Identifying hub genes of papillary thyroid carcinoma in the TCGA and GEO database using bioinformatics analysis

Ying Wan [Corresp., Equal first author, 1] , Xiaolian Zhang [Equal first author, 2] , Huilin Leng [3] , Weihua Yin [4] , Wenxing Zeng [1] , Congling Zhang [1]

[1] Department of Inspection, People's Hospital of Yichun City, Yichun, China

[2] Department of Blood Transfusion, People's Hospital of Yichun City, Yichun, China

[3] Department of Neurology, People's Hospital of Yichun City, Yichun, China

[4] Department of Oncology, People's Hospital of Yichun City, Yichun, China

Corresponding Author: Ying Wan
Email address: emilywan01@tom.com

Background. Thyroid carcinoma (THCA) is a common endocrine malignant tumor. Papillary carcinoma with low degree of malignancy and good prognosis is the most common. It can occur at any age, but it is more common in young adults. Although the mortality rate is decreased due to early diagnosis, the survival rate varies depending on the type of tumor. Therefore, the purpose of this study is to identify hub biomarkers and novel therapeutic targets for THCA.

Methods. The GSE3467, GSE3678, GSE33630 and GSE53157 were obtained from the GEO database, including 100 thyroid tumors and 64 normal tissues to obtain the intersection of differentially expressed genes, and a protein-protein interaction network was constructed to obtain the HUB gene. The corresponding overall survival information from The Cancer Genome Atlas Project-THCA was then included in this research. The signature mechanism was studied by analyzing the gene ontology and the Kyoto Encyclopedia gene and genome database.

Results. In this research, we identified eight candidate genes (FN1, CCND1, CDH2, CXCL12, MET, IRS1, DCN and FMOD) from the network. Also, expression verification and survival analysis of these candidate genes based on TCGA database indicate the robustness of the above results. Finally, our hospital samples validated the expression levels of these genes.

Conclusion. The research identified eight mRNA including four up-regulated and four down-regulated serve as signature, which could be a potential prognostic marker of THCA.

Identifying hub genes of papillary thyroid carcinoma in the

TCGA and GEO database using bioinformatics analysis

Ying Wan[1*], Xiaolian Zhang[2*], Huilin Leng[3], Weihua, Yin[4], Wenxing Zeng[1], Congling Zhang[1]

1 Department of Inspection, People's Hospital of Yichun City, Yichun, 336000 Jiangxi, China

2 Department of Blood Transfusion, People's Hospital of Yichun City, Yichun, 336000 Jiangxi, China

3 Department of Neurology, People's Hospital of Yichun City, Yichun, 336000 Jiangxi, China

4 Department of Oncology, People's Hospital of Yichun City, Yichun, 336000 Jiangxi, China

* These authors contributed equally to this work.

Corresponding Author:

Ying Wan[1],

No. 1061 Jinxiu road, Yiyang new district, Yichun, 336000 Jiangxi, China

Email: emilywan01@tom.com

Abstract

Background. Thyroid carcinoma (THCA) is a common endocrine malignant tumor. Papillary carcinoma with low degree of malignancy and good prognosis is the most common. It can occur at any age, but it is more common in young adults. Although the mortality rate is decreased due to early diagnosis, the survival rate varies depending on the type of tumor. Therefore, the purpose of this study is to identify hub biomarkers and novel therapeutic targets for THCA.

Methods. The GSE3467, GSE3678, GSE33630 and GSE53157 were obtained from the GEO database, including 100 thyroid tumors and 64 normal tissues to obtain the intersection of differentially expressed genes, and a protein-protein interaction network was constructed to obtain the HUB gene. The corresponding overall survival information from The Cancer Genome Atlas Project-THCA was then included in this research. The signature mechanism was studied by analyzing the gene ontology and the Kyoto Encyclopedia gene and genome database.

Results. In this research, we identified eight candidate genes (FN1, CCND1, CDH2, CXCL12, MET, IRS1, DCN and FMOD) from the network. Also, expression verification and survival analysis of these candidate genes based on TCGA database indicate the robustness of the above results. Finally, our hospital samples validated the expression levels of these genes.

Conclusion. The research identified eight mRNA including four up-regulated and four down-regulated serve as signature, which could be a potential prognostic marker of THCA.

Keywords: thyroid carcinoma, TCGA, GEO, signature, prognosis, biomarker

Introduction

38

39   Thyroid carcinoma (THCA) is that originates from thyroid tissue and has the potential for

40   distal metastasis (Siegel et al. 2019). It is worth noting that the incidence of women is two to four

41   times higher than that of men (Carling & Udelsman 2014). THCA is mainly divided into four

42   categories, including papillary carcinoma (85%), follicular carcinoma (10-15%), medullary

43   carcinoma (5-10%) and undifferentiated THCA (< 5%). As a result of early diagnosis, the

44   mortality rate of THCA is stable or declining, although the incidence of THCA continues to rise

45   (Wang et al. 2019b). Notwithstanding, the high incidence of papillary thyroid cancer (PTC) is

46   depended on many uncontrollable factors, such as specific genetic changes of THCA and

47   traumatic invasive examination has different degrees of damage to the patient's thyroid.

48   Therefore, it is very important to identify hub biomarker and non-invasive methods for PTC.

49   With the development of RNA sequencing and microarray technology, the study of

50   differentially expressed gene (DEG) between cancer and normal tissues has improved our

51   understanding of the molecular mechanism of PTC. It also could help us identify important

52   biomarkers. Previous studies were based on one dataset with limited sample size, and it is not the

53   best way to get DEGs from a single dataset (Liang & Sun 2018; Zhao et al. 2016). Above defects

54   may lead to a deviation in the final result.

55   In this research, we conducted a multi-step analysis through comprehensive bioinformatics

56   analysis to identify hub genes in THCA. We obtained the expression profiles of GSE3467,

57   GSE3678, GSE33630 and GSE53157 with PTC from the GEO database, which included a total

58   of 100 THCA and 64 normal samples. After the intersection, we got a total of 179 DEGs in four

59   datasets. Then, we carried out above genes through bioinformatics analysis, which provided

60   reliable and novel biomarkers for THCA, which will be helpful for further clinical application in

61   the diagnosis, prognosis and targeted treatment of PTC.

62

Methods

63

64   Datasets selecting and DEGs identification

65   We downloaded the four gene expression datasets of PTC from GEO database, including the

66   following criterias: (a) PTC, (b) datasets including tumor and normal tissues, (c) the organism is

67   *Homo sapiens*, (d) Sample size exceeding 10 samples. GSE3467 (n = 18), GSE3678 (n = 14),

68   GSE33630 (n = 60) and GSE53157 (n = 27) were among depended on the GPL570 platform.

69   The *limma* package was used to identify the DEGs in each GEO datasets in R (Version 3.7.1)

70   (Ritchie et al. 2015). The *P*-value is determined by the false discovery rate. The standard for

71   DEG is that the false discovery rate is less than 0.05, and the criteria of the groups were $|\log_2 \text{FC}$

72   (fold change)$| \geqslant 1$.

73   KEGG and GO enrichment analyses of DEGs

74   To explore biological information and obtain more comprehensive gene and protein functions,

75   we used *clusterProfile* packages for GO, which including cellular component (CC) biological

76 process (BP) and molecular function (MF), and KEGG analysis (Yu et al. 2012).

77 PPI network construction

78   The online database is used to predict the PPI network, which is used to retrieve the

79 interacting genes (STRING; http://string-db.org) (Szklarczyk et al. 2017). In this research, the

80 PPI network of genes was established with STRING database, and the interaction with high

81 confidence > 0.4 was statistically significant. Next, we use Cytoscape (Version 3.6.1), a free

82 bioinformatics software, to visualise and analyse the molecular interaction networks.

83 TCGA database validation and survival analysis

84   In order to verify the results of the hub genes, we also obtained the gene expression profiles

85 from The Cancer Genome Atlas (TCGA) (https://cancergenome.nih.gov/) database to validate

86 them, including 502 tumor and 58 non-tumor samples. SurvExpress, a free online website, uses

87 log-rank tests to draw Kaplan-Meier plots that contain the "Survival" package of R software.

88 Clinical patient samples

89   A total of 39 patients with THCA diagnosed in People's Hospital of Yichun City from October

90 2018 to August 2019 were included in this study. This study was approved by the Ethical

91 Committee of People's Hospital of Yichun City (PHYC-2018-10-8-1). All patients had not

92 received any radiotherapy or chemotherapy before sampling and signed a written informed

93 consent form. Immediately after surgical resection, the tumor and adjacent non-tumor tissues

94 were transferred to liquid nitrogen and stored in the refrigerator at -80 ℃ until RNA was

95 extracted.

96 RNA extraction and qRT-PCR analysis

97   Total RNA was extracted by TRIzol reagent (Invitrogen, CA, USA) from tissues. After

98 dissolution, NanoDrop Lite spectrophotometer to evaluate the concentration and purity of total

99 RNA. Then, the total RNA reverse transcription synthesis of cDNA and used to conduct qRT-

100 PCR assays for genes. GAPDH was used as internal control, and PCR experiments were repeated

101 three times. Fold change ($2^{-\Delta\Delta CT}$) was represented relative gene expression levels. Table 1 was

102 listed all the primer sequences and small interfernce sequences, and divergent primer was

103 designed for each gene.

104

Results

106 Intersection of DEGs in four GEO datasets

107   As shown in Figure. 1, we conducted a multi-step analysis to study hub DEGs and its

108 important biological functions and prognosis in THCA through comprehensive bioinformatics

109 methods. To begin with, we chose and obtained a total of four gene expression profiles of THCA

110 from GEO datasets (GSE3467, GSE3678, GSE33630 and GSE53157). In this research, 100 cases

111 of thyroid tumors and 64 cases of normal tissues were obtained.

112   The DEG distribution in every GEO datasets demonstrated in the form of volcano plots

113 (Figure. 2a-d), including 851 DEGs in GSE3467, 785 DEGs in GSE3678, 1404 DEGs in

114 GSE33630, and 858 DEGs in GSE53157. The intersection of 179 DEGs in four GEO datasets

115   was displayed in Figure. 2e.

116   Functional annotation of DEGs

117   GO enrichment analysis was carried out to research the biological functions of all the DEGs.

118   The BP classification of GO analysis demonstrated that the DEGs obviously enriched the

119   skeletal system development, muscle cell differentiation and regulation of chromosome

120   organization. For CC, these DEGs are rich in chromatin, adherens junction and nuclear

121   chromosome part. In addition they are significantly rich in transcription coregulator activity,

122   proximal promoter sequence-specific DNA binding and RNA polymerase II proximal from

123   promoter in the MF category (Figure. 3a).

124   The enrichment analysis of KEGG pathway displayed that DEGs were obviously abundant in

125   metabolic pathways, TGF-beta signaling pathway, central carbon metabolism in cancer and

126   proteoglycans in cancer (Figure. 3b). Pathways of PTC were also contained, including metabolic

127   pathways, neuroactive ligand−receptor interaction, rap1 signaling pathway, thyroid hormone

128   signaling pathway.

129   Construction of PPI network

130   In order to determine the interaction of these DEGs in THCA, we used the STRING database

131   to build a PPI network in Cytoscape software (Figure. 4a), which included 127 nodes and 199

132   edges. Next, we use the plug-in cytoHubba in Cytoscape to screen out the first eight hub genes.

133   They are FN1, CCND1, CDH2, CXCL12, MET, IRS1, DCN and FMOD using the MCC method

134   in Figure 4b.

135   Prognostic gene screening

136   In order to verify above results of all the DEGs and candidate genes obtained in other datasets,

137   we obtained the gene expression profiles of all 179 DEGs in THCA from the TCGA database.

138   Additionally, a distinctly different expression pattern between the tumor and normal samples

139   were found (Figure. 5a-h). We explored the association between gene expression and survival,

140   and found that only DCN was significantly correlated with the overall survival (OS) of THCA

141   samples trough the GEPIA database (log-rank p = 0.01892, Figure 5i).

142   Validation of differentially expressed levels for hub genes by qRT-PCR

143   The levels of eight hub genes expression (FN1, CCND1, CDH2, CXCL12, MET, IRS1, DCN

144   and FMOD) in PTC and para-cancerous non-tumor tissues were detected by qRT-PCR. As

145   shown in Figure. 6, FN1 (fibronectin 1), CCND1 (Cyclin D1), CDH2 (cadherin 2), MET (MET

146   proto-oncogene, receptor tyrosine kinase), were significantly upregulated in THCA tissues (n =

147   87) compared with normal samples. Additionally, IRS1 (insulin receptor substrate 1), DCN

148   (decorin), FMOD (fibromodulin), and CXCL12 (chemokine C-X-C motif ligand 12) were down-

149   regulated in tumor tissues. The above results confirmed the accuracy of our analysis and showed

150   that these eight key genes can be considered as a potential signature or biomarker of THCA.

151

152   **Discussion**

153   In the present study, we screened 179 robust DEGs in PTC's four GEO datasets. These DEGs

154   were obviously enriched in a variety of cancer-related functions and pathways by GO and KEGG

155 pathway enrichment analysis. Use the STRING database to build the PPI network, and carry on
156 the hub genes. We finally screened eight central genes from the whole network, including FN1,
157 CCND1, CDH2, CXCL12, MET, IRS1, DCN and FMOD, which genes are involved in the
158 regulation of cyclin gene or cadherin gene, and could participate in the development of PTC. The
159 expression verification and prognosis analysis of these candidate genes depended on TCGA
160 database showed the robustness and reliability of the above results.

161 Several studies had been published about DEGs in human malignant tumors, such as THCA.
162 Liu *et al.* identified 359 DEGs between tumor and normal thyroid tissues from TCGA database
163 (Liu et al. 2019). Among them, five DEGs (LPAR5, NMU, FN1, NPY1R, and CXCL12) were
164 screened with higher degrees and LPAR5 was associated with OS. Compared with our research,
165 among the top eight hub genes (FN1, CCND1, CDH2, CXCL12, MET, IRS1, DCN and FMOD),
166 we found that only DCN existed significance of OS in this study.

167 Compared with previously published researches on THCA, our research fully integrates
168 multiple databases, datasets, and large sample sizes, including tumors and normal tissues. The
169 results we got were robust and powerful, which could have non-invasive examination for patients
170 with PTC.

171 Through palpation, doctors can find thyroid nodules, which are hard in texture and not smooth
172 on the surface. At the same time, if the nodules do not move with the trachea when swallowing,
173 and the enlarged and hard lymph nodes in the neck can be touched, thyroid cancer should be
174 taken into account. If the tumor is in a small stage, ultrasound has become a very good screening
175 tool for thyroid space-occupying lesions, which is widely used in thyroid physical examination
176 and clinical differential diagnosis. However, the ultrasonic features of benign and malignant
177 nodules overlap, and ultrasound doctors' interpretations and suggestions on these features are
178 quite different in clinical practice, which requires further fine needle puncture and genetic testing
179 at the same time. We obtained eight hub genes in the network, which can be used for unique
180 expression levels in PTC, and we can even diagnose PTC in early diagnosis. Some of which have
181 been proved to be associated with PTC, depended on PPI network and candidate genes
182 evaluation. Low fibronectin 1 (FN1) expression in tumorous tissues was an independent worse
183 prognostic factor for progression-free survival in sporadic medullary thyroid carcinoma (Zhan et
184 al. 2018). Additionally, Ouchi *et al*. identified FN1 as a fusion partner of ALK by 5'RACE, FN1-
185 ALK, resulted in ALK overexpression in the inflammatory myofibroblastic tumors, and this
186 finding should facilitate development of novel therapeutics (Ouchi et al. 2015). Also, FN1 could
187 encode fibronectin, which was associated with various cancers precession, including cervical
188 cancer (Wang et al. 2019a), gastric cancer (Jiang et al. 2019), colorectal cancer (Wu et al. 2016).
189 Cyclin D1 (CCND1) functions as regulators of CDK kinases, and forms a complex with and
190 functions as a regulatory subunit of CDK4 or CDK6, whose activity is required for cell cycle
191 G1/S transition. It was proven as a DEG and used as a diagnostic and predictive biomarker in
192 thyroid carcinoma and may be functionally involved in the development and progression of the
193 disease (Jeon et al. 2018). In addition, down regulation of CCND1 may be an important
194 mechanism by which everolimus increases the therapeutic window of paclitaxel in cervical
195 cancers (Yilmaz et al. 2016). The biological role of cadherin 2 (CDH2) was also mostly reported

196 in glioma with prognostic significance (Chen et al. 2018b). Previous studies showed that
197 chemokine (C-X-C motif) ligand 12 (CXCL12) had reported in different human tumors.
198 CXCL12 could activate CXCR4 as well as multiple downstream multiple tumorigenic signaling
199 pathways, promoting expression of various oncogenes. Activation of the CXCL12-CXCR4
200 signaling axis promotes epithelial-mesenchymal transition (EMT) (Sleightholm et al. 2017; Yang
201 et al. 2019). The biological roles of CXCL12 also were founded in THCA, in which
202 CXCR4/CXCR7/CXCL12 axis offer new valuable insight into the oncogenesis of metastatic
203 follicular thyroid carcinoma (Werner et al. 2018). The amplification of MET proto-oncogene,
204 receptor tyrosine kinase (MET) is confirmed as a resistance factor in gastric cancer cells (Ebert
205 et al. 2019). And, MET may serve a role in regulating PD-L1 expression in hepatocellular
206 carcinoma (Chun & Hong 2019) Meanwhile, MET is also involved in high-risk metastasis
207 progression in thyroid cancer (Garcia et al. 2019). Insulin receptor substrate 1 (IRS1) can be
208 phosphorylated to activate AKT kinase, which promotes the development of lung cancer
209 (Gorgisen et al. 2019). The low expression of decorin (DCN) has been demonstrated to
210 contribute to be as a potential biomarker in colon cancer (Li et al. 2017). DCN was found
211 underexpressed in thyroid tumors (Arnaldi et al. 2005). Qian *et al.* had shown that the methylated
212 +58CpG in DCN 5'-UTR involved in reduced the mRNA expression of DCN (Qian et al. 2014).
213 In our research, Fibromodulin (FMOD) acted as an important mediator in VEGF expression and
214 angiogenesis induced by GDNF in human glioblastoma (Chen et al. 2018a).
215   The expression of DCN gene in cancer tissues was lower than that in para-cancerous tissues.
216 However, the lower the expression in cancer tissue, the better the prognosis was. This DCN is
217 still an oncogene, but its pathway may be suppressed as a whole, with low expression or low
218 levels of other regulation. It can still have an effect on tumor growth, but it is not the main
219 driving gene. This can also happen if the prognosis of the tumor is good.
220   However, functional experiments should be carried out to verify our results. In addition,
221 further research on a larger sample size will be required.
222

Conclusion

224   We found eight hub genes including four up-regulated and four down-regulated in PTC, and
225 disclosed that DCN gene is an important gene related to prognosis through TCGA database to
226 verify the accuracy of our finding. This hub genes may be considered as a new and potential
227 signature or biomarker of PTC, which will further aid in contributing to the diagnosis, prognosis
228 and clinical application of therapeutic targets for PTC.
229
230

231 Arnaldi LA, Borra RC, Maciel RM, and Cerutti JM. 2005. Gene expression profiles reveal that DCN, DIO1, and DIO2
232   are underexpressed in benign and malignant thyroid tumors. *Thyroid* 15:210-221.
233   10.1089/thy.2005.15.210
234 Carling T, and Udelsman R. 2014. Thyroid cancer. *Annu Rev Med* 65:125-137. 10.1146/annurev-med-061512-
235   105739

236   Chen M, Ba H, Lu C, Dai J, and Sun J. 2018a. Glial Cell Line-Derived Neurotrophic Factor (GDNF) Promotes
237   Angiogenesis through the Demethylation of the Fibromodulin (FMOD) Promoter in Glioblastoma. *Med Sci*
238   *Monit* 24:6137-6143. 10.12659/MSM.911669
239   Chen Q, Cai J, and Jiang C. 2018b. CDH2 expression is of prognostic significance in glioma and predicts the efficacy
240   of temozolomide therapy in patients with glioblastoma. *Oncol Lett* 15:7415-7422. 10.3892/ol.2018.8227
241   Chun HW, and Hong R. 2019. Significance of PD-L1 clones and C-MET expression in hepatocellular carcinoma.
242   *Oncol Lett* 17:5487-5498. 10.3892/ol.2019.10222
243   Ebert K, Mattes J, Kunzke T, Zwingenberger G, and Luber B. 2019. MET as resistance factor for afatinib therapy and
244   motility driver in gastric cancer cells. *PLoS One* 14:e0223225. 10.1371/journal.pone.0223225
245   Garcia C, Buffet C, El Khattabi L, Rizk-Rabin M, Perlemoine K, Ragazzon B, Bertherat J, Cormier F, and Groussin L.
246   2019. MET overexpression and activation favors invasiveness in a model of anaplastic thyroid cancer.
247   *Oncotarget* 10:2320-2334. 10.18632/oncotarget.26798
248   Gorgisen G, Hapil FZ, Yilmaz O, Cetin Z, Pehlivanoglu S, Ozbudak IH, Erdogan A, and Ozes ON. 2019. Identification of
249   novel mutations of Insulin Receptor Substrate 1 (IRS1) in tumor samples of non-small cell lung cancer
250   (NSCLC): Implications for aberrant insulin signaling in development of cancer. *Genet Mol Biol* 42:15-25.
251   10.1590/1678-4685-gmb-2017-0307
252   Jeon S, Kim Y, Jeong YM, Bae JS, and Jung CK. 2018. CCND1 Splice Variant as A Novel Diagnostic and Predictive
253   Biomarker for Thyroid Cancer. *Cancers (Basel)* 10. 10.3390/cancers10110437
254   Jiang K, Liu H, Xie D, and Xiao Q. 2019. Differentially expressed genes ASPN, COL1A1, FN1, VCAN and MUC5AC are
255   potential prognostic biomarkers for gastric cancer. *Oncol Lett* 17:3191-3202. 10.3892/ol.2019.9952
256   Li G, Li M, Liang X, Xiao Z, Zhang P, Shao M, Peng F, Chen Y, Li Y, and Chen Z. 2017. Identifying DCN and HSPD1 as
257   Potential Biomarkers in Colon Cancer Using 2D-LC-MS/MS Combined with iTRAQ Technology. *J Cancer*
258   8:479-489. 10.7150/jca.17192
259   Liang W, and Sun F. 2018. Identification of key genes of papillary thyroid cancer using integrated bioinformatics
260   analysis. *J Endocrinol Invest* 41:1237-1245. 10.1007/s40618-018-0859-3
261   Liu L, He C, Zhou Q, Wang G, Lv Z, and Liu J. 2019. Identification of key genes and pathways of thyroid cancer by
262   integrated bioinformatics analysis. *J Cell Physiol*. 10.1002/jcp.28932
263   Ouchi K, Miyachi M, Tsuma Y, Tsuchiya K, Iehara T, Konishi E, Yanagisawa A, and Hosoi H. 2015. FN1: a novel fusion
264   partner of ALK in an inflammatory myofibroblastic tumor. *Pediatr Blood Cancer* 62:909-911.
265   10.1002/pbc.25424
266   Qian Q, Shi X, Lei Z, Zhan L, Liu RY, Zhao J, Yang B, Liu Z, and Zhang HT. 2014. Methylated +58CpG site decreases
267   DCN mRNA expression and enhances TGF-beta/Smad signaling in NSCLC cells with high metastatic
268   potential. *Int J Oncol* 44:874-882. 10.3892/ijo.2014.2255
269   Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, and Smyth GK. 2015. limma powers differential expression
270   analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43:e47. 10.1093/nar/gkv007
271   Siegel RL, Miller KD, and Jemal A. 2019. Cancer statistics, 2019. *CA Cancer J Clin* 69:7-34. 10.3322/caac.21551
272   Sleightholm RL, Neilsen BK, Li J, Steele MM, Singh RK, Hollingsworth MA, and Oupicky D. 2017. Emerging roles of
273   the CXCL12/CXCR4 axis in pancreatic cancer progression and therapy. *Pharmacol Ther* 179:158-170.
274   10.1016/j.pharmthera.2017.05.012
275   Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ,
276   and von Mering C. 2017. The STRING database in 2017: quality-controlled protein-protein association

277   networks, made broadly accessible. *Nucleic Acids Res* 45:D362-D368. 10.1093/nar/gkw937

278   Wang S, Gao B, Yang H, Liu X, Wu X, and Wang W. 2019a. MicroRNA-432 is downregulated in cervical cancer and
279   directly targets FN1 to inhibit cell proliferation and invasion. *Oncol Lett* 18:1475-1482.
280   10.3892/ol.2019.10403

281   Wang Y, Huang H, Hu F, Li J, Zhang L, and Pang H. 2019b. CITED1 contributes to the progression of papillary thyroid
282   carcinoma via the Wnt/beta-catenin signaling pathway. *Onco Targets Ther* 12:6769-6777.
283   10.2147/OTT.S215025

284   Werner TA, Forster CM, Dizdar L, Verde PE, Raba K, Schott M, Knoefel WT, and Krieg A. 2018.
285   CXCR4/CXCR7/CXCL12-Axis in Follicular Thyroid Carcinoma. *J Cancer* 9:929-940. 10.7150/jca.23042

286   Wu J, Wang Y, Xu X, Cao H, Sahengbieke S, Sheng H, Huang Q, and Lai M. 2016. Transcriptional activation of FN1
287   and IL11 by HMGA2 promotes the malignant behavior of colorectal cancer. *Carcinogenesis* 37:511-521.
288   10.1093/carcin/bgw029

289   Yang P, Hu Y, and Zhou Q. 2019. The CXCL12-CXCR4 signaling axis plays a key role in cancer metastasis and is a
290   potential target for developing novel therapeutics against metastatic cancer. *Curr Med Chem*.
291   10.2174/0929867326666191113113110

292   Yilmaz A, Alp E, Onen HI, and Menevse S. 2016. Reduced BCL2 and CCND1 mRNA expression in human cervical
293   cancer HeLa cells treated with a combination of everolimus and paclitaxel. *Contemp Oncol (Pozn)* 20:28-
294   32. 10.5114/wo.2016.58498

295   Yu G, Wang LG, Han Y, and He QY. 2012. clusterProfiler: an R package for comparing biological themes among gene
296   clusters. *OMICS* 16:284-287. 10.1089/omi.2011.0118

297   Zhan S, Li J, Wang T, and Ge W. 2018. Quantitative Proteomics Analysis of Sporadic Medullary Thyroid Cancer
298   Reveals FN1 as a Potential Novel Candidate Prognostic Biomarker. *Oncologist* 23:1415-1425.
299   10.1634/theoncologist.2017-0399

300   Zhao M, Wang KJ, Tan Z, Zheng CM, Liang Z, and Zhao JQ. 2016. Identification of potential therapeutic targets for
301   papillary thyroid carcinoma by bioinformatics analysis. *Oncol Lett* 11:51-58. 10.3892/ol.2015.3829

302

303

304

Table 1 mRNA primers

1   **Table1**   mRNA PCR primer

Gene name	Primer sequence
FN1	F: CGGTGGCTGTCAGTCAAAG
	R: AAACCTCGGCTTCCTCCATAA
CCND1	F: GCTGCGAAGTGGAAACCATC
	R: CCTCCTTCTGCACACATTTGAA
CDH2	F: TCAGGCGTCTGTAGAGGCTT
	R: ATGCACATCCTTCGATAAGACTG
CXCL12	F: ATTCTCAACACTCCAAACTGTGC
	R: ACTTTAGCTTCGGGTCAATGC
MET	F: AGCAATGGGGAGTGTAAAGAGG
	R: CCCAGTCTTGTACTCAGCAAC
IRS1	F: ACAAACGCTTCTTCGTACTGC
	R: AGTCAGCCCGCTTGTTGATG
DCN	F: ATGAAGGCCACTATCATCCTCC
	R: GTCGCGGTCATCAGGAACTT
FMOD	F: GAGACCTACGAGCCTTACCC
	R: TTGAGGTTGCGATTGTCACAG

Figure 1

Figure_1

The development and verification process of the assay method is summarized and detailed in the following sections. Gene Expression Omnibus; DEG, differentially expressed genes; PPI, protein–protein interaction; TCGA, The Cancer Genome Atlas

GEO Detasets selection

Four datasets with gene expression profiles

DEGs identification in each dataset

Based on limma package in R

Robust Rank aggregation method

179 DEGs were selected

Funtional enrichment analysis

Pathway enrichment analysis

Based on STRING database

PPI network construction and modules analysis

Hub genes of indentification

Biological characteristics of hub gene

Based on TCGA database

TCGA database valiadation analysis

Survival analysis

Figure 2

Figure_2

Filter and identify DEGs in various GEO datasets. (a - d) The volcanic map of the DEGs distribution was shown in GSE3467 (a), GSE3678 (b), GSE33630 (c) and GSE53157 (d). (e) Intersection of four datasets. DEG, differentially expressed gene
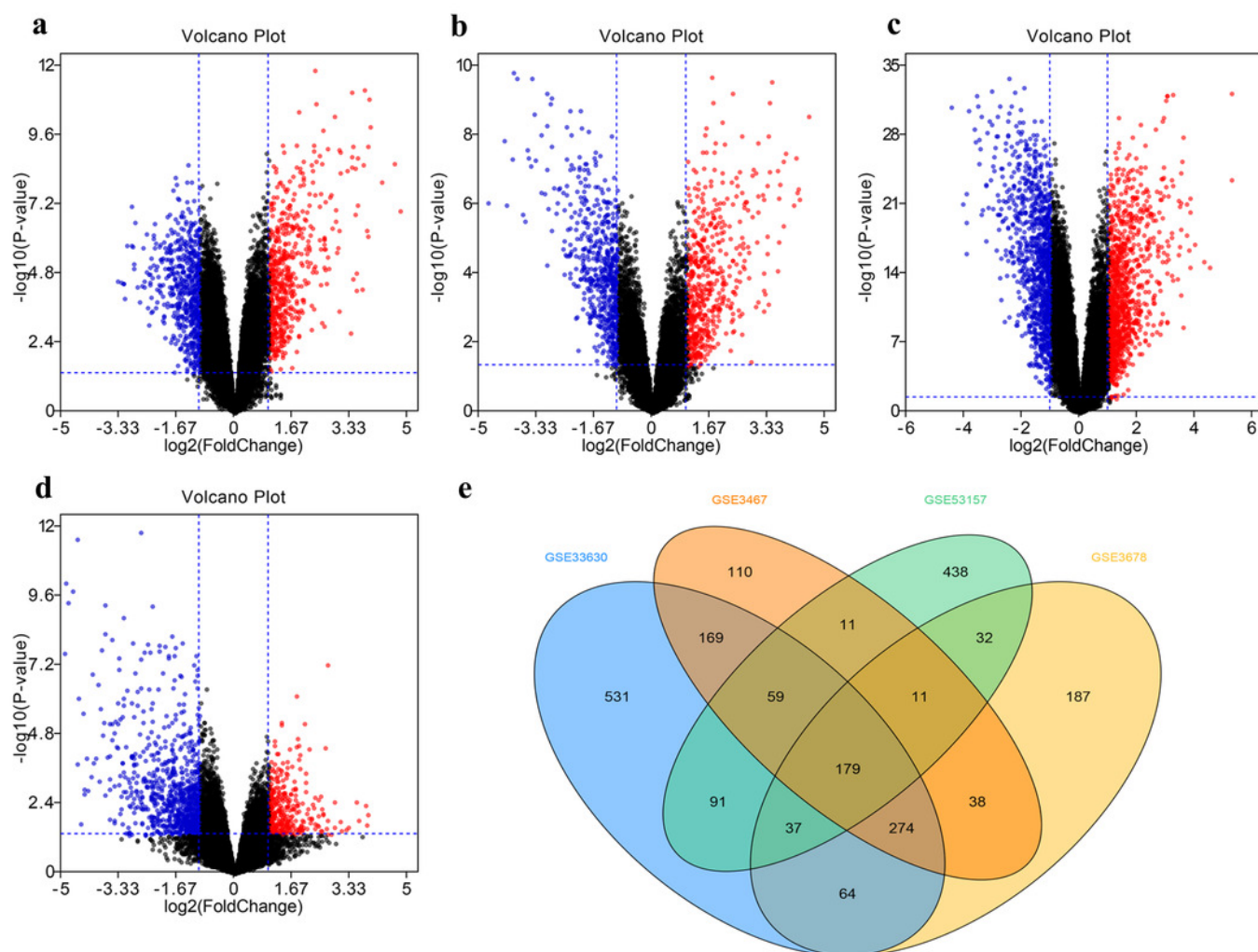
Figure 3

Figure_3

Gene Ontology and Kyoto Encyclopaedia of Genes and Genomes pathway enrichment analyses. (a) Genes were assigned to GO categories and the terms were summarized into three main GO categories. (b) Enrichment Analysis of abnormal regulation DEGs in Bubble Diagram. DEG, differentially expressed gene.

a

The Most Enriched GO Terms

b
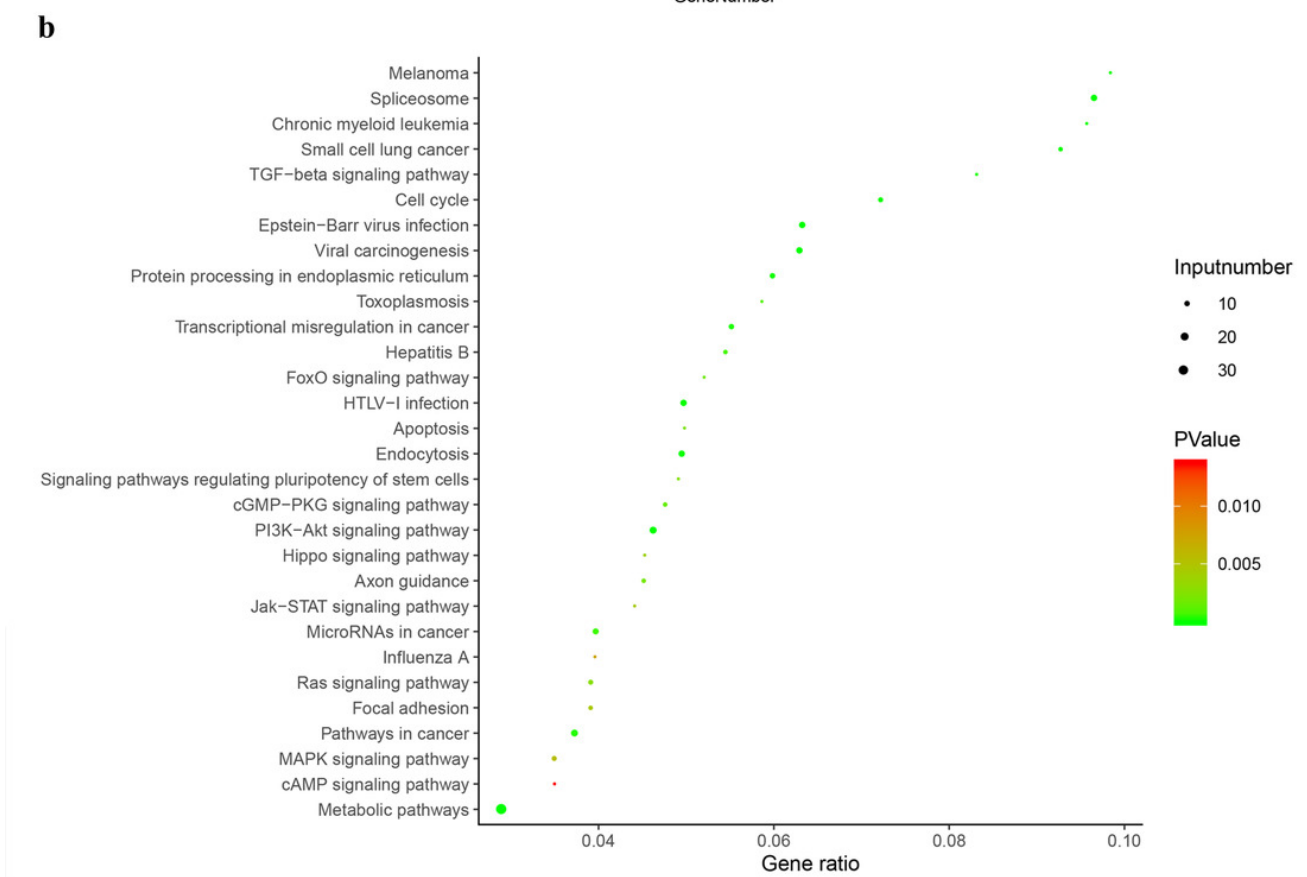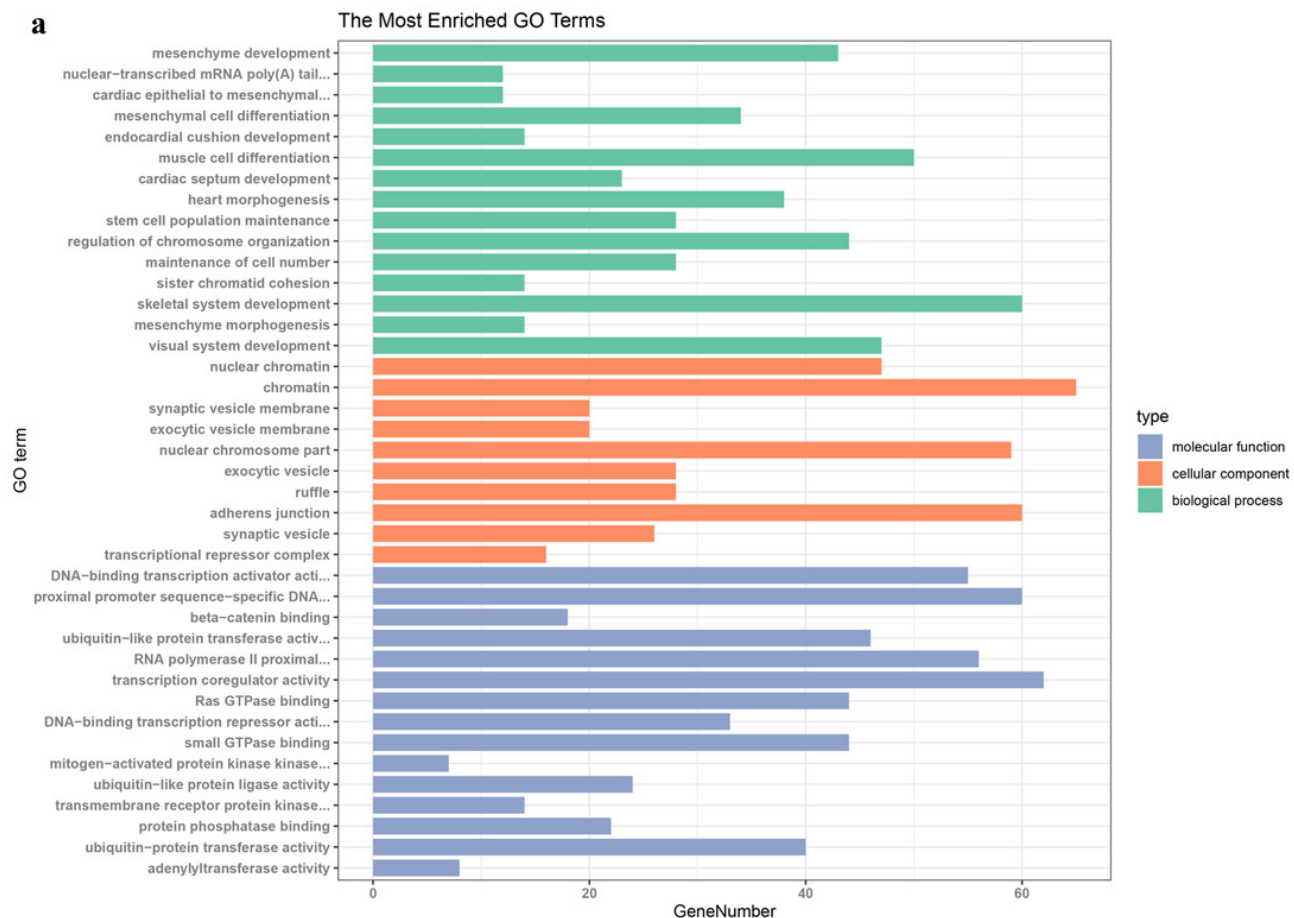
Figure 4

Figure_4

PPI of DEGs. (a) The red nodes represent the hub genes. (b) A redder node means a higher degree.
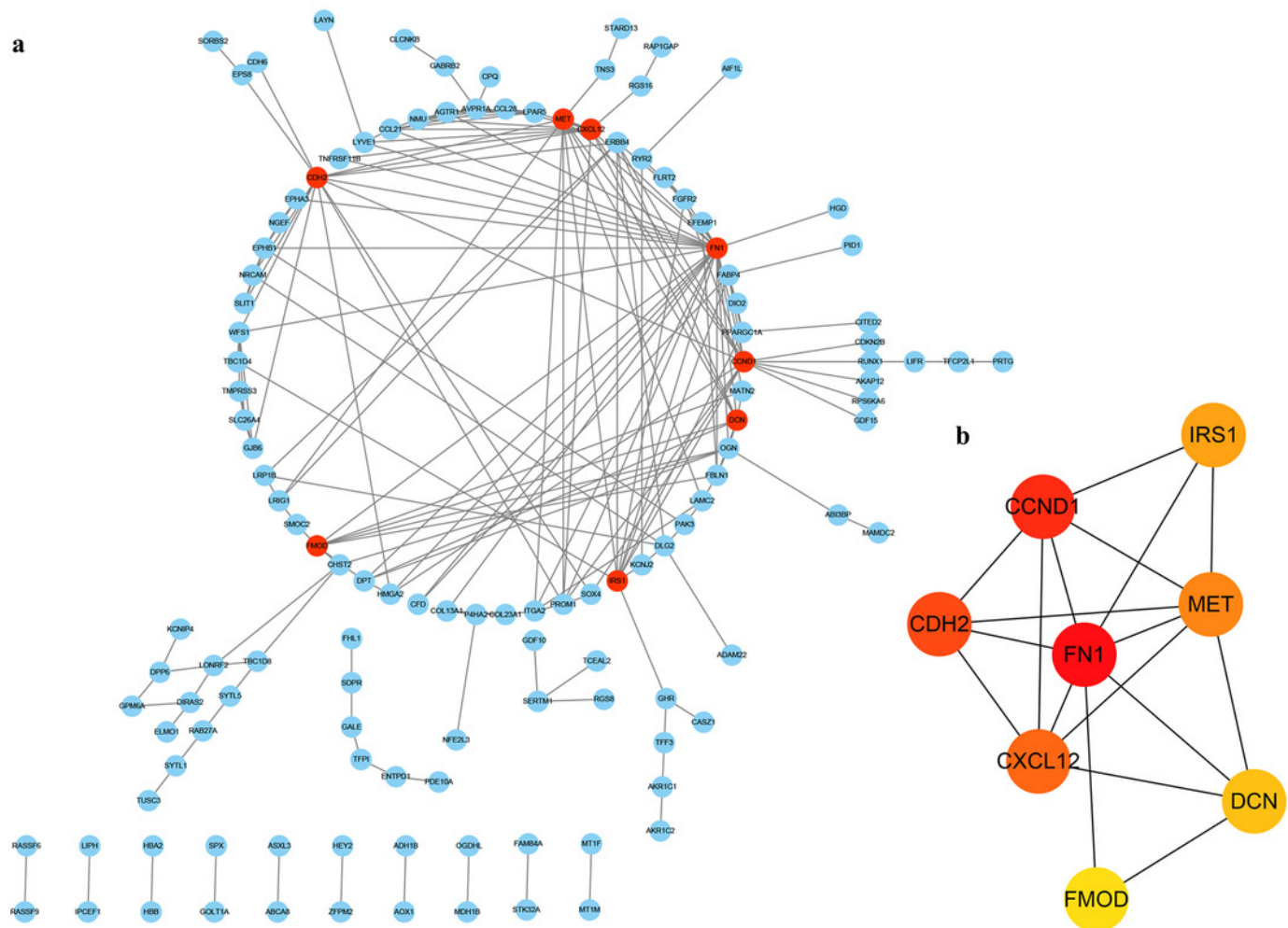
Figure 5

Figure_5

The expression of genes on TCGA. (a) FN1. (b) CCND1 (c) CDH2 (d) CXCL12 (e) MET (f) IRS1 (g) DCN (h) FMOD (i) The OS of DCN in TCGA database. Red represent cancer tissues, grey represent normal tissues. * means P < 0.05. TCGA, The Cancer Genome Atlas; OS, overall survival.
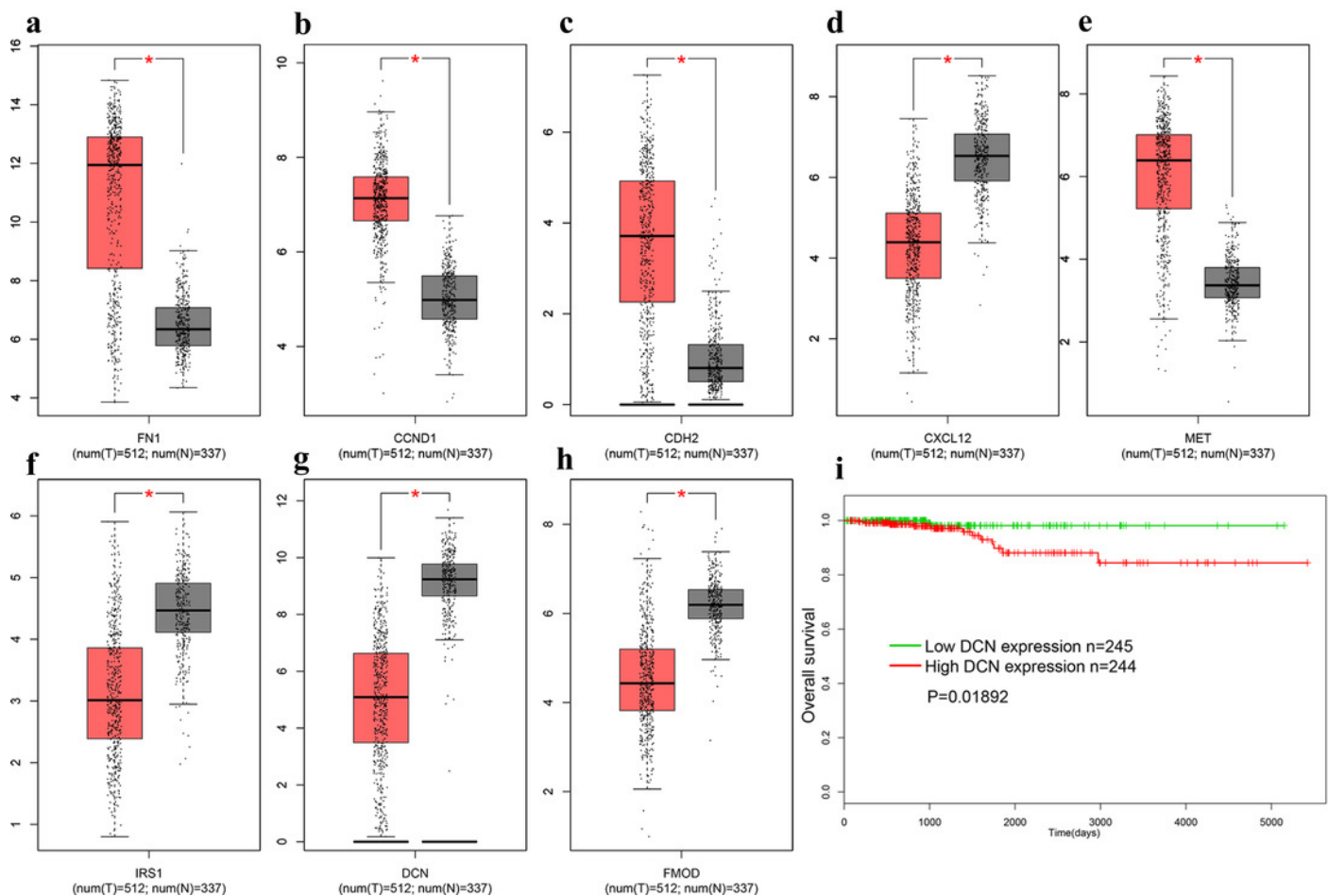
Figure 6

Figure_6

Relative expression levels of hub genes detected by qRT-PCR. *p < 0.05, **p < 0.01 and ***p < 0.001. (a) FN1. (b) CCND1 (c) CDH2 (d) CXCL12 (e) MET (f) IRS1 (g) DCN (h) FMOD. qRT-PCR, quantitative real time polymerase chain reaction