*, *MDK*, *TIMP1*, and *FN1* [8]. Large-scale, high-throughput, high-sensitivity, and high-precision gene sequence analysis and biological function analysis at the whole genome level can be realized using bioinformatics technology to analyze gene expression profile chip data. The process of somatic reprogramming, especially on induced pluripotent stem cells (iPSCs), was very similar to that of carcinogenesis [9]; IPSCs and cancer cells had similar signaling pathways and cell metabolism characteristics [10, 11]. The most important characteristic of iPSCs was their ability to self-renew and differentiate, which also occurred during tumorigenesis. Moreover the abnormal reprogramming of iPSCs caused tumors. In this study, DEGs in GEO and TCGA were excavated. Additionally, the expression, gene function, and signaling pathway between THC and iPSCs were compared by integrated expression profiling. Furthermore, DEGs was screened as THC biomarkers based on clinical information, and we preformed biological function analysis and methylation level analysis of CpG islands.

In this study, we explored the pathogenesis of THC, and discovered the potential therapeutic targets at the molecular level by comparing iPSCs with the high-throughput expression data. Hopefully, the results of this study could provide a new perspective for the research of THC pathogenesis.

## Materials And Methods

### Microarray Datasets of GEO database

Datasets were Conditions for the microarray datasets selected from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) and used in the analysis were described as follows: (1) iPSCs must be compared to normal somatic cells and induced by at least 3 factors (such as Oct4, Sox2, and c-Myc), (2) a case-control study of THC gene expression profiles based on peripheral blood and/or tissue samples, (3) the total number of samples available for each study ≥ 2, and (4) the strength of raw data cell file (.CEL) was available. The research designed by methods such as siRNA interference, gene overexpression, gene knockout, any drug therapy, and cancer cell lines was excluded. We downloaded the Affymetrix Human Genome U133 Plus 2.0 platform file and carried out normalization of raw microarray data using the Robust Multichip Averaging (RMA) method of affy packages of R (v. 3.6.3). Batch calibration was performed by the ComBat function of the sva package. Two independent sample t-test was used to analyze the gene differential expression, and a false discovery rate (FDR) was used for multiple test correction. After DEGs were screened with adjust probability (p) values (p□0.05), |logFC| ≥ 2 indicated up-regulated gene expression, and |logFC| ≤ 0.5 suggested down-regulation of gene expression.

#### Data collection from the TCGA

The expression profiles and clinical data of THC were downloaded from the TCGA database (http://xena.ucsc.edu/public). DEGs were screened using the Illumina HiSeq 2000 RNA Sequencing platform. Differential expression analysis was performed with two independent sample t-testing and multiple test correction. DEGs were screened with $P < 0.05$. Among them, |logFC| ≥ 2 demonstrated up-regulated gene expression, and |logFC| ≤ 0.5 referred to down-regulation of gene expression.

## Functional annotation analysis of DEGs

Gene enrichment analysis (GO function annotation and KEGG pathway analysis) was analyzed using DAVID (http://david.abcc.ncifcrf.gov/). GO functions were annotated from three categories: cellular components (CC), biological processes (BP), and molecular functions (MF). GO terms or KEGG pathways with $P < 0.05$ were selected as the significant enrichment functions or pathways.

## Clinical prognosis analysis of DEGs

To analyze whether the DEGs were related to clinical pathology, we performed Kaplan-Meier survival analysis and One-way Analysis of Variance (ANOVA) to identify the candidate genes with the $P < 0.05$ as the cutoff. Clinical information, such as clinical staging (initial weight, positive rate of lymph node metastasis, neoplasm depth, and extracorporeal extension), TNM staging, pathologic stage, histological type, prognosis status, residual tumor, the primary site of the tumor, and survival time were analyzed and adjusted for gender and age in this study.

## Differential DNA methylation analysis

The methylation profile from the Illumina Infinium Human Methylation 450 was downloaded from the TCGA database to analyze the differential methylation levels of CpG islands in DEGs. The Beta-value for quantifying the methylation level was calculated as Methylation/ (Methylation+Unmethylation+100), which was also the ratio of the methylation degree to the total extent within a range from 0 to 1. The samples containing more than 20% of the missing values were excluded, and the remaining samples adopted the median of the total β-value to replace the missing values.

## Results

# Screening of THC DEGs

The differential expression analysis on the datasets, including GSE65144 [12], GSE53157 [13], GSE3678, and GSE29265, revealed 3,982 DEGs with statistical significance. Compared with the control group, 599 up-regulated DEGs (*FN1*, *RUNX1*, and *TIMP1*) and 283 down-regulated genes (*DPP6*, *MPPED2*, and *PID1* ) were obtained. Meanwhile, 13,438 DEGs were obtained in TCGA, including 507 cases and 59 controls. Of these DEG's, 1,556 genes were found to be up-regulated and 1,099 genes down-regulated. Finally, 538 overlapping DEGs were observed in the two datasets, including 202 down-regulated genes and 336 up-regulated genes (Fig. 1; Fig. 2).

# Screening Of Reprogramming Degs

Ultimately, 14 valid GSE series were included: 148 iPSCs and 60 somatic cells. The details of the collected datasets are illustrated in Table 1. After we normalized and transformed the microarray data, we preformed two-tailed Student's t-tests for gene differential expression analysis. Among iPSCs and somatic cells, 15,466 DEGs were statistically significant, 1,589 genes were up-regulated in iPSCs and 2,408 genes were down-regulated.

Table 1
Reprogramming-GEO. Microarray datasets of the different induced pluripotent stem cells and the somatic cells were included in our study from GEO database

| GEO | Submission | Method of | Reprogramming | Donor cells | iPSCs | SCs |
|---|---|---|---|---|---|---|
| number | date | induction | factors | | | |
| GSE9832 | 10-Dec,2007 | Inducible system | OSKM | Fibroblasts | 8 | 7 |
| GSE12390 | 8-Aug,2008 | Retroviral | OSKM, OSKMN | Fibroblasts and K | 15 | 3 |
| GSE12583 | 27-Aug,2008 | Retroviral | OSK, OSKM | FFs and K | 3 | 3 |
| GSE15148 | 6-Mar,2009 | Retroviral | OSKM | FFs | 16 | 2 |
| GSE9709 | 28-Nov,2007 | Retroviral | OSKM | Skin-derived cells | 8 | 2 |
| GSE18111 | 15-Sep,2009 | Inducible system | OSKM | Fibroblasts | 6 | 6 |
| GSE22499 | 22-Jun,2010 | Inducible system | OSK | Fibroblasts | 22 | 3 |
| GSE22246 | 9-Jun,2010 | Retroviral | OSKM | Fibroblasts | 8 | 2 |
| GSE23583 | 12-Aug,2010 | Retroviral | OSKM | Fibroblasts | 21 | 9 |
| GSE24182 | 16-Sep,2010 | Retroviral | OSKM | Fibroblasts | 7 | 5 |
| GSE26672 | 18-Jan,2011 | Retroviral | OSKMNL | FFs | 12 | 2 |
| GSE27186 | 9-Feb,2011 | Retroviral | OSKM | CB and neonatal K | 4 | 4 |
| GSE48830 | 12-Jul,2013 | Retroviral | OSKM | HDFs | 4 | 2 |
| GSE76832 | 13-Jan,2016 | Retroviral | OSKM | Adult HDFs | 14 | 10 |

iPSCs, induced pluripotent stem cells; SCs, somatic cells; K, keratinocyte; FFs, foreskin fibroblasts; CB, cord blood; HDFs, human dermal fibroblasts; K, KLF4; L, LIN28; M, c-MYC; N, NANOG; O, OCT4; S, SOX2.

## Similarities And Differences Between Tumorigenesis And Somatic Reprogramming

The similarities and differences between tumorigenesis and somatic reprogramming were compared to further investigate the pathogenesis of THC. From the intersection of THC and iPSCs, a total of 237 DEGs were overlapped (Fig. 3).

Additionally, GO and KEGG functional enrichment analyses were performed on the 237 DEGs to analyze their biological functions. Results from GO enrichment analysis indicated that the 47 common up-regulated DEGs were mainly concentrated in areas responsible for transferase activity, positive regulation of synaptic transmission, negative regulation of keratinocyte proliferation, cell migration, and signal transduction (Fig. 4). On the other hand, The 48 down-regulated genes were concentrated in areas regarding mesenchymal cell differentiation, epithelial to mesenchymal transition (EMT), positive regulation of Wnt signaling pathway, and negative regulation of epithelial cell proliferation. 113 genes were up-regulated in THC but down-regulated in iPSCs. Of these genes, significant GO terms were concentrated in the extracellular matrix (ECM), proteinaceous extracellular matrix, transcriptional regulatory region, and DNA-template. The KEGG pathway annotation demonstrated that these genes were concentrated in ECM-receptor interaction, Focal adhesion, Pathways in cancer, and Small cell lung cancer (Fig. 5). 29 genes were down-regulated in THC but up-regulated in iPSCs, of which, significant GO terms were concentrated in morphogenesis of epithelium, tissue morphogenesis, epithelium development, cell surface GO, extracellular region, and calcium ion binding. The KEGG pathway annotation illustrated that these genes were concentrated in the Hedgehog signaling pathway.

## Clinical Significance Of Degs

Combined with the clinicopathological and prognostic features of THC, a total of 153 DEGs were obtained and related with at least one clinical characteristic uncorrelated with age and gender (Table 2).

Table 2
The results of clinicopathological features and prognostic analysis for 153 DEGs

| Gene | OS. Time | RFS. Time | Age | Gender | Weight | positive rate of lymph node metastasis | extracorporeal extension | pathologic stage | histological type | neoplasm depth | TNM M | TNM N |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABCA8 | 0.115 | 0.334 | 0.870 | 0.652 | 0.974 | 0.360 | * | 0.709 | * | 0.402 | 0.698 | * |
| GFRA1 | * | * | 0.748 | 0.532 | 0.494 | 0.178 | 0.927 | 0.945 | 0.757 | 0.352 | 0.337 | 0.618 |
| PID1 | 0.108 | 0.133 | 0.720 | 0.960 | 0.726 | 0.889 | 0.703 | 0.387 | 0.440 | 0.058 | * | 0.261 |
| NELL2 | 0.908 | 0.712 | 0.095 | 0.355 | 0.920 | 0.944 | 0.286 | 0.397 | * | 0.140 | 0.128 | 0.314 |
| RGS4 | * | * | 0.497 | 0.503 | 0.147 | 0.919 | 0.600 | 0.895 | * | 0.962 | 0.188 | 0.496 |
| RELN | 0.989 | 0.922 | 0.889 | 0.403 | 0.636 | 0.443 | 0.560 | 0.769 | 0.218 | * | * | * |
| FAM43A | * | * | 0.984 | 0.955 | 0.761 | * | * | * | * | 0.847 | * | * |
| HBB | * | * | 0.917 | 0.190 | 0.121 | 0.270 | 0.885 | 0.890 | 0.540 | 0.748 | 0.880 | 0.325 |
| SLCO4C1 | 0.622 | 0.540 | 0.997 | 0.506 | 0.193 | 0.638 | 0.978 | 0.784 | * | 0.996 | 0.768 | 0.455 |
| S100A2 | * | * | 0.124 | 0.717 | 0.119 | 0.264 | 0.346 | * | * | 0.499 | 0.734 | 0.063 |
| EPPK1 | 0.079 | 0.164 | 0.207 | 0.810 | 0.686 | 0.221 | 0.994 | 0.470 | * | 0.088 | 0.103 | 0.199 |
| APOC1 | 0.726 | 0.833 | 0.129 | 0.761 | 0.619 | 0.685 | 0.120 | 0.410 | * | 0.381 | 0.486 | * |
| CCDC146 | * | * | 0.984 | 0.455 | 0.947 | * | * | 0.197 | * | 0.132 | 0.948 | * |
| MKX | 0.864 | 0.649 | 0.186 | 0.911 | 0.411 | 0.582 | 0.712 | 0.488 | 0.410 | 0.582 | 0.446 | * |
| G0S2 | 0.231 | 0.381 | 0.547 | 0.543 | 0.617 | 0.319 | * | 0.646 | 0.117 | 0.549 | 0.381 | * |
| DTL | 0.893 | 0.468 | 0.224 | 0.251 | 0.298 | 0.091 | 0.562 | 0.114 | * | 0.720 | 0.627 | 0.428 |
| RHOBTB3 | 0.540 | 0.633 | 0.686 | 0.231 | 0.818 | 0.220 | 0.104 | 0.315 | 0.238 | 0.272 | 0.650 | 0.163 |
| DIRAS3 | 0.852 | 0.930 | 0.970 | 0.167 | 0.949 | 0.306 | 0.612 | 0.741 | * | * | 0.845 | 0.090 |
| CYP1B1 | 0.107 | 0.293 | 0.962 | 0.086 | 0.116 | * | * | 0.786 | * | 0.246 | 0.083 | * |
| ARHGAP6 | 0.263 | 0.291 | 0.547 | 0.395 | 0.384 | 0.104 | * | * | * | * | * | * |
| C19orf33 | 0.276 | 0.479 | 0.685 | 0.819 | 0.163 | 0.411 | 0.116 | 0.089 | * | 0.470 | 0.860 | * |
| GPC3 | 0.088 | 0.143 | 0.162 | 0.728 | 0.188 | 0.543 | 0.912 | 0.794 | * | 0.187 | * | 0.195 |
| CSPG4 | 0.446 | 0.674 | 0.111 | 0.063 | 0.516 | 0.535 | 0.078 | * | * | 0.995 | 0.730 | * |
| FAM129A | 0.273 | 0.339 | 0.660 | 0.714 | 0.450 | 0.963 | * | 0.596 | 0.059 | 0.310 | 0.735 | 0.578 |
| AOX1 | * | * | 0.089 | 0.202 | 0.431 | 0.834 | 0.161 | * | * | 0.212 | 0.582 | * |
| CUX2 | * | 0.015 | 0.061 | 0.076 | 0.621 | * | 0.052 | * | * | 0.524 | * | * |
| CD109 | 0.074 | 0.129 | 0.311 | 0.688 | 0.379 | * | 0.094 | 0.861 | * | 0.195 | 0.172 | * |
| RNF150 | 0.600 | 0.930 | 0.077 | 0.898 | 0.467 | * | * | * | * | 0.759 | * | * |
| ETV4 | 0.959 | 0.813 | 0.424 | 0.510 | 0.597 | 0.335 | 0.157 | 0.092 | * | * | * | 0.116 |
| APOE | 0.579 | 0.313 | 0.083 | 0.531 | 0.289 | 0.745 | * | 0.447 | * | 0.411 | 0.462 | 0.317 |
| RNASE2 | 0.088 | 0.118 | 0.816 | 0.291 | 0.210 | 0.054 | * | 0.095 | * | 0.351 | 0.484 | * |
| GLDC | * | * | 0.291 | 0.097 | 0.777 | 0.800 | * | * | * | 0.996 | * | * |
| FRAS1 | 0.416 | 0.544 | 0.740 | 0.091 | 0.672 | 0.157 | 0.745 | * | 0.475 | 0.782 | * | 0.102 |
| CRABP1 | 0.242 | 0.258 | 0.067 | 0.997 | 0.895 | 0.062 | 0.000 | 0.390 | * | 0.984 | 0.456 | 0.001 |
| CRYAB | 0.215 | 0.387 | 0.366 | 0.644 | 0.642 | 0.512 | 0.717 | 0.771 | * | 0.982 | 0.254 | 0.007 |
| SCUBE3 | 0.780 | 0.410 | 0.109 | 0.399 | 0.246 | * | * | * | * | 0.254 | * | * |
| POPDC3 | 0.061 | 0.090 | 0.588 | 0.279 | 0.196 | 0.056 | 0.719 | 0.972 | * | 0.423 | 0.670 | 0.323 |
| BEX1 | 0.445 | 0.816 | 0.072 | 0.664 | 0.930 | * | * | * | * | 0.613 | * | * |

*P < 0.05

| Gene | OS. | RFS. | Age | Gender | Weight | positive rate | extracorporeal | pathologic | histological | neoplasm | TNM | TNM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Time | Time | | | | of lymph node metastasis | extension | stage | type | depth | M | N |
| SMAD9 | 0.371 | 0.241 | 0.805 | 0.726 | 0.755 | * | * | * | * | 0.320 | * | * |
| METTL7B | 0.075 | * | 0.848 | 0.546 | 0.671 | * | * | 0.172 | * | * | 0.302 | * |
| ZFPM2 | 0.489 | 0.823 | 0.596 | 0.341 | 0.349 | 0.840 | 0.531 | 0.704 | 0.276 | 0.234 | 0.695 | * |
| MEOX2 | 0.086 | 0.134 | 0.722 | 0.621 | 0.165 | 0.915 | 0.366 | 0.536 | * | * | 0.149 | 0.338 |
| ZNF883 | 0.208 | 0.340 | 0.077 | 0.697 | 0.968 | 0.541 | 0.650 | 0.870 | * | 0.479 | 0.647 | 0.502 |
| PLCH1 | 0.040 | 0.153 | 0.249 | 0.175 | 0.461 | 0.056 | * | * | * | 0.093 | * | * |
| ADH1B | 0.307 | 0.481 | 0.062 | 0.275 | 0.802 | 0.681 | 0.380 | 0.113 | 0.617 | * | 0.051 | 0.054 |
| FHDC1 | * | * | 0.423 | 0.142 | 0.641 | * | * | 0.097 | * | 0.318 | * | * |
| PDE5A | * | * | 0.361 | 0.156 | 0.899 | * | * | * | * | 0.141 | 0.070 | * |
| HEY2 | 0.765 | 0.721 | 0.233 | 0.771 | 0.274 | 0.077 | 0.264 | 0.054 | * | 0.082 | 0.113 | * |
| LRRK2 | * | 0.528 | 0.236 | 0.595 | 0.238 | 0.949 | 0.605 | 0.410 | * | 0.476 | 0.437 | * |
| NUP62CL | 0.109 | 0.090 | 0.092 | 0.909 | 0.912 | * | * | 0.113 | * | 0.773 | * | * |
| RYR2 | * | 0.065 | 0.108 | 0.358 | 0.639 | * | * | * | * | 0.413 | 0.050 | * |
| NMU | 0.578 | 0.868 | 0.150 | 0.134 | 0.432 | * | * | * | * | 0.854 | 0.896 | * |
| CSGALNACT1 | 0.228 | 0.641 | 0.530 | 0.746 | 0.940 | 0.160 | * | 0.122 | * | * | * | * |
| DEPDC1B | 0.292 | 0.534 | 0.238 | 0.967 | 0.167 | 0.718 | 0.229 | 0.126 | * | 0.243 | 0.147 | 0.372 |
| IER5L | * | 0.106 | 0.504 | 0.955 | 0.577 | * | 0.068 | 0.098 | * | 0.195 | * | * |
| CCDC85A | 0.674 | 0.563 | 0.448 | 0.639 | 0.844 | 0.592 | * | 0.089 | * | 0.705 | 0.944 | 0.370 |
| ZMAT3 | 0.255 | 0.508 | 0.660 | 0.963 | 0.716 | 0.763 | 0.404 | 0.907 | 0.356 | 0.173 | 0.555 | 0.972 |
| SLC4A4 | * | * | 0.091 | 0.500 | 0.482 | * | * | * | * | 0.157 | * | * |
| CLUL1 | 0.051 | 0.082 | 0.224 | 0.600 | 0.698 | * | * | * | * | 0.394 | 0.225 | * |
| CHI3L1 | 0.070 | 0.115 | 0.729 | 0.565 | 0.269 | * | * | * | * | 0.571 | 0.764 | * |
| FHL1 | * | 0.067 | 0.302 | 0.789 | 0.506 | * | * | * | * | 0.690 | * | * |
| LMOD1 | 0.069 | 0.239 | 0.144 | 0.368 | 0.629 | * | 0.068 | 0.242 | * | * | 0.140 | * |
| FBLN5 | 0.742 | 0.768 | 0.313 | 0.071 | 0.602 | 0.205 | 0.244 | 0.132 | 0.231 | * | * | * |
| SFN | * | * | 0.853 | 0.153 | 0.333 | 0.243 | * | * | * | 0.250 | 0.942 | * |
| MAMLD1 | * | * | 0.114 | 0.169 | 0.293 | * | * | * | * | 0.201 | 0.459 | * |
| GOLT1A | 0.587 | 0.504 | 0.653 | 0.532 | 0.143 | 0.143 | 0.054 | * | * | * | 0.534 | * |
| TMEM173 | 0.289 | 0.317 | 0.618 | 0.143 | 0.168 | * | * | * | * | 0.752 | * | * |
| GLRB | 0.512 | 0.740 | 0.683 | 0.205 | 0.389 | 0.142 | * | * | * | 0.115 | 0.291 | 0.186 |
| C15orf48 | 0.506 | 0.721 | 0.589 | 0.323 | 0.191 | 0.054 | * | * | * | 0.975 | 0.247 | * |
| ULBP2 | 0.101 | * | 0.705 | 0.287 | 0.110 | 0.237 | * | 0.097 | * | 0.154 | 0.645 | 0.077 |
| CITED2 | 0.701 | 0.744 | 0.187 | 0.227 | 0.654 | 0.566 | 0.054 | 0.421 | * | 0.301 | 0.527 | * |
| SPOCK1 | 0.470 | 0.580 | 0.933 | 0.147 | 0.387 | 0.777 | * | * | 0.097 | 0.268 | 0.409 | 0.485 |
| CORO2A | 0.143 | 0.185 | 0.066 | 0.103 | 0.613 | * | * | * | * | * | 0.067 | * |
| ZDHHC11 | * | * | 0.617 | 0.421 | 0.472 | * | * | * | * | 0.986 | 0.258 | * |
| SYNE1 | * | * | 0.840 | 0.591 | 0.562 | * | * | * | * | * | 0.371 | * |
| AIF1L | 0.747 | 0.470 | 0.921 | 0.820 | 0.533 | 0.161 | * | * | * | 0.712 | 0.164 | * |
| FGFR2 | 0.387 | 0.302 | 0.860 | 0.916 | 0.855 | 0.169 | * | * | * | * | 0.465 | * |

*P < 0.05

| Gene | OS. | RFS. | Age | Gender | Weight | positive rate | extracorporeal | pathologic | histological | neoplasm | TNM | TNM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Time | Time | | | | of lymph node metastasis | extension | stage | type | depth | M | N |
| HS6ST2 | * | 0.063 | 0.135 | 0.110 | 0.645 | 0.152 | 0.169 | 0.730 | 0.097 | 0.141 | 0.661 | 0.151 |
| MFAP4 | 0.587 | 0.385 | 0.064 | 0.547 | 0.687 | 0.423 | * | 0.099 | * | 0.513 | 0.401 | 0.059 |
| KLK10 | * | * | 0.197 | 0.637 | 0.242 | * | 0.105 | * | * | 0.095 | 0.405 | * |
| TDRD9 | * | * | 0.068 | 0.538 | 0.153 | * | 0.085 | * | * | 0.478 | * | * |
| MDK | * | * | 0.588 | 0.234 | 0.931 | 0.606 | 0.080 | * | * | 0.162 | * | * |
| TIMP1 | 0.139 | 0.276 | 0.472 | 0.987 | 0.409 | * | * | 0.139 | * | 0.270 | 0.124 | * |
| IGF2BP3 | 0.336 | 0.459 | 0.752 | 0.778 | 0.382 | 0.591 | 0.615 | 0.762 | * | 0.677 | 0.689 | 0.061 |
| TCEAL2 | 0.073 | 0.189 | 0.209 | 0.687 | 0.908 | * | * | 0.343 | * | 0.198 | 0.502 | * |
| CDH3 | 0.083 | 0.140 | 0.256 | 0.691 | 0.808 | 0.082 | * | 0.160 | * | * | 0.073 | * |
| TGFBI | 0.285 | 0.309 | 0.181 | 0.868 | 0.235 | 0.681 | * | 0.079 | * | 0.464 | 0.521 | 0.265 |
| SLC34A2 | * | * | 0.737 | 0.364 | 0.487 | * | * | * | * | 0.524 | 0.210 | * |
| GALNT7 | 0.086 | 0.150 | 0.096 | 0.896 | 0.601 | * | * | * | * | 0.092 | 0.275 | * |
| SGCD | 0.126 | 0.153 | 0.844 | 0.857 | 0.343 | 0.276 | 0.282 | 0.500 | * | 0.784 | 0.143 | 0.243 |
| GGCT | * | * | 0.983 | 0.738 | 0.447 | * | * | * | * | * | * | * |
| FXYD5 | 0.226 | 0.179 | 0.837 | 0.267 | 0.477 | 0.090 | 0.089 | * | * | 0.566 | 0.052 | * |
| DDX25 | 0.565 | 0.749 | 0.963 | 0.915 | 0.697 | 0.270 | * | * | 0.303 | 0.861 | 0.099 | 0.053 |
| WISP1 | 0.147 | 0.173 | 0.076 | 0.794 | 0.093 | * | * | * | * | 0.943 | 0.936 | * |
| CCNA1 | 0.053 | 0.084 | 0.856 | 0.988 | 0.360 | * | * | * | * | 0.806 | 0.839 | * |
| TNC | * | * | 0.328 | 0.178 | 0.109 | * | * | 0.075 | * | 0.451 | 0.151 | * |
| DPT | 0.835 | 0.810 | 0.888 | 0.804 | 0.732 | 0.752 | * | 0.505 | 0.258 | * | * | 0.113 |
| DYNLRB2 | 0.227 | 0.393 | 0.159 | 0.847 | 0.485 | * | * | * | * | 0.713 | 0.194 | * |
| SOD3 | * | * | 0.145 | 0.675 | 0.720 | * | * | * | * | 0.653 | 0.345 | * |
| RUNX2 | * | 0.079 | 0.264 | 0.342 | 0.493 | * | * | * | * | 0.301 | * | * |
| P4HA2 | 0.551 | 0.703 | 0.860 | 0.449 | 0.216 | 0.171 | * | * | * | 0.149 | 0.333 | * |
| SHANK2 | * | 0.080 | 0.819 | 0.280 | 0.743 | * | * | * | * | 0.177 | 0.085 | * |
| PRTG | 0.074 | 0.142 | 0.436 | 0.504 | 0.553 | * | * | * | * | 0.448 | 0.391 | * |
| CDC42EP5 | * | * | 0.624 | 0.814 | 0.259 | * | * | * | * | 0.078 | * | * |
| PROS1 | 0.010 | * | 0.734 | 0.629 | 0.972 | * | * | * | * | 0.173 | 0.067 | * |
| ESM1 | 0.793 | 0.583 | 0.848 | 0.112 | 0.400 | 0.241 | 0.131 | * | * | 0.161 | * | * |
| ALDH1A1 | 0.086 | 0.111 | 0.944 | 0.127 | 0.217 | * | * | * | * | 0.842 | 0.570 | * |
| IGFBP6 | 0.675 | 0.589 | 0.367 | 0.382 | 0.995 | * | * | * | * | 0.206 | 0.819 | * |
| LTBP1 | 0.847 | 0.930 | 0.648 | 0.283 | 0.168 | * | * | 0.066 | * | 0.611 | * | * |
| SLC22A4 | * | * | 0.360 | 0.713 | 0.310 | * | * | 0.155 | * | 0.104 | * | * |
| CAMK2N1 | 0.103 | 0.258 | 0.718 | 0.734 | 0.695 | * | * | * | * | 0.137 | * | * |
| HRH1 | * | * | 0.293 | 0.860 | 0.858 | * | * | * | * | 0.354 | 0.190 | * |
| PHLDA2 | 0.330 | 0.374 | 0.642 | 0.696 | 0.701 | * | * | * | * | 0.314 | 0.583 | * |
| CA4 | 0.182 | 0.351 | 0.592 | 0.890 | 0.177 | 0.104 | * | * | * | 0.560 | 0.392 | * |
| NFE2L3 | 0.125 | 0.126 | 0.511 | 0.988 | 0.975 | * | * | * | * | 0.772 | 0.540 | * |
| NT5E | 0.305 | 0.399 | 0.544 | 0.548 | 0.383 | * | * | 0.083 | * | 0.927 | 0.124 | * |

*P < 0.05

| Gene | OS. | RFS. | Age | Gender | Weight | positive rate | extracorporeal | pathologic | histological | neoplasm | TNM | TNM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Time | Time | | | | of lymph node metastasis | extension | stage | type | depth | M | N |
| PDLIM4 | 0.006 | * | 0.855 | 0.528 | 0.261 | * | * | * | * | 0.146 | 0.171 | * |
| SLC35F2 | 0.107 | 0.156 | 0.510 | 0.672 | 0.525 | * | * | * | * | 0.058 | * | * |
| CA12 | 0.111 | 0.148 | 0.232 | 0.791 | 0.796 | 0.117 | 0.173 | 0.139 | * | 0.412 | 0.206 | * |
| TMEM92 | 0.062 | 0.095 | 0.156 | 0.671 | 0.165 | * | * | * | * | 0.514 | 0.295 | * |
| F2RL2 | 0.063 | 0.114 | 0.056 | 0.122 | 0.280 | * | * | * | * | 0.886 | 0.258 | * |
| KDELR3 | 0.702 | 0.906 | 0.348 | 0.449 | 0.276 | * | * | * | * | 0.601 | 0.340 | * |
| S100A11 | 0.118 | 0.205 | 0.398 | 0.302 | 0.090 | * | * | * | * | 0.303 | 0.115 | * |
| CLDN10 | 0.195 | 0.284 | 0.065 | 0.638 | 0.236 | * | * | * | * | 0.449 | 0.313 | * |
| LGALS3 | 0.080 | 0.201 | 0.938 | 0.083 | 0.238 | * | * | * | * | 0.193 | * | * |
| ABI3BP | 0.221 | 0.135 | 0.843 | 0.242 | 0.941 | 0.330 | 0.233 | 0.763 | 0.630 | 0.234 | 0.747 | 0.250 |
| FAM20A | 0.746 | 0.871 | 0.703 | 0.075 | 0.794 | * | * | * | * | 0.258 | 0.338 | * |
| MXRA8 | * | * | 0.956 | 0.338 | 0.473 | * | * | * | * | 0.194 | 0.097 | * |
| CDC42EP3 | 0.281 | 0.428 | 0.903 | 0.876 | 0.485 | * | * | * | * | 0.883 | 0.312 | * |
| MRC2 | 0.067 | * | 0.312 | 0.764 | 0.856 | * | * | 0.137 | * | * | * | * |
| ABCC3 | 0.272 | 0.310 | 0.971 | 0.561 | 0.967 | * | 0.210 | 0.466 | * | 0.102 | 0.056 | * |
| KCNN4 | * | 0.091 | 0.488 | 0.843 | 0.104 | * | * | * | * | 0.326 | 0.474 | * |
| B4GALT6 | 0.436 | 0.567 | 0.338 | 0.080 | 0.528 | 0.974 | 0.156 | 0.600 | 0.339 | 0.127 | * | 0.220 |
| ROR1 | * | 0.070 | 0.620 | 0.106 | 0.105 | * | * | * | * | 0.395 | 0.557 | * |
| LAMB3 | * | * | 0.829 | 0.805 | 0.784 | * | * | * | * | 0.143 | 0.321 | * |
| SCN3A | * | 0.078 | 0.926 | 0.128 | 0.216 | 0.248 | * | 0.067 | * | 0.301 | 0.414 | 0.095 |
| PLAUR | * | * | 0.865 | 0.203 | 0.109 | * | * | * | * | 0.785 | 0.488 | * |
| TGM2 | 0.174 | 0.147 | 0.238 | 1.000 | 0.533 | * | * | * | * | 0.970 | 0.371 | * |
| NID1 | * | * | 0.897 | 0.122 | 0.135 | * | * | 0.170 | * | 0.811 | 0.329 | * |
| CDH11 | 0.093 | 0.119 | 0.290 | 0.708 | 0.089 | * | * | * | * | 0.979 | 0.346 | * |
| IL17RD | * | * | 0.487 | 0.091 | 0.448 | * | * | 0.313 | * | * | * | * |
| IGFBP3 | 0.634 | 0.714 | 0.755 | 0.438 | 0.764 | 0.488 | 0.384 | 0.723 | 0.280 | 0.357 | 0.085 | 0.431 |
| COMP | * | * | 0.130 | 0.791 | 0.420 | 0.131 | 0.076 | 0.161 | * | 0.561 | 0.904 | 0.275 |
| TNFRSF21 | * | * | 0.115 | 0.698 | 0.657 | * | * | 0.112 | * | 0.219 | 0.570 | * |
| FN1 | 0.127 | 0.332 | 0.212 | 0.696 | 0.084 | * | * | * | * | 0.684 | 0.327 | * |
| SPP1 | 0.315 | 0.531 | 0.214 | 0.316 | 0.093 | 0.517 | 0.050 | * | 0.069 | 0.937 | 0.717 | 0.272 |
| MET | * | 0.082 | 0.093 | 0.595 | 0.636 | * | * | * | * | 0.219 | 0.107 | * |
| AHNAK2 | * | * | 0.302 | 0.640 | 0.364 | * | * | * | * | 0.310 | * | * |
| MICAL2 | 0.117 | 0.112 | 0.487 | 0.946 | 0.264 | * | * | * | * | 0.483 | 0.321 | * |
| COL8A2 | 0.093 | 0.208 | 0.196 | 0.742 | 0.916 | * | * | * | * | 0.130 | 0.253 | * |
| NPTX2 | 0.601 | 0.441 | 0.103 | 0.078 | 0.411 | 0.152 | 0.962 | 0.991 | 0.799 | 0.786 | 0.553 | 0.159 |
| LOX | 0.089 | 0.222 | 0.108 | 0.315 | 0.263 | * | * | * | * | 0.584 | 0.297 | * |
| RUNX1 | * | 0.081 | 0.743 | 0.782 | 0.073 | * | * | * | * | 0.417 | * | * |

*P < 0.05

There were 129 significant DEGs related to TNM stagings, such as *RELN*, *IL17RD*, *AHNAK2*, and *RUNX1*. Among them, these DEGs were clearly associated with a positive rate of lymph node metastasis, extracorporeal extension, and neoplasm depth, which were *IL17RD*, *GGCT*, and *MRC2*. In comparison, the number of

DEGs related to histological type and pathologic stage was 131 (such as *COMP*, *FN1*, *LOX*, and *COL8A2*) and 83, respectively. 78 DEGs including *CLUL1*, *CHI3L1*, *FHL1*, and *SFN* were associated with TNM staging, histological type, and pathologic stage. There were only 40 DEGs significantly related to prognosis status as they must involve one of the above three clinical conditions, such as *LAMB3*, *MET*, *AHNAK2*, *COMP*, and *PLAUR*. In addition, 17 DEGs, such as *CYP1B1*, *GPC3*, and *AOX1*, affected the primary site of the tumor. Finally, 42 DEGs have the potential influence residual tumors, such as *PHLDA2*, *NFE2L3*, and *S100A2*.

GO enrichment analysis result demonstrated that DEGs were mainly concentrated in extracellular space, regulation of cell growth, cell adhesion, collagen binding, negative regulation of protein, extracellular matrix disassembly, negative regulation of keratinocyte proliferation, and signal transduction. KEGG enrichment analysis result verified that the DEGs were mainly concentrated in ECM-receptor interaction, Focal adhesion, and PI3K-Akt signaling pathway.

# The Methylation Levels Of 153 Degs

DNA methylation occurring on CpG sites maintained proper regulation of gene expression and acted as a silencing mechanism during tumorigenesis. Kaplan-Meier survival analysis was conducted on 153 DEGs CpG sites to verify the significance of gene expression in predicting the prognosis of THC. There were 5912 CpG sites with significant differences. Among them, only six CpG sites were up-regulated: cg00508334, cg05000748, cg09799983, cg20170028, cg22861369, and cg23358564. They were located on *IGF2BP3*, *RUNX1*, *CYP1B1*, *SHANK2*, and *PDLIM4*. Compared with normal samples,these genes were significantly hypermethylated and low expressed in thyroid cancer samples. They might be tumor suppressors in thyroid cancer. In contrast, 68 CpG sites were down-regulated, including cg03625911, cg19310148, cg22700686, and cg07168232. The genes involved included *CHI3L1*, *NFE2L3*, *S100A2*, *LAMB3, etc*. They were hypomethylated,high expressed and caused the possibly carcinogenic risk of thyroid. Additionally, three genes (*ANHAK2*, *IL17RD*, and *PLAUR*) were significantly negatively correlated with the prognosis (Fig. 6).

## Discussion

The process of reprogramming somatic cells into iPSCs was similar to that of cancer pathogenesis, namely the ability to obtain self-renewal and poor differentiation. Therefore, in this study, the expression profiles of both were integrated to explore the pathogenesis of THC and new biomarkers of THC.

According to the analysis of GEO and TCGA database genes, 237 DEGs overlapped in tumor and reprogramming groups. After clinical prognosis analysis, enrichment analysis was performed. The result indicated that the proteins of these genes were concentrated in the extracellular space, mesenchymal cell differentiation, proteinaceous extracellular matrix, ECM-receptor interaction, and PI3K-Akt signaling pathway. These functional annotations suggested that the proteinaceous extracellular matrix and EMT may play important roles in these biological processes, confirming that EMT is one of the key steps of cancer cell metastasis [14]. Compared with epithelial cancer cells, mesenchymal cancer cells reduced the expression of cell adhesion proteins (such as E-cadherin and G-catenin) and activated mesenchymal markers (such as N-cadherin and vimentin) to increase cells ability to migrate and invade; consequently, changes in cell adhesion proteins were used as markers of the EMT process [15]. EMT also conferred cancer cells with stem cell-like properties, resistance to targeted therapies, and the ability to evade immune surveillance by the host [16]. The tumor microenvironment was a dynamic and constantly changing environment, including cancer cells, fibroblasts, immune cells, endothelial cells, and ECM; among them, cancer-associated fibroblasts or tumor-associated fibroblasts (CAF or TAF) were major cellular components in the environment [17]. Additionally, more and more evidence verified that CAF played a vital role in tumor development and metastasis through the synthesis of ECM protein [18]. Moreover, tumor cells and some immune cells in the tumor microenvironment may secrete TGF-β to induce the EMT process, leading to the spread of tumor cells and cancer metastasis [19]. The PI3K/AKT pathway regulated cell signaling and many other processes (such as proliferation, angiogenesis, migration, and survival) [20]. ECM protein can activate the PI3K/AKT signaling pathway through integrin, resulting in continuous tumor growth. Furthermore, the PI3K/AKT signal can cause the down-regulation of adhesion protein expression by inducing the EMT process and transform those epithelioid cancer cells into mesenchymal cancer cells, bringing about cancer cell metastasis [21]. Therefore, a better understanding of proteinaceous extracellular matrix and EMT will contribute to discovering new prognostic and diagnostic indicators, as well as treatment opportunities.

After methylation analysis, 4 DEGs (*NFE2L3, LAMB3, S100A2*, and *CHI3L1*) involving more down-regulated CpG sites were considered to have a relationship with THC progression significantly. *CHI3L1* was located on chromosome 1q32.1 and encoded a protein YKL-40 related to inflammation and cancer. The up-regulation of this gene was closely related to the metastasis of various cancers (such a thyroid cancer, liver cancer, bladder cancer, and lung cancer); it could be used as a biomarker for independent prognosis of cancer [22]. Human *NFE2L3* was a transcription factor of the basic region of the cap 'n' on chromosome 7p15-p14, and the overexpression of this gene independently enhanced the growth and invasion of thyroid cancer cells [23, 24]. *S100A2*, on chromosome 1q21, a member of the S100 protein family, bound to calcium ions and regulated the transmission of cellular signals [25, 26]. It can affect the progression of carcinogenesis by regulating cell migration and proliferative activity [25, 26]. Moreover, it was demonstrated to play a different role in tumor suppressor or tumor promoter in cancer. For example, *S100A2* was down-regulated in breast cancer, lung cancer, and gastric cancer while it was up-regulated in ovarian cancer, head and neck squamous cell carcinoma, and pancreatic cancer [25, 26]. The data suggested that *S100A2* can promote tumor cell invasion and metastasis by activated transforming growth factor-β (TGF-β) and inducing effective EMT [27]. However, the mechanism of the in thyroid cancer remained unclear, even though *LAMB3* was positively involved in the metastasis of multiple tumor types, including colon, pancreas, lung, cervix, stomach, and prostate [28]. Our results indicated that *LAMB3* expression was up-regulated in THC, promoting cell migration or invasion by affecting the PI3K-Akt signaling pathway and EMT-related proteins. Furthermore, clinicopathological features and prognostic analysis illustrated that *LAMB3* can regulate lymphatic metastasis in thyroid cancer.

## Conclusion

In this study, we revealed four genes that are closely related to the development of thyroid cancer from GEO and TCGA. This would contribute to predicting the prognosis of patients with thyroid cancer.

## Declarations

**Conflict of Interest:** None of the authors had a commercial association or a competing financial interest, either actual or potential, which might create a conflict of interest in connection with the submitted article.

**Author Contributions:** HZ, PL, JL and LD conceived and designed the overall study. YJ,SL, YB, and XZ retrieved the date. HZ, JL and LD analyzed the data. HZ and LD wrote the manuscript. ML and XH revised the manuscript. All authors read and approved the final manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China [Grant No. 81572528] and Outstanding Youth Fund of Heilongjiang Province [Grant No. JC2018024] to Xiaoyi Huang.

**Acknowledgments:** The authors are grateful to all subjects in the study.

**Data Availability Statement:** The raw sequencing data can be found in the GEO database and TCGA database.

## References

1. Antonelli A, La Motta C. Novel therapeutic clues in thyroid carcinomas: The role of targeting cancer stem cells [J]. *Med Res Rev*. 37, 1299-317 (2017).
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. *CA Cancer J Clin*. 68, 394-424 (2018).
3. Colin IM, Denef JF, Lengele B, Many MC, Gerard AC. Recent insights into the cell biology of thyroid angiofollicular units [J]. *Endocr Rev*. 34, 209-38 (2013).
4. Saiselet M, et al. miRNA expression and function in thyroid carcinomas: a comparative and critical analysis and a model for other cancers [J]. *Oncotarget*. 7, 52475-92 (2016).
5. Schlumberger M, Chevillard S, Ory K, Dupuy C, Le Guen B, de Vathaire F. Thyroid cancer following exposure to ionising radiation [J]. *Cancer Radiother*. 15, 394-9 (2011).
6. Ferrari SM, et al. Lenvatinib in the Therapy of Aggressive Thyroid Cancer: State of the Art and New Perspectives with Patents Recently Applied [J]. *Recent Pat Anticancer Drug Discov*. 13, 201-8 (2018).
7. Makki F, et al. Serum biomarkers of papillary thyroid cancer. [J]. *J Otolaryngol Head Neck Surg*. 42, 16 (2013).
8. Huang Y, et al. Gene expression in papillary thyroid carcinoma reveals highly consistent profiles [J]. *Proc Natl Acad Sci U S A*. 98, 15044-9 (2001).
9. Yamanaka S. A fresh look at iPS cells [J]. *Cell*. 137, 13-7 (2009).
10. Krizhanovsky V, Lowe S. Stem cells The promises and perils of p53 [J]. *Nature*. 460, 1085-6 (2009).
11. Varum S, et al. Energy metabolism in human pluripotent stem cells and their differentiated counterparts [J]. *PLoS One*. 6, e20914 (2011).
12. von Roemeling CA, et al. Aberrant lipid metabolism in anaplastic thyroid carcinoma reveals stearoyl CoA desaturase 1 as a novel therapeutic target [J]. *J Clin Endocrinol Metab*. 100, E697-709 (2015).
13. Pita JM, Banito A, Cavaco BM, Leite V. Gene expression profiling associated with the progression to poorly differentiated thyroid carcinomas [J]. *Br J Cancer*. 101, 1782-91 (2009).
14. Zheng X, et al. Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer [J]. *Nature*. 527, 525-30 (2015).
15. Li C, et al. Excess PLAC8 promotes an unconventional ERK2-dependent EMT in colon cancer [J]. *J Clin Invest*. 124, 2172-87 (2014).
16. Chockley PJ, Chen J, Chen G, Beer DG, Standiford TJ, Keshamouni VG. Epithelial-mesenchymal transition leads to NK cell–mediated metastasis-specific immunosurveillance in lung cancer [J]. *Journal of Clinical Investigation*. 128, 1384-96 (2018).
17. Carstens JL, Lovisa S, Kalluri R. Microenvironment-dependent cues trigger miRNA-regulated feedback loop to facilitate the EMT/MET switch [J]. *J Clin Invest*. 124, 1458-60 (2014).
18. Jang I, Beningo KA. Integrins, CAFs and Mechanical Forces in the Progression of Cancer [J]. *Cancers (Basel)*. 11, E721 (2019).
19. David CJ, et al. TGF-beta Tumor Suppression through a Lethal EMT [J]. *Cell*. 164, 1015-30 (2016).
20. Han B, et al. Coptisine-induced apoptosis in human colon cancer cells (HCT-116) is mediated by PI3K/Akt and mitochondrial-associated apoptotic pathway [J]. *Phytomedicine*. 48, 152-60 (2018).
21. Wu X, Cai J, Zuo Z, Li J. Collagen facilitates the colorectal cancer stemness and metastasis through an integrin/PI3K/AKT/Snail signaling pathway [J]. *Biomed Pharmacother*. 114, 108708 (2019).
22. Luo D, et al. CHI3L1 overexpression is associated with metastasis and is an indicator of poor prognosis in papillary thyroid carcinoma [J]. *Cancer Biomark*. 18, 273-84 (2017).
23. Chevillard G, Blank V. NFE2L3 (NRF3): the Cinderella of the Cap'n'Collar transcription factors [J]. *Cell Mol Life Sci*. 68, 3337-48 (2011).
24. Yu MM, Feng YH, Zheng L, Zhang J, Luo GH. Short hairpin RNA-mediated knockdown of nuclear factor erythroid 2-like 3 exhibits tumor-suppressing effects in hepatocellular carcinoma cells [J]. *World J Gastroenterol*. 25, 1210-23 (2019).
25. Bai Y, Li L, Li J, Lu X. Prognostic values of S100 family members in ovarian cancer patients. [J]. *BMC Cancer*. 18, 1256 (2018).

26. Woo T, et al. Up-Regulation of S100A11 in Lung Adenocarcinoma - Its Potential Relationship with Cancer Progression [J]. *PLoS One*. 10, e0142642 (2015).
27. Naz S, Bashir M, Ranganathan P, Bodapati P, Santosh V, Kondaiah P. Protumorigenic actions of S100A2 involve regulation of PI3/Akt signaling and functional interaction with Smad3 [J]. *Carcinogenesis*. 35, 14-23 (2014).
28. Jung S, et al. LAMB3 mediates metastatic tumor behavior in papillary thyroid cancer by regulating c-MET/Akt signals. [J]. *Sci Rep*. 8, 2718 (2018).

## Figures

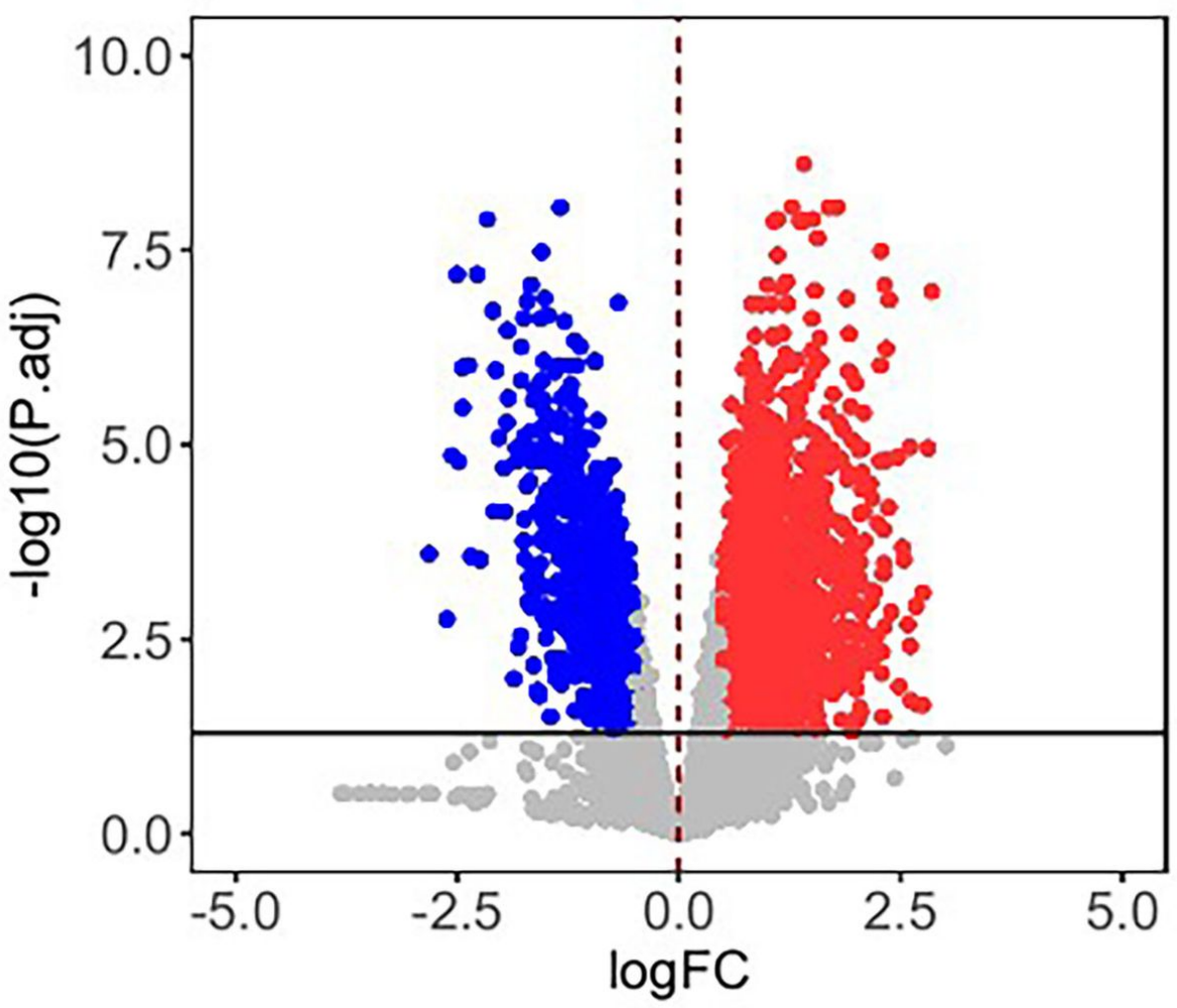


**Figure 1**

Vocalno plot indicated DEGS of THC and ctr from GEO Database. Red up-regulation; blue down-regulation

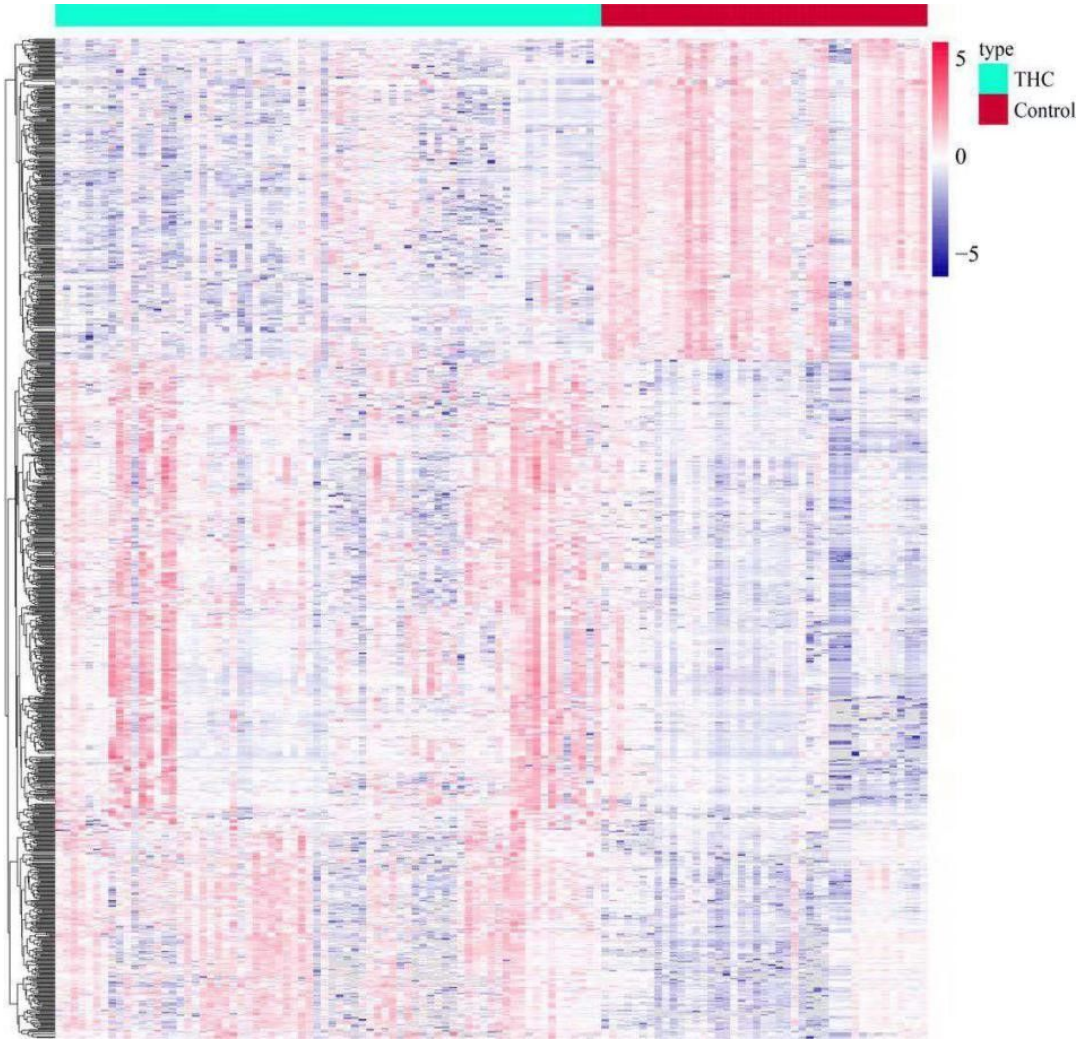


**Figure 2**

Representative heat map of DEGS of THC and ctr from GEO Database.

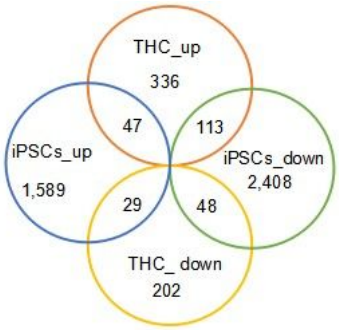


**Figure 3**

Venn diagram for THC and iPSCs common DEGS.

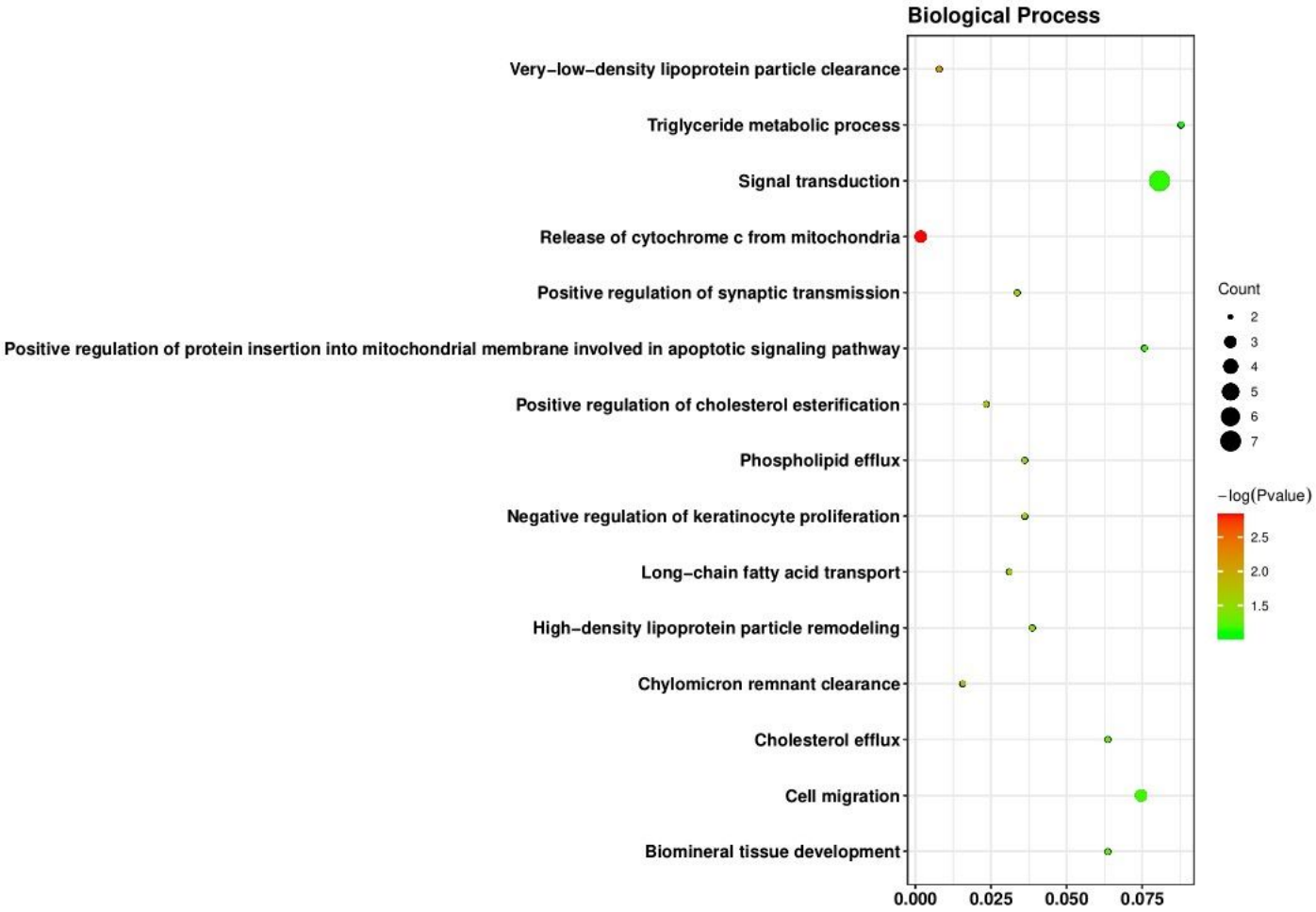


Figure 4

Bubble chart showed the Gene ontology analysis of THC and iPSCs jointly up-regulate DEGS.

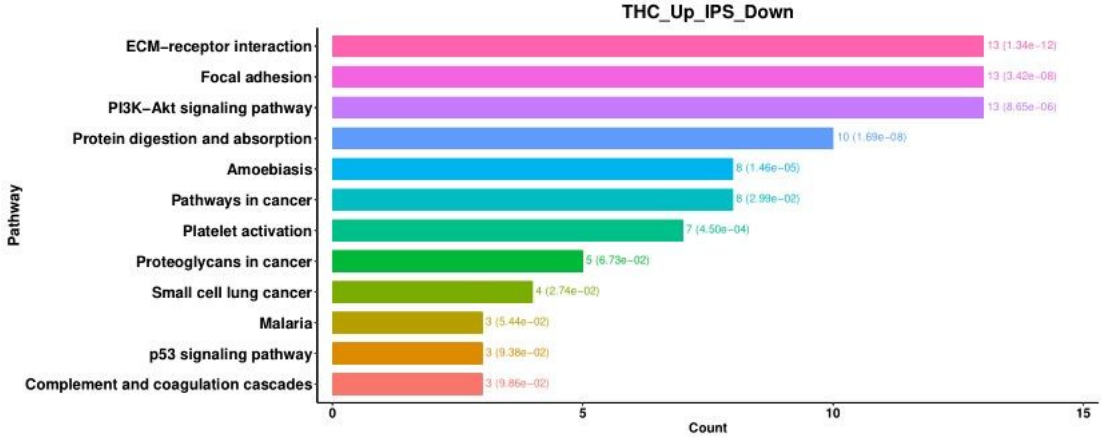


Figure 5

Histogram showed the KEGG pathyway analysis of THC up-regulated but iPSCs down-regulated DEGS.

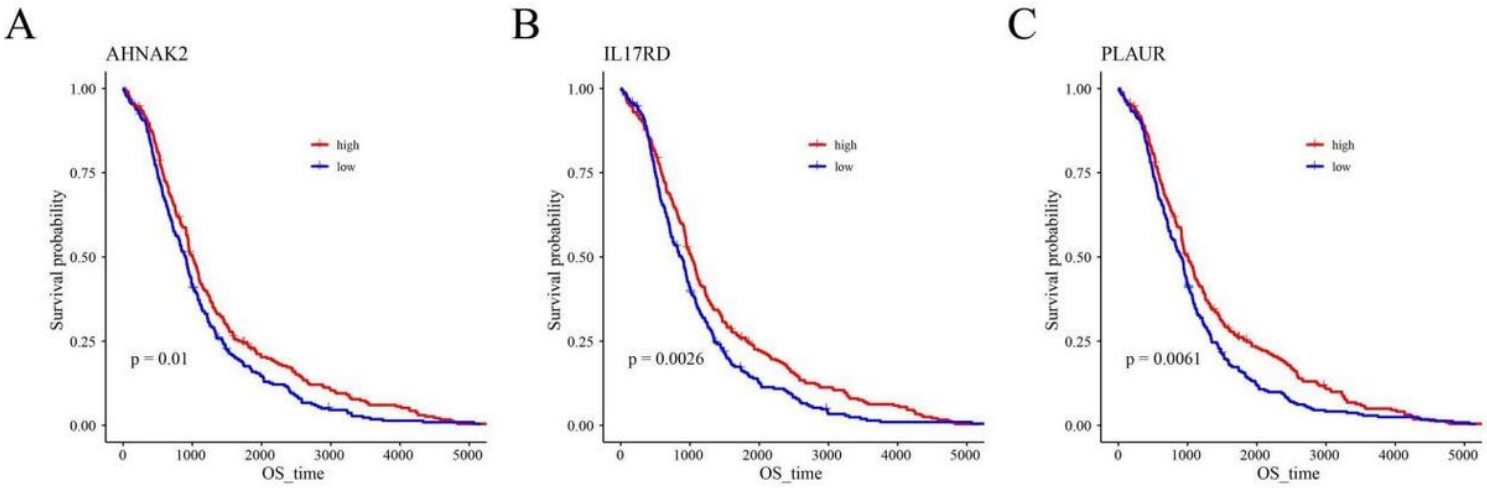


**Figure 6**

AHNAK2 survival curve, IL17RD survival curve, and PLAUR survival curve. Red high; blue low