Identification of potential gene signatures associated with osteosarcoma by integrated bioinformatics analysis

Yutao Jia [1], **Yang Liu** [1], **Zhihua Han** [Corresp., 2], **Rong Tian** [1]

[1] Department of Spine Surgery, Tianjin Union Medical Center, Tianjin, China

[2] Department of Anesthesiology, Tianjin Union Medical Center, Tianjin, China

Corresponding Author: Zhihua Han
Email address: hanzhihua1996@163.com

Background. Osteosarcoma (OS) is the most primary malignant bone cancer in children and adolescents with a high mortality rate. This work aims to screen novel potential gene signatures associated with OS by integrated microarray analysis of Gene Expression Omnibus (GEO) database. Material and Methods. The OS microarray datasets were searched and downloaded from GEO database to identify differentially expressed genes (DEGs) between OS and normal samples. Afterwards, the functional enrichment analysis, protein-protein interaction (PPI) network analysis and transcription factor (TF)-target gene regulatory network were applied to uncover the biological function of DEGs. Finally, two published OS datasets (GSE39262 and GSE126209) were obtained from GEO database for evaluating the expression level and diagnostic values of key genes. Results. Totally, 1059 DEGs (569 up-regulated DEGs and 490 down-regulated DEGs) between OS and normal samples were screened. Functional analysis showed that these DEGs were markedly enriched in 214 GO terms and 54 KEGG pathways such as pathways in cancer. Five genes (CAMP, METTL7A, TCN1, LTF and CXCL12) acted as hub genes in PPI network. Besides, METTL7A, CYP4F3, TCN1, LTF and NETO2 were key genes in TF-gene network. Moreover, Pax-6 regulated four key genes (TCN1, CYP4F3, NETO2 and CXCL12). The expression levels of four genes (METTL7A, TCN1, CXCL12 and NETO2) in GSE39262 set were consistent with our integration analysis. The expression levels of two genes (CXCL12 and NETO2) in GSE126209 set were consistent with our integration analysis. ROC analysis of GSE39262 set revealed that CYP4F3, CXCL12, METTL7A, TCN1 and NETO2 had good diagnostic values for OS patients. ROC analysis of GSE126209 set revealed that CXCL12, METTL7A, TCN1 and NETO2 had good diagnostic values for OS patients.

Identification of potential gene signatures associated with osteosarcoma by integrated bioinformatics analysis

Yutao Jia[1], Yang Liu[1], Zhihua Han[2#], Rong Tian[1]

[1.] Department of Spine Surgery, Tianjin Union Medical Center, Tianjin, China;

[2.] Department of Anesthesiology, Tianjin Union Medical Center, Tianjin, China;

[#]**Correspondence to:**

Zhihua Han;

Address: Department of Anesthesiology, Tianjin Union Medical Center; No.190, Jieyuan Road, Hongqiao District, Tianjin 300121, China;

Tel: (+86)022-27557374

Email: hanzhihua1996@163.com

Running title: Potential gene with osteosarcoma

Abstract

Background. Osteosarcoma (OS) is the most primary malignant bone cancer in children and adolescents with a high mortality rate. This work aims to screen novel potential gene signatures associated with OS by integrated microarray analysis of Gene Expression Omnibus (GEO) database.

Material and Methods. The OS microarray datasets were searched and downloaded from GEO database to identify differentially expressed genes (DEGs) between OS and normal samples.. Afterwards, the functional enrichment analysis, protein-protein interaction (PPI) network analysis and transcription factor (TF)-target gene regulatory network were applied to uncover the biological function of DEGs. Finally, two published OS datasets (GSE39262 and GSE126209) were obtained from GEO database for evaluating the expression level and diagnostic values of key genes.

Results. Totally, 1059 DEGs (569 up-regulated DEGs and 490 down-regulated DEGs) between OS and normal samples were screened. Functional analysis showed that these DEGs were markedly enriched in 214 GO terms and 54 KEGG pathways such as pathways in cancer. Five genes (*CAMP*, *METTL7A*, *TCN1*, *LTF* and *CXCL12*) acted as hub genes in PPI network. Besides, *METTL7A*, *CYP4F3*, *TCN1*, *LTF* and *NETO2* were key genes in TF-gene network. Moreover, Pax-6 regulated four key genes (*TCN1*, *CYP4F3*, *NETO2* and *CXCL12*). The expression levels of four genes (*METTL7A*, *TCN1*, *CXCL12* and *NETO2*) in GSE39262 set were consistent with our integration analysis. The expression levels of two genes (*CXCL12* and *NETO2*) in GSE126209 set were consistent with our integration analysis. ROC analysis of GSE39262 set revealed that *CYP4F3*, *CXCL12*, *METTL7A*, *TCN1* and *NETO2* had good diagnostic values for OS patients. ROC analysis of GSE126209 set revealed that *CXCL12*, *METTL7A*, *TCN1* and *NETO2* had good diagnostic values for OS patients.

Conclusion. Four genes (*CXCL12*, *METTL7A*, *TCN1* and *NETO2*) may be involved in the development of OS, which will contribute to develop novel therapeutic strategies for OS management.

45 **Keywords**: Osteosarcoma, Diagnosis, Genes, Activating transcription factor, Bioinformatic.

46

Introduction

Osteosarcoma (OS) is a type of primary malignant bone cancer that causes public health concern, especially in children and adolescents (Isakoff MS 2015; Lindsey et al. 2017). Several treatment strategies of OS such as surgical resection, traditional adjuvant chemotherapy and radiotherapy have been favored by clinical oncologists in the past few decades (Nagarajan et al. 2011). Accordingly, the 5-year survival rate has been raised to approximately 70% (Bielack SS 2002). However, it is estimated that 80% of OS patients may suffer from the micro-metastasis, which cannot be detected at early diagnosis (Messerschmitt et al. 2009). Although multiple methods for the diagnosis and treatment of OS have been developed, new methods for the prevention and treatment of OS are still needed. The pathogenesis of OS progression remains not fully understood. Therefore, the identification of effective diagnostic makers and exploring underlying molecular etiology of OS is a pressing need.

The emergence of high-throughput sequencing technology has become an effective way to illuminate the pathogenic genes in a variety of human diseases, which help to explore pathogenesis and develop biomarkers. Many research groups have screened multiple biomarkers associated with OS by analyzing gene expression data. For example, Sun *et al* evaluated the difference of genes in the expression level between OS metastasis and OS non-metastasis and discovered that *TGFB1*, *LFT3*, *KDM1A*, and *KRAS* may participate in the occurrence of OS. Xiong *et al* found that *CCT3*, *COPS3* and *WWP1* were involved in the OS development by integrating gene expression and genomic aberration data (Xiong Y 2015). Liu *et al* constructed a co-expression network based on a Gene Expression Omnibus (GEO) dataset and identified many potential biomarkers such as *CTLA4* and *PBF* for diagnosis and treatment of OS (Liu X 2017). However, the molecular mechanisms of OS initiation and development have not been fully explored.

In this study, we retrieved GEO database and obtained four OS datasets. Subsequently, the differentially expressed genes (DEGs) between OS samples and normal samples were obtained and subjected to functional analysis. A protein-protein interaction (PPI) network was constructed

followed by the establishment of transcription factor (TF)-target regulatory network. Following this, we downloaded two published GEO datasets of OS as the validation set for assessing the expression levels of key candidate genes. Finally, the receiver operating characteristic (ROC) analysis was conducted to evaluate the diagnostic values of key candidate genes. This study will discover novel gene signatures associated with OS, providing new trains of thought for the diagnosis and treatment of OS.

Materials and methods

Data acquisition

The datasets were retrieved from the National Center for Biotechnology Information-GEO (http://www.ncbi.nlm.nih.gov/geo/) repository using the key terms of 'osteosarcoma' AND 'Homo sapiens'[porgn]. All selected datasets in this study should meet the following criteria: ⅰ) the datasets contained genome-wide expression data of tumor tissues and normal control tissues of OS patients; and ⅱ) all data were standardized or raw data. As shown in table S1, a total of five datasets were obtained. Notably, the GSE9508 dataset contained over 50% missing data and was subsequently removed from the following analysis. Eventually, four datasets (GSE12865, GSE19276, GSE87624 and GSE99671) were included in this study, which included 118 OS tissues and 28 normal bone tissues. The GSE12865 series (GPL6244 platform) included a total of 14 samples (12 OS and 2 normal control tissues). The platform for GSE19276 was GPL6848 including 44 OS and 5 normal control bone tissue. The platform for GSE87624, consisting of 44 OS and 3 normal control bone tissue samples, was GPL11154. GSE99671 was in GPL20148 platform, which contained 18 OS and 18 normal control bone tissue samples. The platform and series matrix files were downloaded. The dataset information was listed in table S1. The impact of different platforms on the results, we normalized the data through the log function, and centralized and standardized the scale function to eliminate the impact of the dimension on the data structure.

Data pre-processing and DEGs identification

The standardized data from included datasets were firstly processed as follows: ⅰ) the probes

that mapped to several genes were deleted; and ⅱ) if the gene was matched by multiple probes, the probe with the greatest gene expression value would be retained. Overall, there were overlapping 14981 genes among four datasets. Subsequently, MetaMA (https://cran.r-project.org/web/packages/metaMA/), an R package, was used to combine data from four GEO datasets. Individual p-values were obtained using Limma R package. The inverse normal method was used to combine *P* values in meta-analysis (Marot G 2009). We carried out the multiple comparison correction by Benjamini & Hochberg approach to acquire false discovery rate (FDR). Herein, the DEGs between OS tissues and normal controls were defined according to the cutoff of false discovery rate (FDR) < 0.05 and those DEGs with different directionality were removed from this study. Finally, the hierarchical clustering analysis of top 100 DEGs was also carried out by pheatmap package in R software.

Functional enrichment analyses

To systematically explore the underlying biological functions of identified DEGs, the Metascape (http://metascape.org/gp/index.html), an online tool that integrates multiple data resources such as Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) and Universal Protein (Uniprot) database, was used to perform GO and KEGG pathway enrichment analysis of DEGs. GO analysis involved three categories: biological process (BP), cellular component (CC) and molecular function (MF). *P* values < 0.05 was set as the thresholds for significant enrichment analyses.

Protein-protein interaction (PPI) network analysis

The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database, a freely web-based analytic tool, can predict the interactions among proteins (Szklarczyk et al. 2019). Here, a PPI analysis was conducted to examine the interactive associations between protein products of DEGs. The Cytoscape software (http://www.cytoscape.org) was utilized to establish a PPI network. In addition, the CytoNCA (http://apps.cytoscape.org/apps/cytonca) was used to analyze topological characteristics of PPI network. The top 15 nodes were considered as hub genes according to the degree value.

TF-candidate gene network analysis

TFs can bind to specific DNA sequences in promoter region of target gene to regulate gene expression. The top 20 up- and down-regulated genes were regarded as the candidate genes. The DNA sequences (2 kb) in the upstream promoter region of these candidate genes were firstly downloaded from the University of California, Santa Cruz (http://www.genome.ucsc.edu/) databases. Then, we employed online match tool from TRANSFAC (http://genexplain.com/transfac) to predict potential TFs that targeted candidate genes. Notably, the TFs that had only one binding site with target genes were retained in this study. Finally, the Cytoscape software was used to build a transcriptional regulatory network and perform node degree analysis.

Evaluation the expression level and diagnostic values of key genes

Two published OS datasets were obtained from GEO database for the expression level evaluation of seven key DEGs (*CAMP*, *METTL7A*, *TCN1*, *LTF*, *CXCL12*, *CYP4F3* and *NETO2*). Then, we performed a ROC analysis using the pROC package in R software (http://web.expasy.org/pROC/) to evaluate the diagnostic value of these seven DEGs. Accordingly, the area under the curve (AUC) was computed and the ROC curve was built. The AUC value > 0.8 showed a good diagnostic value for OS.

Results

Identification of DEGs

After data pre-processing, a total of 1059 DEGs (569 up-regulated genes and 490 down-regulated genes) were identified between OS samples and normal controls according to methods described above. The clustering analysis indicated that top 100 DEGs could distinguish OS samples and controls from four datasets (Figure S1). The top 20 up- and down-regulated genes were listed in table 1.

GO and KEGG enrichment analysis of DEGs

The GO enrichment analysis of DEGs showed that a total of 214 GO terms were enriched, including 194 GO-BP terms, 15 GO-CC terms and 5 GO-MF terms (Table S1). Specifically, for

GO-BP analysis, these DEGs were strongly associated with positive regulation of cell death, myeloid cell activation involved in immune response and regulation of protein kinase activity. Meanwhile, protein domain specific binding and transcription factor binding were significantly enriched GO-MF terms. Many DEGs were primarily involved in multiple GO-CC terms, such as anchored component of membrane and tertiary granule. The top 20 clusters of significantly enriched GO terms were displayed in figure 1. In addition, these DEGs were markedly enriched in 20 KEGG pathways such as regulation of lipolysis in adipocytes, protein processing in endoplasmic reticulum and pathways in cancer (Table 2). Notably, the top 20 up-regulated genes did not enrich in any KEGG pathway. However, four of top 20 down-regulated genes played vital roles in multiple significantly enriched KEGG pathways, including *FABP4* (fatty acid binding protein 4), *CXCL12* (C-X-C motif chemokine ligand 12), *CXCL12* (C-X-C motif chemokine ligand 12) and *CAT* (Catalase). More specifically, *FABP4* was involved in regulation of lipolysis in adipocytes and *CXCL12* participated in pathways in cancer and axon guidance (Table 2). *CYP4F3* was closely correlated with arachidonic acid metabolism pathway and *CAT* was significantly enriched in biosynthesis of amino acids and AMPK signaling pathway (Table 2).

PPI network analysis

To determine the relationships among DEGs, a PPI network was built based on the STRING database, which included 109 nodes and 196 protein pairs (Figure 2). The top 15 hub genes contain *PPBP* (pro-platelet basic protein; degree = 13), *CAMP* (cathelicidin antimicrobial peptide, degree = 13), *LTF* (lactotransferrin, degree = 12), *BST1* (bone marrow stromal cell antigen 1, degree = 12), *CXCR2* (C-X-C motif chemokine receptor 2, degree = 10), *OLFM4* (olfactomedin 4, degree = 10), *STOM* (stomatin, degree = 10), *TCN1* (transcobalamin 1, degree = 9), *SLC4A1* (solute carrier family 4 member 1, degree = 9), *LTA4H* (leukotriene A4 hydrolase, degree = 9), *S100A9* (S100 calcium binding protein A9, degree = 9), *CXCL12* (degree = 8), *CLEC12A* (C-type lectin domain family 12 member A, degree = 8), *RAB37* (RAB37, member RAS oncogene family, degree = 8) and *METTL7A* (methyltransferase like 7A, degree = 8). More

notably, these genes were all down-regulated.

TF-target network analysis

TF exerts crucial roles in regulating the expression of target gene. Herein, we employed TRANSFAC to predict TFs that regulated 20 up-regulated and down-regulated DEGs. As shown in Figure 3, the TF-target gene regulatory network contained 95 nodes (55 TFs and 40 genes) and 275 TF-gene pairs. The top 15 genes in TF-target network were *CHL1* (cell adhesion molecule L1 like, down-regulation, degree = 16), *SERPINB2* (serpin family B member 2, down-regulation, degree = 15), *SLC28A3* (solute carrier family 28 member 3, down-regulation, degree = 11), *ZC3H8* (zinc finger CCCH-type containing 8, up-regulation, degree = 10), *LAPTM4B* (lysosomal protein transmembrane 4 beta, up-regulation, degree = 10), *DYRK4* (dual specificity tyrosine phosphorylation regulated kinase 4, up-regulation, degree = 9), *METTL7A* (solute carrier family 28 member 3, down-regulation, degree = 9), *KCNG1* (potassium voltage-gated channel modifier subfamily G member 1, down-regulation, degree = 9), *CYP4F3* (down-regulation, degree = 9), *GNL2* (G protein nucleolar 2, up-regulation, degree = 8), *HBD* (hemoglobin subunit delta, down-regulation, degree = 8), *TCN1* (down-regulation, degree = 8), *ZZZ3* (zinc finger ZZ-type containing 3, up-regulation, degree = 8), *NETO2* (neuropilin and tolloid like 2, up-regulation, degree = 8), and *LTF* (lactotransferrin, down-regulation, degree = 8). In addition, the top 6 TFs that covered the most downstream genes were exhibited in table 3, which contained Pax-4, 1-Oct, Nkx2-5, HNF-4, FOXD3 and Pax-6. Interestingly, *TCN1*, *CYP4F3*, *NETO2* and *CXCL12* were regulated by Pax-6 (Table 3).

Evaluation the expression level and diagnostic values of key genes

An external dataset (GSE39262) was obtained from GEO database, which contained 10 human osteosarcoma cell lines and 5 untransformed cell lines samples. The platform for this dataset was GPL96 [HG-U133A] Affymetrix Human Genome U133A Array. GSE126209 was downloaded from GEO database, which inclued 12 osteosarcomas tumors and 11 adjacent normal tissues samples. The platform for this dataset was GPL20301 Illumina HiSeq 4000. Seven key genes (*CAMP*, *METTL7A*, *TCN1*, *LTF*, *CXCL12*, *CYP4F3* and *NETO2*) were selected to verify in

GSE39262. Among them, *CAMP*, *METTL7A*, *TCN1*, *LTF* and *CXCL12* acted as hub genes in PPI network. *METTL7A*, *CYP4F3*, *TCN1*, *LTF* and *NETO2* were key genes in TF-gene network. Moreover, Pax-6 regulated four key genes (*TCN1*, *CYP4F3*, *NETO2* and *CXCL12*). The gene differential expression analysis of GSE39262 dataset revealed that *NETO2* was significantly up-regulated while *CXCL12, METTL7A* and *TCN1* were significantly down-regulated, which were consistent with our integration analysis (Figure 4; Table S2). The gene differential expression analysis of GSE126209 dataset displayed that *NETO2* was significantly up-regulated while *CXCL12* was significantly down-regulated, which were consistent with our integration analysis (Figure 5; Table S3). ROC analysis is a commonly used method to evaluate the value of genetic diagnosis and has been used in previously biomedical works (Le et al. 2019; Le et al. 2019; Thi et al. 2020). Additionally, the results of GSE39262 dataset showed that five genes had good diagnostic values for OS (*CXCL12*, *CYP4F3*, *METTL7A*, *NETO2* and *TCN1*; Figure 6). The AUC of *CXCL12* was 1.000 and the specificity and sensitivity of this model were 100.0% and 100%, respectively. The AUC of *CYP4F3* was 0.840 and the specificity and sensitivity of this model were 80.0% and 80.0%, respectively. The AUC of *METTL7A* was 0.900 and the specificity and sensitivity of this model were 100.0% and 70.0%, respectively. The AUC of *NETO2* was 0.860 and the specificity and sensitivity of this model were 100.0% and 80.0%, respectively. The AUC of *TCN1*was 0.820 and the specificity and sensitivity of this model were 80.0% and 90.0%, respectively. Ultimately, the results of GSE126209 dataset showed that four genes had good diagnostic values for OS (*CXCL12*, *METTL7A*, *NETO2* and *TCN1*; Figure 7). *CXCL12* was 1.000 and the specificity and sensitivity of this model were 100.0% and 100%, respectively. The AUC of *METTL7A* was 0.856 and the specificity and sensitivity of this model were 100.0% and 83.3%, respectively. The AUC of *NETO2* was 0.833 and the specificity and sensitivity of this model were 100.0% and 75.0%, respectively. The AUC of *TCN1*was 0.879 and the specificity and sensitivity of this model were 81.8% and 83.3%, respectively.

Discussion

OS is a common malignant bone tumor and originates from mesenchymal stromal cells (MSCs)

(Xiao W 2013). The heterogeneous histopathological characteristics and complex genomic landscape of OS have been major challenges for elaborating underlying the molecular pathogenesis of OS. In this study, we included four OS datasets and identified 1059 DEGs (569 up-regulated DEGs and 490 down-regulated DEGs) between OS and normal samples. These genes were significantly enriched in 54 KEGG pathways such as pathways in cancer. Moreover, *CAMP*, *METTL7A*, *TCN1*, *LTF* and *CXCL12* served as hub genes in PPI network. *METTL7A*, *CYP4F3*, *TCN1*, *LTF* and *NETO2* were key players in TF-target gene regulatory network. Interestingly, *TCN1*, *CYP4F3*, *NETO2* and *CXCL12* were all regulated by Pax-6. Additionally, the expression patterns of key genes (*CAMP*, *METTL7A*, *TCN1*, *LTF*, *CXCL12*, *CYP4F3* and *NETO2*) were selected to verify in two published OS datasets (GSE39262 and GSE126209).

CAMP, also known as *hCAP18* or *LL37*, is an antimicrobial peptide gene in human (Larrick et al. 1995). The C-terminal of the protein product of *CAMP* contains a 37-amino acid-long peptide with broad spectrum-antibacterial activity (Vandamme D 2012). There are positive expressions of *CAMP* in the multiple cell systems, such as epithelial cells, neutrophils and macrophages (Dhawan P 2015; Frew L 2014; Li Y 2018). Wu *et al* suggested that bone marrow stroma could express *CAMP*, which may be a potential *ex vivo* priming factor for hematopoietic stem progenitor cells to promote hematopoietic reconstitution after transplantation (Wu et al. 2012). Later, Coffelt *et al* discovered that *CAMP* expression level was elevated in MSCs compared to that in ovarian cancer cells (Coffelt SB 2019). Herein, our analysis showed that *CAMP* was the most down-regulated gene in patients suffering from OS. Besides, *CAMP* acted as a hub gene in PPI network, suggesting that this gene may be involved in the pathologic mechanism of OS. Although the underlying role of *CAMP* on the initiation and progression of OS has not been investigated, available evidence showes that *CAMP* plays significant roles in several cancers, including breast cancer, lung cancer and pancreatic cancer (García-Quiroz J 2016; Sainz et al. 2015; von Haussen et al. 2008). More notably, existing data indicated that *CAMP* had either carcinogenic or anti-cancer effects (Chen X 2018; Wu WK 2010). Therefore, the influence of *CAMP* on OS occurrence and development needs to be further clarified in future.

Our gene differential expression revealed that *CXCL12* and *TCN1* were down-regulated in OS patients, which were verified in a validation dataset. Moreover, these two genes also acted as hub genes in PPI network. In addition, up-regulated *NETO2* and down-regulated *CYP4F3* had high degree in TF-gene regulatory network. Interestingly, *CXCL12*, *TCN1*, *NETO2* and *CYP4F3,* regulated by Pax-6, exhibited important diagnostic values for OS. *CXCL12* is also called stromal cell-derived factor-1 (*SDF-1)* and can bind to G-protein-coupled chemokine receptor CXCR4 (T 2014). Increasing studies suggested that *CXCL12*/CXCR4 axis played pivotal roles in tumor growth and development (F 2004; Lu Y 2015; Perissinotto E 2005). Li *et al* highlighted that epigenetic regulation of *CXCL12* by DNA methyltransferase 1 was associated with the metastasis and immune response in OS (Li B 2018). Previous reports also indicated that down-regulation of CXCR4 induced OS cell apoptosis via suppressing PI3K/Akt/NF-κβ pathway (Pollino S 2019). However, there is no directive evidence to support the involvement of *TCN1*, *NETO2* and *CYP4F3* in OS. Notably, Pax-6 is a highly conserved evolutionarily TF and belongs to paired box TF family (Mansouri A 1996). Several studies have pointed out that Pax6 participated in the regulation of cancer cell proliferation and progression (Shyr et al. 2010; Zong X 2011). Yang *et al* established a TF-top 20 DEGs regulatory network by integrating and analyzing three GEO datasets (GSE66673, GSE49003 and GSE37552), and found that Pax-6 down-regulated *BMP6* expression in non-metastatic OS samples (Yang C 2019). Taken together, we inferred that *CXCL12*/*TCN1*/*NETO2*/*CYP4F3*-Pax-6 axis may be implicated in the pathogenesis of OS, and four genes (*CXCL12*, *TCN1*, *NETO2* and *CYP4F3*) were novel diagnostic biomakers for OS.

METTL7A and *LTF* are reported to act as tumor suppressor genes (Qi et al. 2017; Zhang et al. 2011; Zhang et al. 2015). Similarly, our findings showed that the expressions of *METTL7A* and *LTF* were decreased in OS samples. Moreover, these two genes were both hub genes in PPI analysis and key gene nodes in TF-gene regulatory analysis. These results implied that *METTL7A* and *LTF* may be correlated with underlying mechanisms of OS. However, the potential effects of *METTL7A* and *LTF* down-regulation on OS progression needs to be further

investigated.

Although we have identified multiple novel gene signatures associated with OS, there are still limitations in this work. Our conclusion was drawn based on an integrated bioinformatic analysis. Therefore, additional experiments are required to confirm our findings. In addition, a larger sample size verification will also improve the reliability of our conclusion. Moreover, the clinical information should be collected to evaluate the diagnostic value of biomarkers for OS patients. Finally, the biological significances of key biomarkers will be investigated in model systems or cell lines.

In summary, a total of 1059 DEGs were identified between OS and normal samples. Among them, up-regulation of *NETO2* and down-regulation of *METTL7A*, *TCN1*, and *CXCL12* may be potential gene signatures related to OS. Pax-6 was also probably associated with the pathological process of OS. However, a comprehensive bioinformatics analysis with larger sample size and *in vivo* or *in vitro* assays should be performed to confirm our results.

Acknowledgement

Not applicable.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Bielack SS K-BB, Delling G, Exner GU, Flege S, Helmke K, Kotz R, Salzer-Kuntschik M, Werner M, Winkelmann W, Zoubek A, Jürgens H, Winkler K. 2002. Prognostic factors in high-grade osteosarcoma of the extremities or trunk: an analysis of 1,702 patients treated on neoadjuvant cooperative osteosarcoma study group protocols. *J Clin Oncol* 20:776-790.

Chen X ZX, Qi G, Tang Y, Guo Y, Si J, Liang L. 2018. Roles and Mechanisms of Human Cathelicidin LL-37 in Cancer. *Cell Physiol Biochem* 47:1060-1073.

Coffelt SB MF, Watson K, Zwezdaryk KJ, Dembinski JL, LaMarca HL, Tomchuck SL, Honer zu Bentrup K, Danka ES, Henkle SL, Scandurro AB. 2019. The pro-inflammatory peptide LL-37 promotes ovarian tumor progression through recruitment of multipotent mesenchymal stromal cells. *Proc Natl Acad Sci U S A* 106:3806-3811.

Dhawan P WR, Sun C, Gombart AF, Koeffler HP, Diamond G, Christakos S. 2015. C/EBPα and the Vitamin D Receptor Cooperate in the Regulation of Cathelicidin in Lung Epithelial Cells. *J Cell Physiol* 230:464-472.

F B. 2004. Cancer and the chemokine network. *Nat Rev Cancer* 4:540-550.

Frew L MS, McKinlay AT, McHugh BJ, Doust A, Norman JE, Davidson DJ, Stock SJ. 2014. Human cathelicidin production by the cervix. *PLoS One* 9:e103434.

García-Quiroz J G-BR, Santos-Martínez N, Avila E, Larrea F, Díaz L. 2016. Calcitriol stimulates gene expression of cathelicidin antimicrobial peptide in breast cancer cells with different phenotype. *J Biomed Sci* 23:78.

Isakoff MS BS, Meltzer P, Gorlick R. 2015. Osteosarcoma: Current Treatment and a Collaborative Pathway to Success. *J Clin Oncol* 33:3029-3035.

Larrick J, Hirata M, Balint RF, Lee J, Zhong J, and Wright S. 1995. Human CAP18: a novel antimicrobial lipopolysaccharide-binding protein. *Infect Immun* 63:1291-1297.

Le N Q K, Yapp E K Y, H. Y. Yeh. 2019. ET-GRU: using multi-layer gated recurrent units to identify electron transport proteins. *BMC Bioinformatics* 20.1.

Le N Q K, Yapp, E.K. Y, Nagasundaram N, Chua, M C H,Yeh H. 2019. Computational identification of vesicular transport proteins from sequences using deep gated recurrent units architecture. *Computational and Structural Biotechnology Journal* 17.

Li B WZ, Wu H, Xue M, Lin P, Wang S, Lin N, Huang X, Pan W, Liu M, Yan X, Qu H, Sun L, Li H, Wu Y, Teng W, Wang Z, Zhou X, Chen H, Poznansky MC, Ye Z. 2018. Epigenetic Regulation of CXCL12 Plays a Critical Role in Mediating Tumor Progression and the Immune Response In Osteosarcoma. *Cancer Res* 78:3938-3953.

Li Y ØS, Johnsen IB. 2018. Human Metapneumovirus Infection Inhibits Cathelicidin Antimicrobial Peptide Expression in Human Macrophages. *Front Immunol* 9:902.

Lindsey BA, Markel JE, and Kleinerman ES. 2017. Osteosarcoma Overview. *Rheumatol Ther* 4:25–43.

Liu X HA, Zhao JL, Chen FL. 2017. Identification of Key Gene Modules in Human Osteosarcoma by Co-Expression Analysis Weighted Gene Co-Expression Network Analysis (WGCNA). *J Cell Biochem* 118:3953-3959.

Lu Y HB, Guan GF, Chen J, Wang CQ, Ma Q, Wen YH, Qiu XC, Zhang XP, Zhou Y. 2015.

353 SDF-1/CXCR4 promotes F5M2 osteosarcoma cell migration by activating the Wnt/β-
354 catenin signaling pathway. *Med Oncol* 32:194.
355 Mansouri A HM, Gruss P. 1996. Pax genes and their roles in cell differentiation and
356 development. *Curr Opin Cell Biol* 8:851-857.
357 Marot G FJ, Mayer CD, Jaffrézic F. 2009. Moderated effect size and p-value combinations for
358 microarray meta-analyses. *Bioinformatics* 25:2692-2699.
359 Messerschmitt PJ, Garcia RM, FW A-K, EM G, and Getty PJ. 2009. Osteosarcoma. *J Am Acad*
360 *Orthop Surg* 17:515-527.
361 Nagarajan R, Kamruzzaman A, Ness KK, Marchese VG, Sklar C, Mertens A, Yasui Y, Robison
362 LI, and Marina N. 2011. Twenty years of follow-up of survivors of childhood
363 osteosarcoma: a report from the Childhood Cancer Survivor Study. *Cancer* 117:625-634.
364 Perissinotto E CG, Leone F, Fonsato V, Mitola S, Grignani G, Surrenti N, Sangiolo D, Bussolino
365 F, Piacibello W, Aglietta M. 2005. Involvement of chemokine receptor 4/stromal cell-
366 derived factor 1 system during osteosarcoma tumor progression. *Clin Cancer Res* 11:490-
367 497.
368 Pollino S PE, Dozza B, Bientinesi E, Piccinni-Leopardi M, Lucarelli E, Righi A, Benassi MS,
369 Pazzaglia L. 2019. CXCR4 in human osteosarcoma malignant progression. The response
370 of osteosarcoma cell lines to the fully human CXCR4 antibody MDX1338. *J Bone Oncol*
371 17:100239.
372 Qi L, Song Y, Chan THM, Yang H, Lin CH, Tay DJT, Hong H, Tang SJ, Tan KT, Huang XX,
373 Lin JS, Ng VHE, Maury JJP, Tenen DG, and Chen L. 2017. An RNA editing/dsRNA

binding-independent gene regulatory mechanism of ADARs and its clinical implication in cancer. *Nucleic Acids Res* 45:10436–10451.

Sainz B, Jr., Alcala S, Garcia E, Sanchez-Ripoll Y, Azevedo MM, Cioffi M, Tatari M, Miranda-Lorenzo I, Hidalgo M, Gomez-Lopez G, Cañamero M, Erkan M, Kleeff J, García-Silva S, Sancho P, Hermann PC, and Heeschen C. 2015. Microenvironmental hCAP-18/LL-37 promotes pancreatic ductal adenocarcinoma by activating its cancer stem cell compartment. *Gut* 64:1921-1935.

Shyr C, Tsai MY, Yeh S, HY K, YC C, Wong Pl, CC H, KE H, and Chang C. 2010. Tumor suppressor PAX6 functions as androgen receptor co-repressor to inhibit prostate cancer growth. *Prostate* 70:190–199.

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, and Mering Cv. 2019. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research* 47:D607-D613. 10.1093/nar/gky1131

Thi D, Trang L, Khanh L. 2020. Using deep neural networks and biological subwords to detect protein S-sulfenylation sites. *Briefings in Bioinformatics*.

T N. 2014. CXC chemokine ligand 12 (CXCL12) and its receptor CXCR4. *J Mol Med (Berl)* 92:433–439.

Vandamme D LB, Luyten W, Schoofs L. 2012. A comprehensive summary of LL-37, the factotum human cathelicidin peptide. *Immunol* 280:22-35.

von Haussen J, Koczulla R, Shaykhiev R, Herr C, Pinkenburg O, Reimer D, Wiewrodt R, Biesterfeld S, Aigner A, Czubayko F, and Bals R. 2008. The host defence peptide LL-37/hCAP-18 is a growth factor for lung cancer cells. *Lung Cancer* 59:12–23.

Wu W, Kim CH, Liu R, Kucia M, Marlicz W, Greco N, Ratajczak J, MJ L, and Ratajczak M. 2012. The bone marrow-expressed antimicrobial cationic peptide LL-37 enhances the responsiveness of hematopoietic stem progenitor cells to an SDF-1 gradient and accelerates their engraftment after transplantation. *Leukemia* 26:736-745.

Wu WK WG, Coffelt SB, Betancourt AM, Lee CW, Fan D, Wu K, Yu J, Sung JJ, Cho CH. 2010. Emerging roles of the host defense peptide LL-37 in human cancer and its potential therapeutic applications. *Int J Cancer* 127:1741-1747.

Xiao W MA, Hogendoorn PC, Cleton-Jansen AM. 2013. Mesenchymal stem cell transformation and sarcoma genesis. *Clin Sarcoma Res* 3:10.

Xiong Y WS, Du Q, Wang A, Wang Z. 2015. Integrated analysis of gene expression and genomic aberration data in osteosarcoma (OS). *Cancer Gene Ther* 22:524-529.

Yang C HD, Ma C, Ren J, Fu L, Cheng C, Li B, Shi X. 2019. Identification of Pathogenic Genes and Transcription Factors in Osteosarcoma. *Pathol Oncol Res*.

Zhang H, Feng X, Liu W, Jiang X, Shan W, Huang C, Yi H, Zhu B, Zhou W, Wang L, Liu C, Zhang L, Jia W, Huang W, Li G, Shi J, Wanggou S, Yao K, and Ren C. 2011. Underlying mechanisms for LTF inactivation and its functional analysis in nasopharyngeal carcinoma cell lines. *J Cell Biochem* 112:1832-1843.

Zhang J, Ling T, Wu H, and Wang K. 2015. Re-expression of Lactotransferrin, a candidate

416 tumor suppressor inactivated by promoter hypermethylation, impairs the malignance of
417 oral squamous cell carcinoma cells. *J Oral Pathol Med* 44:578-584.
418 Zong X YH, Yu Y, Zou D, Ling Z, He X, Meng X. 2011. Possible role of Pax-6 in promoting
419 breast cancer cell proliferation and tumorigenesis. *BMB Rep* 44:595-600.
420
421

Figure legends

Figure 1 Top 20 significantly enriched Gene Ontology terms of differentially expressed genes.

Figure 2 Protein-protein interaction networks of differentially expressed genes. Red and green ellipses represent up-regulated and down-regulated genes, respectively. The black borders indicate top 20 up-regulated and down-regulated genes.

Figure 3 Transcription factor-top 20 up-regulated and down-regulated genes network. Diamonds and ellipses represent transcription factors and top 20 up-regulated and down-regulated genes, respectively. Red and green ellipses represent up-regulated and down-regulated genes, respectively.

Figure 4 Box plots of seven differentially expressed genes in GSE39262 dataset. The x-axes represent control and case groups while the y-axes represent the relative expression levels of the genes. Seven genes included *NETO2*, *CAMP*, *METTL7A*, *TCN1*, *LTF*, *CXCL12* and *CYP4F3*.

Figure 5 Box plots of seven differentially expressed genes in GSE126209 dataset. The x-axes represent control and case groups while the y-axes represent the relative expression levels of the genes. Seven genes included *NETO2*, *CAMP*, *METTL7A*, *TCN1*, *LTF*, *CXCL12* and *CYP4F3*.

Figure 6 ROC curves of selected differentially expressed genes in GSE39262 dataset. The x-axes and the y-axes show 1-specificity and sensitivity, respectively. *ROC*, receiver operating characteristic. Seven genes included *NETO2*, *CAMP*, *METTL7A*, *TCN1*, *LTF*, *CXCL12* and *CYP4F3*.

Figure 7 ROC curves of selected differentially expressed genes in GSE126209 dataset. The x-axes and the y-axes show 1-specificity and sensitivity, respectively. *ROC*, receiver operating characteristic. Seven genes included *NETO2*, *CAMP*, *METTL7A*, *TCN1*, *LTF*, *CXCL12* and *CYP4F3*.

Figure S1 The heat map of the top 100 differentially expressed genes.

Figure 1

Top 20 significantly enriched Gene Ontology terms of differentially expressed genes.

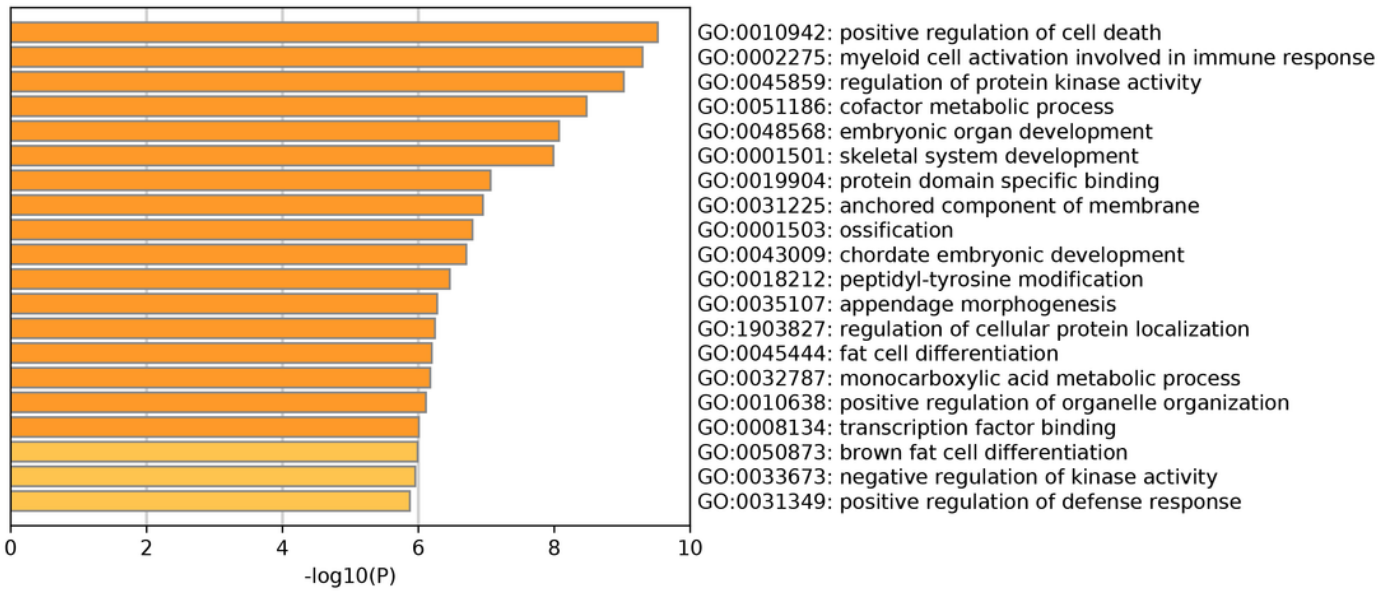

Figure 2

Protein-protein interaction networks of differentially expressed genes. Red and green ellipses represent up-regulated and down-regulated genes, respectively. The black borders indicate top 20 up-regulated and down-regulated genes.

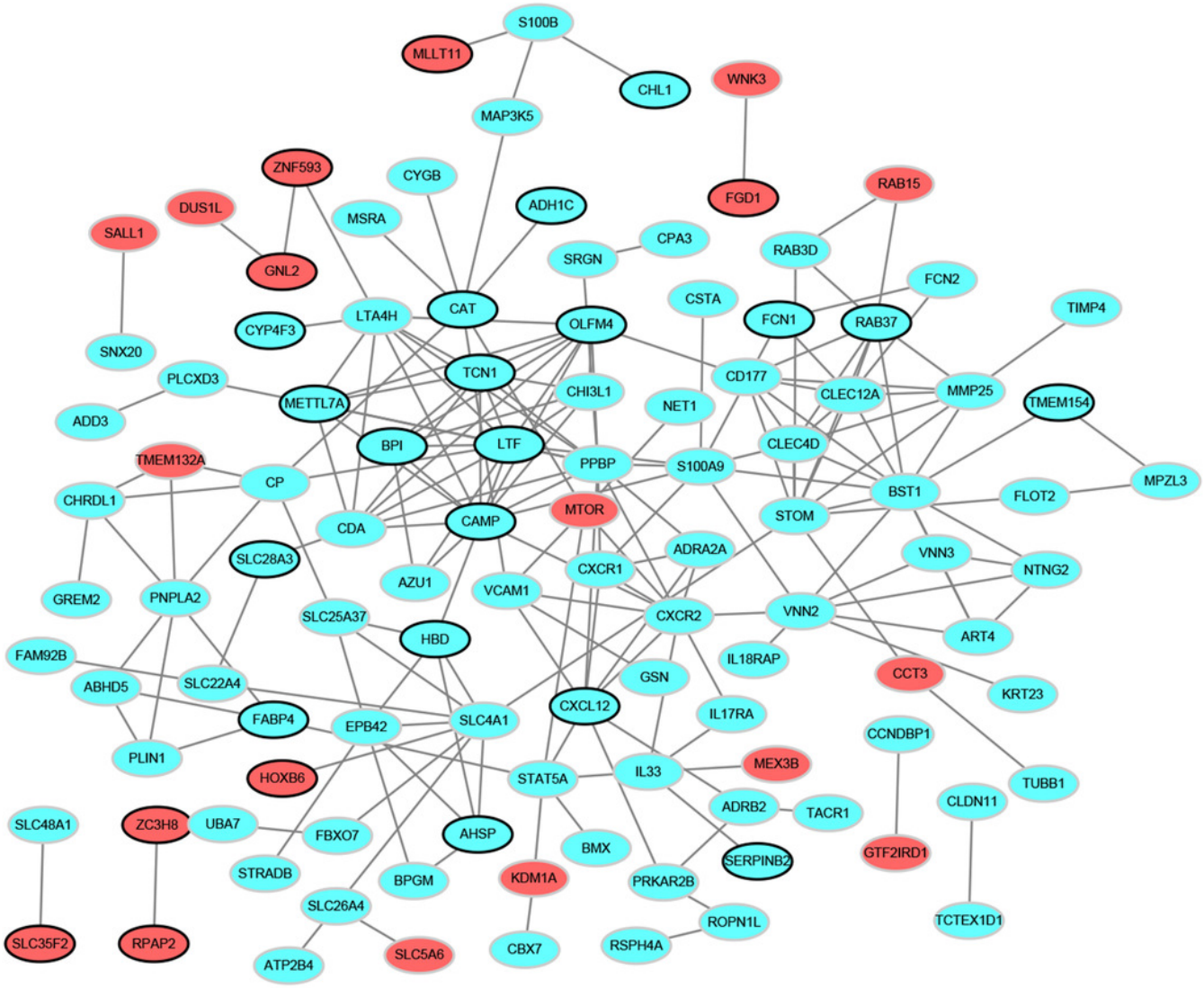

Figure 3

Transcription factor-top 20 up-regulated and down-regulated genes network. Diamonds and ellipses represent transcription factors and top 20 up-regulated and down-regulated genes, respectively. Red and green ellipses represent up-regulated and down-reg

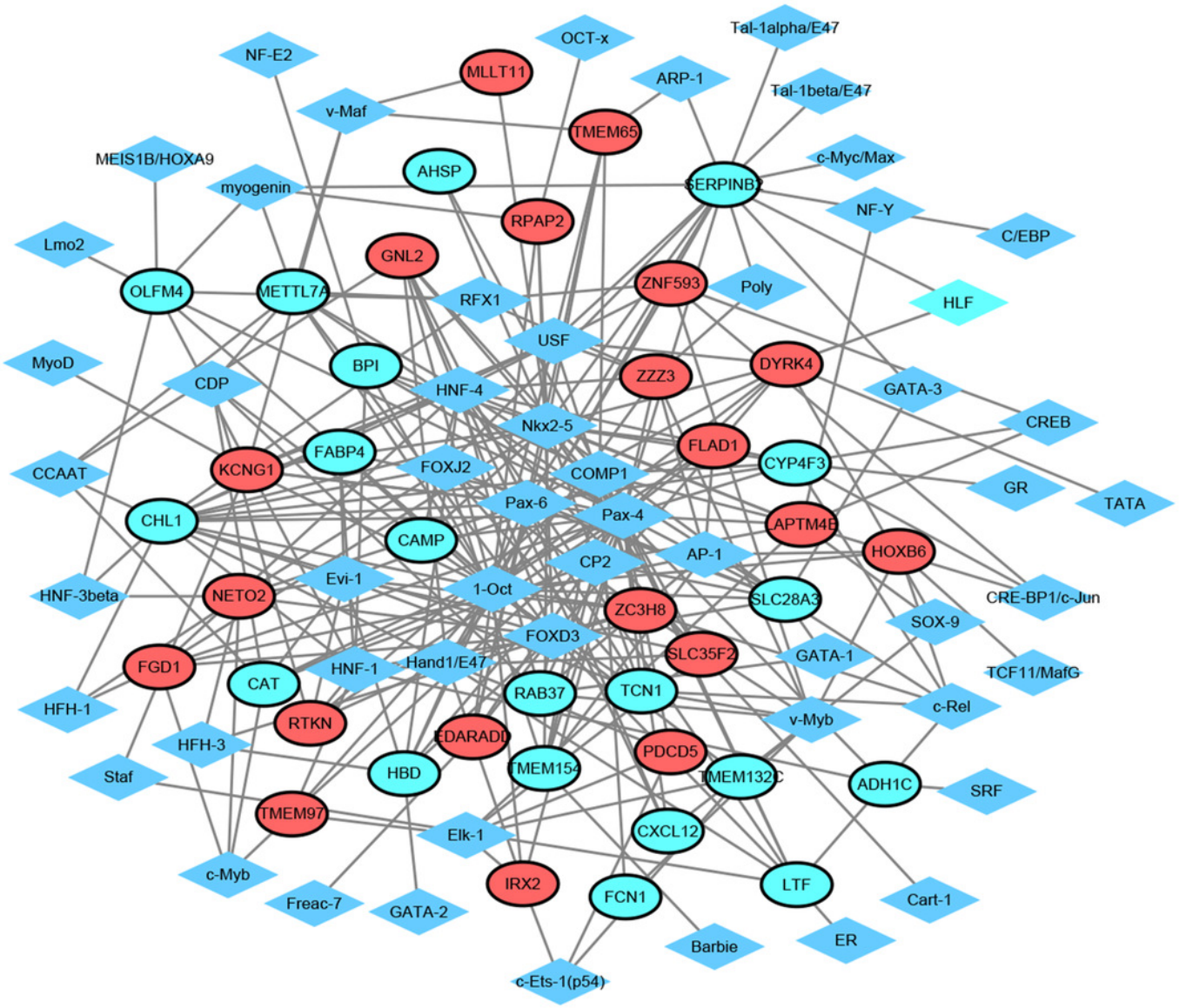

Figure 4

Box plots of seven differentially expressed genes in GSE39262 dataset.

The x-axes represent control and case groups while the y-axes represent the relative expression levels of the genes. Seven genes included *NETO2*, *CAMP*, *METTL7A*, *TCN1*, *LTF*, *CXCL12* and *CYP4F3*.

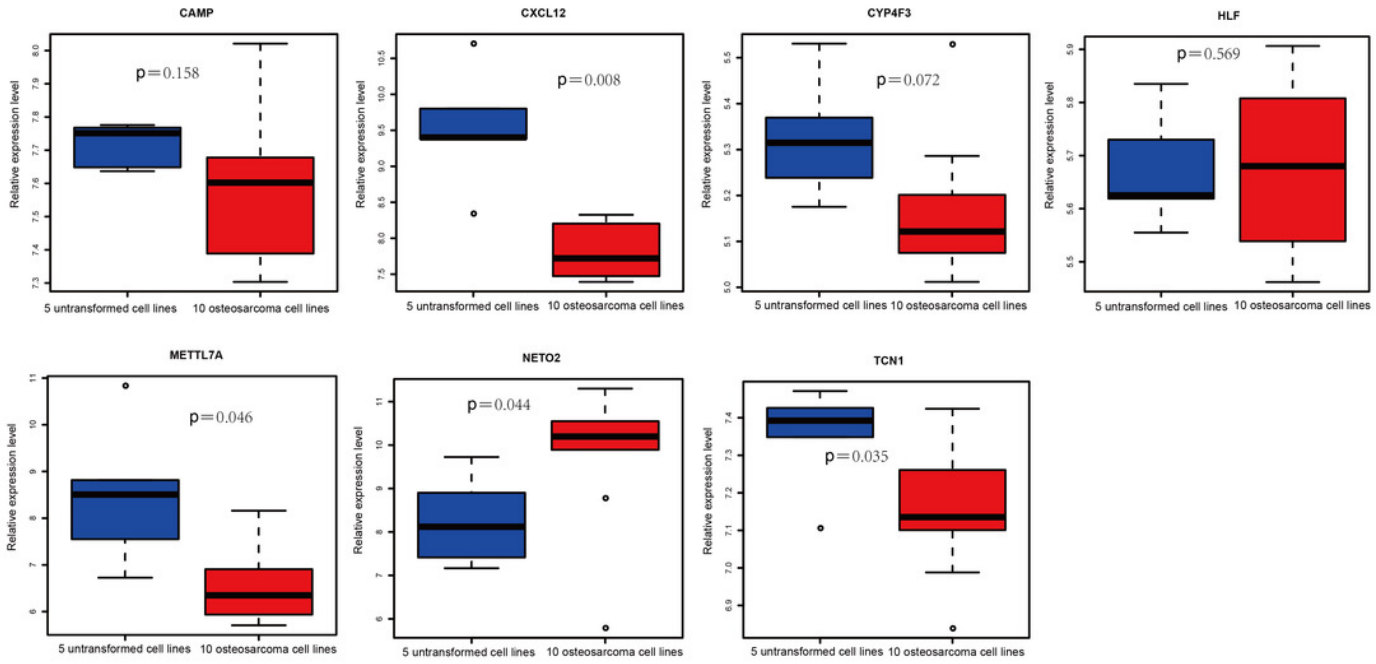

Figure 5

Box plots of seven differentially expressed genes in GSE126209 dataset.

The x-axes represent control and case groups while the y-axes represent the relative expression levels of the genes. Seven genes included *NETO2*, *CAMP*, *METTL7A*, *TCN1*, *LTF*, *CXCL12* and *CYP4F3.*

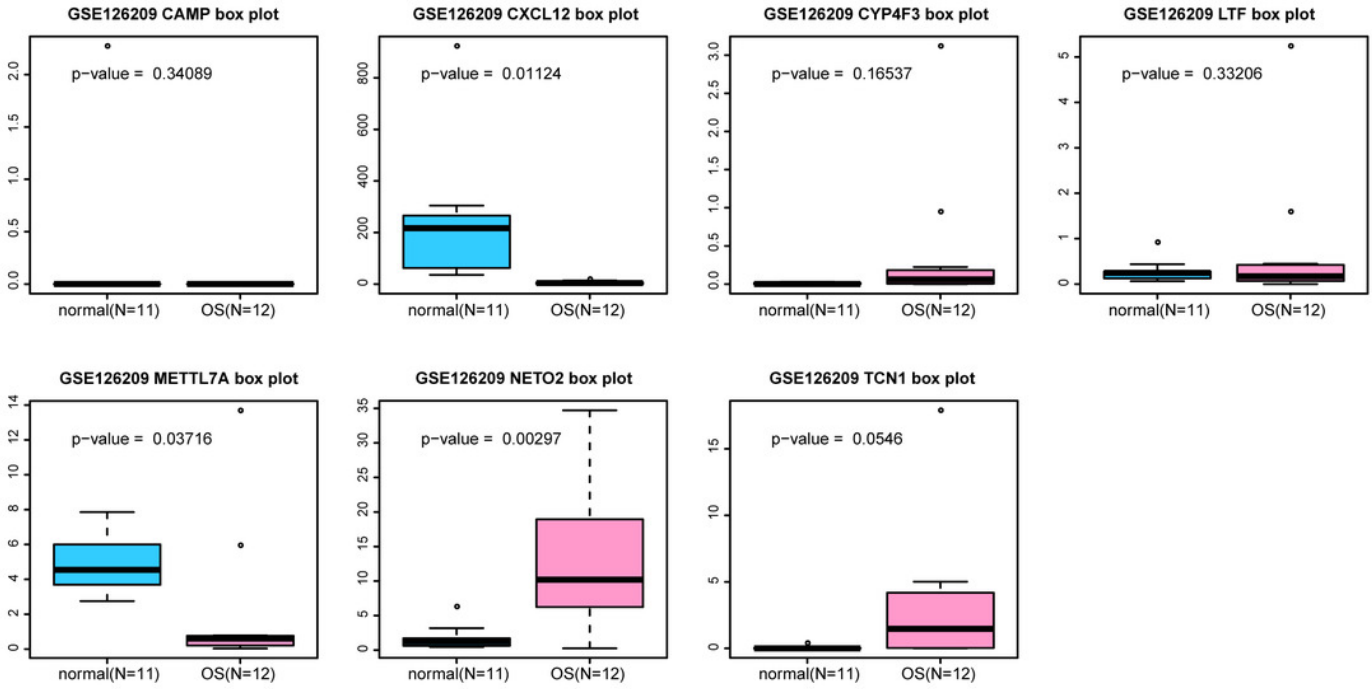

Figure 6

ROC curves of selected differentially expressed genes in GSE39262 dataset.

The x-axes and the y-axes show 1-specificity and sensitivity, respectively. *ROC*, receiver operating characteristic. Seven genes included *NETO2*, *CAMP*, *METTL7A*, *TCN1*, *LTF*, *CXCL12* and *CYP4F3*.

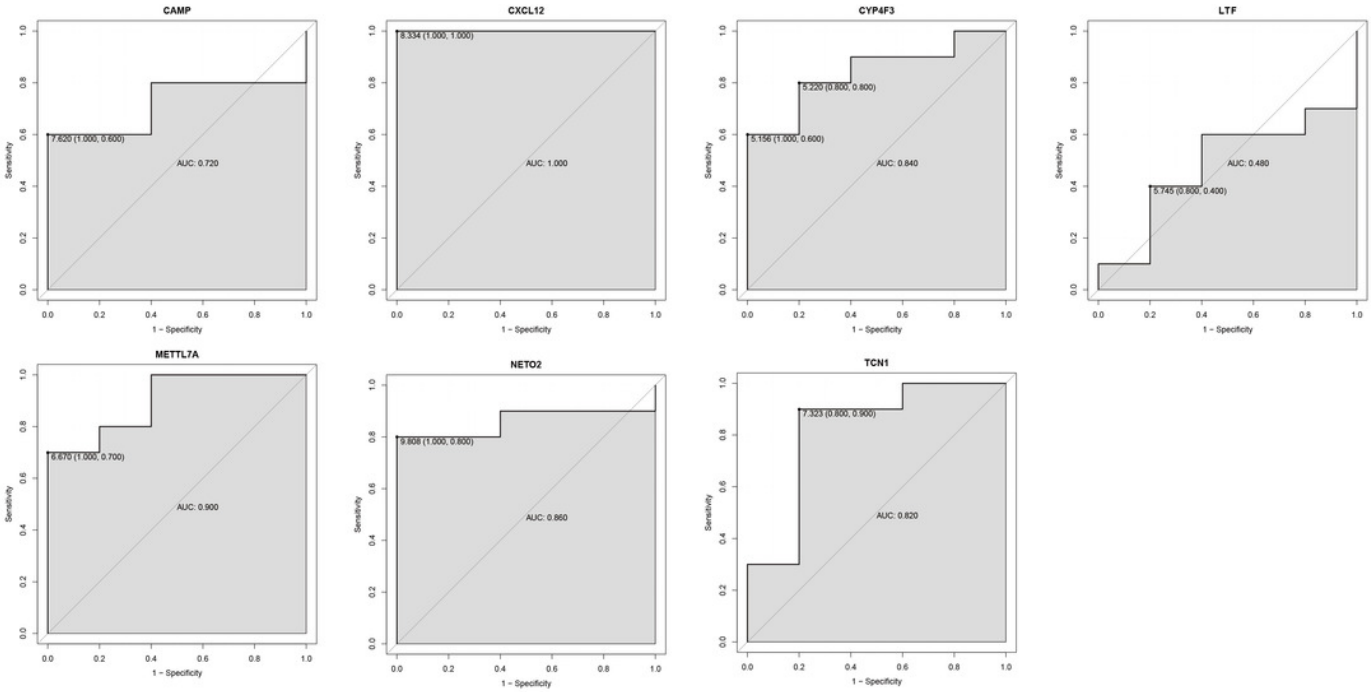

Figure 7

ROC curves of selected differentially expressed genes in GSE126209 dataset.

The x-axes and the y-axes show 1-specificity and sensitivity, respectively. *ROC*, receiver operating characteristic. Seven genes included *NETO2*, *CAMP*, *METTL7A*, *TCN1*, *LTF*, *CXCL12* and *CYP4F3.*

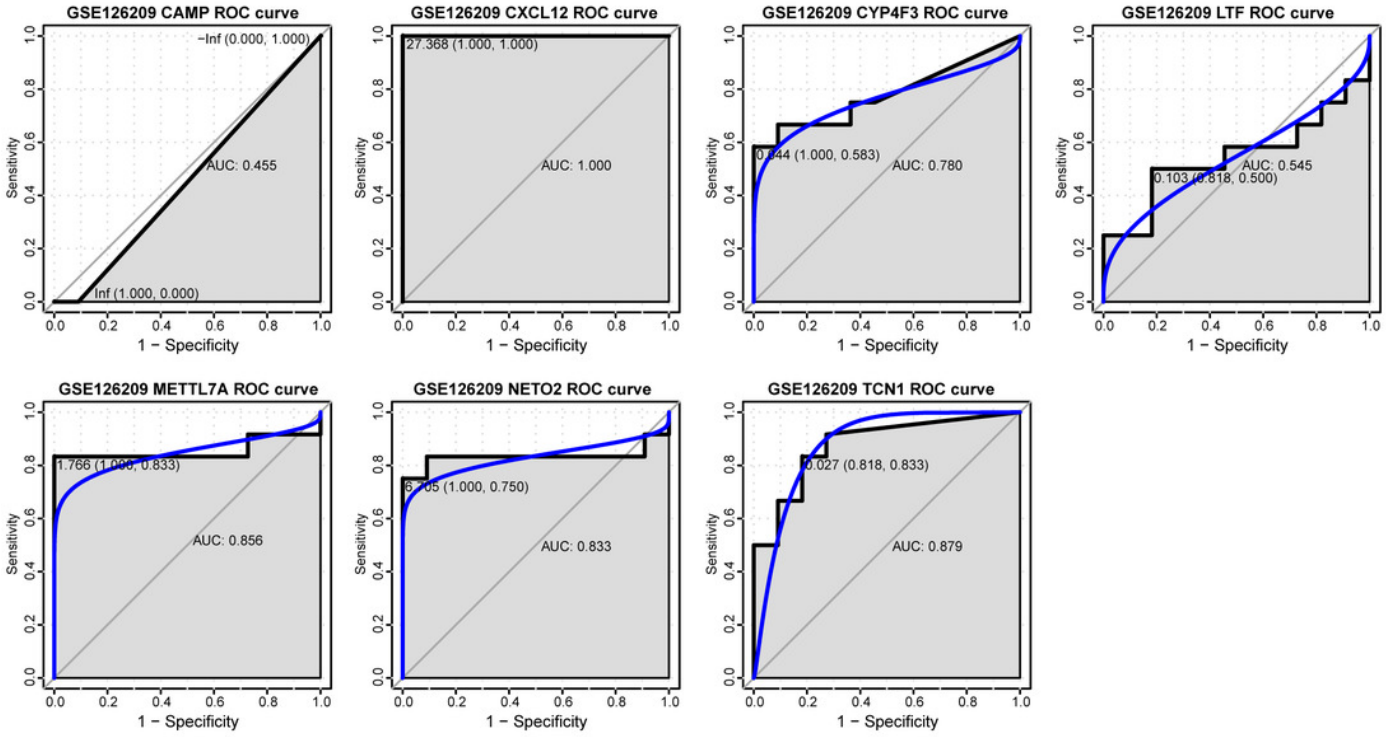

Table 1(on next page)

The list of top 20 up-regulated and down-regulated differentially expressed genes.

Table 1 The list of top 20 up-regulated and down-regulated differentially expressed genes.

Gene Symbol	Combined.ES	*P*_value	FDR	Up/Down-regulation
NETO2	1.773	5.48E-10	1.16E-07	Up-regulation
RTKN	1.664	3.90E-09	6.57E-07	Up-regulation
TMEM65	1.669	1.08E-08	1.54E-06	Up-regulation
MLLT11	1.507	1.41E-08	1.84E-06	Up-regulation
LAPTM4B	1.499	1.78E-08	2.21E-06	Up-regulation
ZC3H8	1.563	1.97E-08	2.41E-06	Up-regulation
SLC35F2	1.570	3.46E-08	3.90E-06	Up-regulation
IRX2	1.534	5.03E-08	5.19E-06	Up-regulation
ZNF593	1.404	7.11E-08	6.78E-06	Up-regulation
FLAD1	1.451	7.54E-08	7.10E-06	Up-regulation
KCNG1	1.456	1.32E-07	1.16E-05	Up-regulation
FGD1	1.304	2.59E-07	2.04E-05	Up-regulation
RPAP2	1.480	2.63E-07	2.06E-05	Up-regulation
DYRK4	1.405	2.83E-07	2.20E-05	Up-regulation
EDARADD	1.360	3.06E-07	2.33E-05	Up-regulation
PDCD5	1.285	3.77E-07	2.80E-05	Up-regulation
TMEM97	1.349	4.04E-07	2.95E-05	Up-regulation
GNL2	1.311	4.14E-07	3.01E-05	Up-regulation
HOXB6	1.527	4.47E-07	3.17E-05	Up-regulation
ZZZ3	1.504	5.79E-07	3.99E-05	Up-regulation
CAMP	-4.158	0	0	Down-regulation
AHSP	-3.156	0	0	Down-regulation
OLFM4	-3.108	0	0	Down-regulation
LTF	-3.006	0	0	Down-regulation
ADH1C	-2.814	0	0	Down-regulation

CXCL12	-2.746	0	0	Down-regulation
BPI	-2.556	0	0	Down-regulation
HBD	-2.615	8.88E-16	5.32E-13	Down-regulation
TCN1	-2.522	1.11E-15	6.40E-13	Down-regulation
FABP4	-2.296	7.77E-15	4.31E-12	Down-regulation
RAB37	-2.279	2.02E-14	1.08E-11	Down-regulation
FCN1	-2.320	2.95E-14	1.53E-11	Down-regulation
TMEM154	-2.279	3.13E-14	1.55E-11	Down-regulation
METTL7A	-2.181	5.44E-14	2.47E-11	Down-regulation
CYP4F3	-2.304	6.66E-14	2.94E-11	Down-regulation
CAT	-2.109	3.44E-13	1.43E-10	Down-regulation
CHL1	-1.974	5.51E-13	2.17E-10	Down-regulation
TMEM132C	-1.860	5.19E-12	1.85E-09	Down-regulation
SERPINB2	-3.082	5.64E-12	1.97E-09	Down-regulation
SLC28A3	-2.118	8.26E-12	2.69E-09	Down-regulation

2

Table 2(on next page)

The top 20 significantly enriched KEGG pathways

Table 2 The top 20 significantly enriched KEGG pathways.

ID	Term	*P*_value	Count	Gene Symbols
hsa04923	Regulation of lipolysis in adipocytes	0.0001	10	*ADRB2*, *FABP4,* *PDE3B*, *PIK3CD*, *PLIN1*, *PTGER3*, *IRS2*, *ABHD5*, *PNPLA2*, *ADCY4*
hsa04141	Protein processing in endoplasmic reticulum	0.0001	19	*BAG1*, *EIF2S1*, *HSPA2*, *LMAN1*, *MAN1A1*, *MAP3K5*, *P4HB*, *EIF2AK2*, *RPN2*, *RRBP1*, *SSR2*, *PREB*, *SEC24A*, *SEC61G*, *UBQLN2*, *DNAJB11*, *UBQLN4*, *DNAJC1*, *UBE2J2*
hsa05200	Pathways in cancer	0.0002	34	*BCL2L1*, *CASP8*, *CDKN2A*, *CEBPA*, *CKS1B*, *COL4A1*, *DVL1*, *E2F1*, *EPAS1*, *FGF7*, *FGF13*, *FLT3*, *MTOR*, *GLI2*, *GLI3*, *GNAQ*, *PIK3CD*, *PTEN*, *PTGER2*, *PTGER3*, *RARB*, *RXRA*, *CXCL12*, *SKP2*, *STAT5A*, *TGFB3*, *TGFBR2*, *VEGFA*, *NCOA4*, *FGF16*, *PIAS2*, *ARHGEF11*, *LEF1*, *ADCY4*, *SHC1*, *CALM1*, *CHEK1*, *ELK1*, *FDPS*, *GPS2*, *MSX2*, *RANBP1*, *TLN1*, *VCAM1*, *TLN2*
hsa01230	Biosynthesis of amino acids	0.0004	11	*CTH*, *ENO1*, *GAPDH*, *PC*, *SHMT1*, *TALDO1*, *TKT*, *TPI1*, *SDS*, *RPIA*, *AADAT*, *CAT*, *MDH2*, *ME1*, *MCEE*
hsa04974	Protein digestion and absorption	0.0019	11	*COL2A1*, *COL4A1*, *COL5A2*, *COL10A1*, *COL11A1*, *COL17A1*, *CPA3*, *SLC7A8*, *COL18A1*, *COL27A1*, *SLC16A10*
hsa04722	Neurotrophin signaling pathway	0.0022	13	*CALM1*, *MAPK14*, *MAP3K5*, *NFKBIE*, *NTF3*, *NTRK2*, *PIK3CD*, *PRKCD*, *RPS6KA2*, *SHC1*, *MAGED1*, *PRDM4*, *IRAK3*

				CPT1A, *ELAVL1*, *MTOR*, *LEP*, *PIK3CD*, *PPP2R5A*, *PPP2R5D*, *PRKAB2*, *IRS2*, *CREB5*, *CAMKK2*, *STRADB*, *CREB3L1*, *PRKCD*, *PRKCE*, *PTEN*, *RPS6KA2*, *TBC1D4*, *JAK2*, *NFKBIE*, *RXRA*, *SOCS2*, *CAT*, *ADCY4*, *CALM1*,
hsa04152	AMPK signaling pathway	0.0023	13	*ELK1*, *FLOT2*, *PDE3B*, *PRKAR2B*, *SHC1*, *INPP5K*, *HSPA2*
M00007	Pentose phosphate pathway, non-oxidative phase, fructose 6P => ribose 5P	0.0026	3	*TALDO1*, *TKT*, *RPIA*
hsa03060	Protein export	0.0028	5	*OXA1L*, *SRP54*, *SRP72*, *SEC61G*, *SRPRB*
hsa04933	AGE-RAGE signaling pathway in diabetic complications	0.0041	11	*COL4A1*, *MAPK14*, *JAK2*, *PIK3CD*, *PRKCD*, *PRKCE*, *STAT5A*, *TGFB3*, *TGFBR2*, *VCAM1*, *VEGFA*, *CD247*, *MTOR*, *NFKBIE*, *RXRA*, *STAT6*, *IL27RA*, *LHB*, *SHC1*, *SOCS2*, *AOX1*, *BCL2L1*, *IL5RA*, *LEP*, *PIAS2*
hsa04360	Axon guidance	0.0045	16	*EFNA1*, *EFNA3*, *EFNA5*, *EFNB1*, *EPHA3*, *EPHB2*, *EPHB6*, *FES*, *PIK3CD*, *CXCL12*, *SEMA7A*, *SEMA3A*, *RHOD*, *SEMA4C*, *NTNG2*, *PLXNA4*
hsa04015	Rap1 signaling pathway	0.0053	18	*CALM1*, *MAPK14*, *EFNA1*, *EFNA3*, *EFNA5*, *FGF7*, *FGF13*, *GNAQ*, *ITGB3*, *PFN2*, *PIK3CD*, *TLN1*, *VEGFA*, *FGF16*, *RAPGEF3*, *APBB1IP*, *TLN2*, *ADCY4*, *BCL2L1*, *COL2A1*, *COL4A1*, *MTOR*, *ITGA7*, *JAK2*, *PPP2R5A*, *PPP2R5D*, *PTEN*, *RXRA*, *TLR4*, *CREB5*, *CREB3L1*, *THEM4*
hsa00590	Arachidonic acid metabolism	0.0055	8	*ALOX15*, *GGT5*, *GPX3*, *GPX7*, *LTA4H*, *CYP4F3*, *PLA2G5*, *PTGES*, *GSTA4*, *GSTM3*, *SRM*, *GGCT*

hsa04932	Non-alcoholic fatty liver disease (NAFLD)	0.0060	14	*CASP8*, *CEBPA*, *EIF2S1*, *LEP*, *MAP3K5*, *NDUFB9*, *NDUFS5*, *PIK3CD*, *PRKAB2*, *RXRA*, *UQCRH*, *IRS2*, *NDUFB11*, *NDUFA4L2*
hsa00620	Pyruvate metabolism	0.0066	6	*GLO1*, *HAGH*, *MDH2*, *ME1*, *PC*, *LDHD*
hsa04924	Renin secretion	0.0073	8	*ADRB2*, *CALM1*, *GNAQ*, *PDE1A*, *PDE1C*, *PDE3B*, *PTGER2*, *CLCA4*
hsa04750	Inflammatory mediator regulation of TRP channels	0.0101	10	*CALM1*, *MAPK14*, *GNAQ*, *PIK3CD*, *PRKCD*, *PRKCE*, *PTGER2*, *TRPA1*, *TRPV4*, *ADCY4*, *ADRB2*, *ATP2B4*, *PPP2R5A*, *PPP2R5D*, *CREB5*, *RAPGEF3*, *CACNG7*, *CREB3L1*, *HSPA2*, *SHC1*, *PLA2G5*, *PTGIR*, *ARHGEF11*, *PPP1R14A*, *MYL6B*, *GPX3*, *GPX7*, *SLC26A4*, *ELK1*, *GNRH2*, *LHB*
hsa01524	Platinum drug resistance	0.0144	8	*BCL2L1*, *CASP8*, *CDKN2A*, *GSTA4*, *GSTM3*, *MAP3K5*, *PIK3CD*, *PMAIP1*, *BCL2A1*, *DFFA*, *DFFB*, *EIF2S1*, *PARP2*, *DIABLO*, *DAB2IP*, *MAPK14*, *VCAM1*, *CREB5*, *CREB3L1*, *MLKL*, *CALM1*, *ITGB3*, *PLAT*, *VEGFA*, *TRPV4*
hsa04071	Sphingolipid signaling pathway	0.0146	11	*MAPK14*, *MS4A2*, *GNAQ*, *MAP3K5*, *PIK3CD*, *PPP2R5A*, *PPP2R5D*, *PRKCE*, *PTEN*, *SPTLC2*, *SGMS1*
hsa00983	Drug metabolism-other enzymes	0.0147	6	*CDA*, *CES1*, *DPYD*, *UCK2*, *CES2*, *UCKL1*

KEGG Kyoto Encyclopedia of Genes and Genomes

Table 3(on next page)

The top 6 TF that has the most downstream genes

Table 3 The top 6 TF that has the most downstream genes.

TF	Number*	Gene Symbol
Pax-4	31	*FLAD1*, *ZZZ3*, *GNL2*, *EDARADD*, *RTKN*, *LTF*, *CAMP*, *CHL1*, *ADH1C*, *TMEM154*, *IRX2*, *FABP4*, *LAPTM4B*, *SLC28A3*, *FGD1*, *CXCL12*, *HBD*, *TCN1*, *CAT*, *SLC35F2*, *DYRK4*, *METTL7A*, *OLFM4*, *AHSP*, *RAB37*, *TMEM97*, *SERPINB2*, *CYP4F3*, *PDCD5*, *BPI*, *KCNG1*
1-Oct	22	*FLAD1*, *GNL2*, *RPAP2*, *EDARADD*, *ZC3H8*, *CHL1*, *CAMP*, *IRX2*, *FABP4*, *TMEM65*, *SLC28A3*, *HBD*, *TCN1*, *CAT*, *TMEM132C*, *DYRK4*, *METTL7A*, *TMEM97*, *HOXB6*, *SERPINB2*, *BPI*, *KCNG1*
Nkx2-5	19	*FLAD1*, *MLLT11*, *RPAP2*, *EDARADD*, *CAMP*, *CHL1*, *TMEM154*, *LAPTM4B*, *TMEM65*, *SLC28A3*, *FGD1*, *TCN1*, *SLC35F2*, *DYRK4*, *AHSP*, *SERPINB2*, *CYP4F3*, *KCNG1*, *BPI*
HNF-4	16	*ZZZ3*, *ZNF593*, *GNL2*, *LTF*, *CHL1*, *CAMP*, *LAPTM4B*, *SLC28A3*, *FGD1*, *CXCL12*, *HBD*, *SLC35F2*, *METTL7A*, *RAB37*, *CYP4F3*, *KCNG1*
FOXD3	14	*GNL2*, *RPAP2*, *EDARADD*, *ZC3H8*, *CHL1*, *LTF*, *FABP4*, *FGD1*, *HBD*, *DYRK4*, *METTL7A*, *OLFM4*, *NETO2*, *RAB37*
Pax-6	13	*EDARADD*, *GNL2*, *CHL1*, *TMEM65*, *SLC28A3*, *CXCL12*, *HBD*, *TCN1*, *DYRK4*, *NETO2*, *SERPINB2*, *CYP4F3*, *KCNG1*

Number* indicates the number of genes regulated by the TF. *TF* transcription factor