Screening key marker molecules of osteosarcoma using bioinformatics analysis technology

Ruiqing Xu
Ningxia Medical University
Ningjun Wan
Ningxia Medical University
Jiandang Shi (shi_jiandang@163.com)
General Hospital of Ningxia Medical University, Ningxia Hui Autonomous Region

Research Article

Keywords: osteosarcoma, bioinformatics analysis, differentially expressed genes, immune infiltration

Posted Date: July 6th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1787159/v1

Abstract

BACKGROUND: The most frequent primary bone tumor in children, adults, and elderly individuals over 60 years of age is osteosarcoma (OS), which is most commonly detected in the long bone epiphysis and has a poor prognosis and a high impairment rate. The precise etiology and underlying molecular pathways, however, remain unknown. Using bioinformatics approaches to evaluate the underlying biological mechanisms, this work attempted to discover differentially expressed genes (DEGs) in OS, implicated cell signaling pathways, and associated immune infiltration.

MATERIALS AND METHODS: The GSE14359 dataset's expression profiles were downloaded from the GEO database, which contains 20 samples, including 18 OS samples and 2 normal osteoblast control samples. Raw data were collected, DEGs were obtained, and bioinformatics was used to further analyze the data. The gene expression matrix was subjected to WGCNA-weighted gene co-expression network analysis using R (version 4.1.1) software, as well as gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of candidate genes. The STRING database was used to create protein-protein interaction (PPI) networks of candidate genes. Core gene network maps were created using the Cytoscape software. The R (version 4.1.1) software was used to rank key genes. Subsequently, plots of individual hub gene survival curves, expression and clinical correlation and promoter methylation and clinical correlation were produced via an online website (http://ualcan.path.uab.edu/). Finally, the infiltration of immune cells in OS was assessed using CIBERSORT.

RESULTS: A total of 625 DEGs were identified, of which 314 were up-regulated and 311 were down-regulated. Principal component analysis showed significant differences between the two tissues. WGCNA-weighted gene co-expression showed the highest absolute value of the MEgrey60 module, and a total of 247 genes were listed in the intersection of DEGs and MEgrey60 module genes taken.KEGG pathway analysis showed that these candidate genes mainly involved adherent spots, Rap1, PI3K-Akt, and MAPK, while GO pathway analysis mainly involved cellular processes such as osteogenesis, epithelial cell proliferation, osteoblast differentiation, and molecular functions such as growth factors and fibronectin binding. The 23 most closely related genes among the candidate genes were identified from the PPI network.

CONCLUSION: Our study shows that screening for DEGs, pathway enrichment and immune infiltration using bioinformatics analysis not only helps to understand the molecular mechanisms of OS occurrence but also facilitates us to better execute target gene interventions. However, this study still has limitations and subsequent clinical validation of the samples is needed.

Introduction

Osteosarcoma (OS) is the most common primary bone tumor in children, adults and the elderly over 60 years of age. It is commonly found in the epiphysis of long bones and has a poor prognosis and high disability rate[1]. Osteosarcoma is also the second leading cause of cancer-related mortality in children and adolescents[2]. The worldwide incidence of OS in 0–14 and 0–19 years old is 4 and 5 cases per million people per year, respectively. The incidence rate is higher in males than in females (5.4 and 4.0 cases per million per year, respectively). Two peaks in OS incidence occurred with increasing age, with the first peak occurring between the ages of 10 and 14 years, coinciding with a period of rapid pubertal development, suggesting a strong association between pubertal growth and OS. The second peak occurs above 65 years of age[3]. Most OS originates in the long bones, with 50% of cases occurring in the knee region, including the distal femur and proximal tibia[4]. As a rare and challenging neoplastic disease to treat, OS imposes a significant financial burden on patients and adds significant psychological stress to patients and physicians.OS is highly aggressive, with a metastasis rate of approximately 20%, and the most common target for metastasis is the lung. Currently, the main treatments for osteosarcoma are surgery, radiotherapy, and immunotherapy[5]. However, there is no effective treatment modality or the current treatments are clinically ineffective. Therefore, there is an urgent need to improve the treatment of OS and to find new therapeutic strategies[6].

The most promising direction now is to research the underlying molecular mechanisms of osteosarcoma to find more effective diagnostic techniques and more reliable molecular markers to detect the occurrence and prognostic assessment of OS. Gene expression microarrays are widely used for the study of gene expression profiles, providing a new way to explore genes and a broad application prospect for molecular targeting of drugs. Currently, considerable data on osteosarcoma have been published publicly on the GEO database. Moreover, the integration of these data allows for a more in-depth study of molecular mechanisms.

In this study, we downloaded a raw microarray dataset GSE14359 from the GEO database, which contained a total of 20 samples, 2 healthy controls and 18 OS samples. The R (version 4.1.1) software was used to screen the differentially expressed genes (DEGs) in OS samples and normal groups, and combined with WGCNA weighted gene co-expression network analysis to filter out the more relevant candidate differential genes. Subsequently and candidate genes were subjected to gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. Protein-protein interaction (PPI) network was used to analyze the association of specific candidate proteomes through the STRING online database. Finally, immune infiltration was analyzed by performing the CIBERSORT algorithm between OS and normal tissues, which was widely used to assess the relationship and relative content of 22 immune cell subpopulations. Due to sampling limitations and subjective selectivity differences, not all screened DEGs, involved pathways and immune infiltrates can be directly used as biomarkers.

Materials And Methods

Search Strategy

We searched the GEO database (https://www.ncbi.nlm.nih.gov/geo/) using the keyword "osteosarcoma" and from the beginning to November 7, 2019, there were 7642 results for "osteosarcoma" in the GEO database. " results. By restricting entry type (series), study type (array expression profile), and tissue source (Homo sapiens), 7427 items not relevant to the purpose of this study were excluded. After further review of dataset-specific information, datasets that did not contain paraneoplastic or normal tissues were further excluded. Finally, GSE14359 was selected on merit and the original GSE14359 gene expression profile data was downloaded. Specific information is shown in Table 1. the technical roadmap of this paper, as shown in Figure 1.

Microarray Data Information

GSE14359 was created by extracting mRNA from 5 frozen conventional osteosarcoma and 4 osteosarcoma lung metastasis tumor samples and mRNA from fresh primary osteoblasts and hybridizing them into HG U133A microarrays to form 2 control, 18 OS samples included. the platform for GSE14359 isGPL96 [HG-U133A] Affymetrix Human Genome U133A Array. Platform and series matrix files were downloaded from GEO and saved as TXT files. r software (version 4.1.1) was used to process the downloaded files. Specific information is in Table I.

Table 1 Characteristics of datasets

		Sample			
Dataset	platform				Country
		Normal	Primary tumor	Metastatic tumor	
GSE14359	Affymetric HG U133A	2 osteoblast	10 samples	8 samples	Germany

Microarray data collation and differential genes

The microarray dataset was processed for ID conversion to Gene symbol using Perl programming language (version 5.30.0) and genome annotation library, 20 samples were grouped and located in sample1.txt, sample2.txt files, limma data package was installed and loaded, gene expression difference analysis was performed using R software, and then saved as TXT file. And the up-or down-regulated genes were obtained individually by corresponding instructions and used for further analysis. Adjusted P values <0.05 and log fold change (logFC) >2 were considered as DEGs.

Data processing and identification of DEGs

The volcano map is used to display all up-and down-tuned DEGs using the same limma package. the results are then visualized using the heatmap package using a heat map.

WGCNA weighted gene co-expression network analysis

Given the excessive number of differential genes, further screening was required, followed by R software, installation and loading of the WGCNA package, checking for missing values, sample clustering, neighbour-joining matrix conversion, gene clustering, module identification, clustering and merging of similar modules, etc. Gene modules were identified, and the genes within the modules were the more strongly associated genes.

DEGs and WGCNA recognition modules take the intersection

With the R software, install and load the VennDiagram package and plot the Venn diagram

GO and KEGG pathway enrichment analysis of candidate differential genes

Analysis of the functional and pathway enrichment of the proteins encoded by the candidate genes and annotation of these genes using R software. Candidate genes were enriched using appropriate packages (biocLite "dose", "cluster profile" and "enrichplot") for differential genes for GO and KEGG pathway enrichment analysis. In this study, we analyzed the significantly up-and down-regulated DEGs determined from the integrated microarray RA data and considered the adjusted P-values <0.05 to be statistically significant.

PPI Network Analysis

The STRING website (https://string-db.org, version 11.5) is a global resource for the exploration and analysis of known and predicted protein-protein interactions. Concerning a specific set of proteins, the network view analysis predicts the association between proteins. Each network node represents a different protein, and the associations between these nodes represent biomolecular interactions that can be used to identify the interactions between these proteins encoded by candidate differential genes in OS and the associated pathways involved. Central nodes that are closely associated with other related proteins may be core or key proteins that play important physiological functions. In this paper, we chose the lowest interaction score of 0.9.

Core gene file preparation and core gene network map construction

The candidate differential gene expression files and STRING table files are generated by the perl software by executing the corresponding instructions to the node up-down gene files and network interoperation files. Further, the core gene network graph and core gene ranking files are generated and saved by cytoscape software. Finally, through R software, the corresponding code is executed to generate the key gene ranking histogram.

Online website analysis of hub gene and clinical correlation

On the UALCAN website (http://ualcan.path.uab.edu), enter the hub gene, select the "TCGA" database, target the cancer Sarcoma, and use the website to automatically generate the relevant profiles.

CIBERSORT analysis of immune infiltration

The normalized data filtered by Perl programming language were analyzed using CIBERSORT to obtain the immune cell infiltration matrix. In this study, 22 immune cells including macrophage M2, plasma cells, neutrophils, mast cell activation, T cell CD8, macrophage M1, T cell γδ, B cell memory, monocytes, undifferentiated B cells, T cell follicular helper cells, NK cell activation, dendritic cell quiescence, T cell CD4 memory activation, T cell CD4 undifferentiated, NK cell quiescence, regulatory T cells (Tregs), dendritic cell activation, eosinophils, macrophage M0, T cell CD4 memory quiescence, and mast cell quiescence. The percentage of immune cells in the gene expression matrix and the relationship between two immune cells were calculated and visualized by installing the e1071, parallel, preprocessCore, corrplot, and vioplot software packages.

Correlation of core genes and immunity with infiltrating cells and immune checkpoints

Through R software, ggplot2, ggpubr, and ggExtra packages were installed and loaded, and the corresponding instructions were executed to analyze the correlation of each core gene with immune infiltrating cells. Similarly, extract the corresponding information and execute the corresponding code to obtain the correlation of each core gene with immune checkpoints.

Results

A total of 625 DEGs were identified, of which 314 were up-regulated and 311 were down-regulated (logFC absolute value > 1, adj. P.Val < 0.05). And accordingly, the volcano map of up-and down-regulated genes, as well as the heat map of differential genes by R software, are shown in Figure 2-a and 2-b. PCA showed no overlap between these two groups and indicated a significant difference in gene expression matrix between OS group and the healthy control group ($p<0.05$), Figure 2-c.

WGCNA weighted gene co-expression network analysis

According to the gene expression matrix, 32 modules were identified in the normal group or tumor group by R software, respectively. The result of the module with the highest absolute value was shown as MEgrey60, and there were 594 genes in MEgrey60 module, as shown in Figure 3-a、3-b、3-c.

Differential gene and WGCNA identification module take the intersection

The intersection of DEG and MEgrey60 was taken by R software to obtain a total of 247 candidate differential genes, which were saved in TXT files and plotted in a Wayne diagram, as in Figure 4.

GO-term and KEGG pathway enrichment analysis of candidate differential genes

GO-term and KEGG pathway enrichment analyses of candidate differential genes with adjusted p-values <0.05 were obtained, respectively. Figure 5-a, Figure 5-b, and Table II show the KEGG enrichment results of the candidate differential genes. Figure 5-c, Figure 5-d, and Table III show the results of GO enrichment analysis of candidate differential genes. the KEGG pathway analysis showed that these candidate differential genes were mainly involved in adhesive spots, Rap1, PI3K-Akt, MAPK, and insulin resistance, while the GO pathway analysis was mainly involved in cellular processes such as osteogenesis, epithelial cell proliferation, osteoblast differentiation, and growth factors, fibronectin-binding and other molecular functions. In addition, adherent spots, insulin resistance, Rap1, PI3K-Akt, and MAPK pathways were relatively significantly enriched on KEGG, so the specific pathways are shown in Figure 6, Figure 7, and Figure 8. these significantly enriched pathways may enlighten our thinking and help us to further investigate the role of DEGs in OS.

Table 2 KEGG pathway of candidate DEGs

ID	Description	p-value	geneID	Count
hsa04510	Focal adhesion	4.97E-05	rapgef1/thbs1/vegfb/spp1/vegfc/ itga7/parva/egfr/ppp1cb/pxn/rapgef1/vav2	12
hsa04931	Insulin resistance	0.000221	tbc1d4/il6/rps6ka1/trib3/pygb/ppp1cb/ppp1r3c/pck2	8
hsa04015	Rap1 signaling pathway	0.000328	RAPGEF4/FLT1/EFNA1/THBS1/VEGFB/FGFR3/ VEGFC/EGFR/FGF7/RAPGEF1/VAV2	11
hsa04010	MAPK signaling pathway	0.000489	flt1/map3k12/efna1/mef2c/vegfb/fgfr3/vegfc/map4k3 /rps6ka1/egfr/ntrk2/fgf7/ntf3	13
hsa04151	PI3K-Akt signaling pathway	0.000897	flt1/efna1/thbs1/vegfb/spp1 fgfr3/vegfc/il6/itga7/egfr/ntrk2/fgf7/pck2/ntf3	14
hsa04014	Ras signaling pathway	0.008546	flt1/efna1/vegfb/fgfr3/vegfc/egfr/ntrk2/fgf7/ntf3	9
hsa04142	Lysosome	0.014937	ppt1/npc1/ctsl/ctsa/sgsh/ctsb	6
hsa04810	Regulation of actin cytoskeleton	0.017551	IQGAP2/FGFR3/ITGA7/EGFR/PPP1CB PXN/FGF7/VAV2	8
hsa00730	Thiamine metabolism	0.021043	ALPL/AK1	2
hsa05219	Bladder cancer	0.023855	THBS1/FGFR3/EGFR	3
hsa05202	Transcriptional misregulation in cancer	0.026486	flt1/igfbp3/mef2c/il6/jmjd1c/bcl11b MMP3	7
hsa04020	Calcium signaling pathway	0.029179	flt1/vegfb/drd1/fgfr3/vegfc/egfr NTRK2/FGF7	8
hsa05205	Proteoglycans in cancer	0.035992	smo/thbs1/ctsl/egfr/ppp1cb/pxn/vav2	7
hsa05144	Malaria	0.039787	HBB/THBS1/IL6	3
hsa04964	Proximal tubule bicarbonate reclamation	0.046882	CA2/PCK2	2

Table 3 GO analysis of candidate DEGs

ONTOLOGY	ID	Description	p.adjust	geneID	Count
BP	GO:0001503	ossification	0.000108	SNX10/SEMA4D/ALPL/PENK/SMO/IGFBP3/IGFBP5/FHL2/MEF2C/PDLIM7/ pth1r/spp1/fgfr3/vegfc/il6/col13a1/egfr/smad3/npr2/suco	20
BP	GO:0001649	osteoblast differentiation	0.000108	SEMA4D/ALPL/PENK/SMO/IGFBP3/IGFBP5/FHL2/MEF2C/PDLIM7 /PTH1R/SPP1/VEGFC/IL6/SMAD3/SUCO	15
BP	GO:0050673	epithelial cell proliferation	0.000198	fst/flt1/odam/smo/igfbp3/igfbp5/mef2c/thbs1/vegfb/vegfc/ IL6/CDH13/RPS6KA1/LOXL2/EGFR/SMAD3/BCL11B/FGF7/CLDN1/MMP12	20
BP	GO:0032963	collagen metabolic process	0.00416	CTSL/ADAMTS2/IL6/COL13A1/TRAM2/MMP3/SUCO/CTSB/MMP12	9
BP	GO:0048771	tissue remodeling	0.00416	SNX10/SEMA3C/IGFBP5/MEF2C/PTH1R/SPP1/CA2/IL6/ADAM8/EGFR/SUCO	11
BP	GO:0045453	bone resorption	0.00416	SNX10/PTH1R/SPP1/CA2/IL6/ADAM8/EGFR	7
BP	GO:0050927	positive regulation of positive chemotaxis	0.00416	VEGFB/VEGFC/CDH13/SMAD3/NTF3	5
BP	GO:0046849	bone remodeling	0.00416	SNX10/PTH1R/SPP1/CA2/IL6/ADAM8/EGFR/SUCO	8
BP	GO:0050926	regulation of positive chemotaxis	0.00419	VEGFB/VEGFC/CDH13/SMAD3/NTF3	5
BP	GO:0050920	regulation of chemotaxis	0.00419	sema4d/sema3e/gpr183/sema3c/thbs1/vegfb/ VEGFC/IL6/DPP4/CDH13/SMAD3/NTF3	12
MF	GO:0001968	fibronectin binding	0.000409	igfbp6/igfbp3/igfbp5/thbs1/ctsl/fbln1	6
MF	GO:0019838	growth factor binding	0.009811	flt1/igfbp6/igfbp3/igfbp5/thbs1/fgfr3/egfr/ntrk2/il11ra	9
MF	GO:0045499	chemorepellent activity	0.047405	SEMA4D/SEMA3E/SEMA3C/DPP4	4

Analysis of candidate differential genes in OS using PPI networks

The candidate DEG expression products in OS were constructed by STRING database to construct the PPI network. After removing some of the isolated and disconnected nodes, the lowest interaction score of 0.9 was chosen to construct an integrated network as shown in Figure 9-b. The 50 most important genes showing statistically significant interactions were ADAMTS2, THBS1,AOX1,MOCOS,CHI3L1,IL13RA2,CNDP2,GATM,COX7B,NDUFA4,CTSB,CTSL,DDX18,POLR1B,EFNA1,

VAV2,EGFR,PXN,SH3BGRL3,FHL2,IGFBP3,ELL,HMGN1,ENTPD1,NT5E,FGF7,FGFR3,FLT1,VEGFC,VEGFB,HEY1,HEY2,IGFBP5,

SPP1,IL6,MMP3,LOXL1,LOXL2,MEF2C,SMAD3,NRIP1,TBC1D4,NTF3,NTRK2,OPTN,SQSTM1,PARVA,PPP1CB,PPP1R3C,SH3BP2.among these genes,EGFR,FLT1,VAV2,and PXN had the highest nodality. As a result, the core gene network map was obtained by CYTOSCAPE software based on the PPI network interactions and the corresponding node files, as shown in Figure 9-c. Finally, the core ranking histogram of key genes was obtained from the above files, as shown in Figure 9-a for the top 23 genes. hub gene survival curves, except for FGF7, were not statistically different, as shown in Figure 10. hub gene expression, clinical correlation, as shown in Figures 11, 12, and 13. EGFR expression analysis in pan-cancer showed a significant difference in BRCA, CSC, CBM , PCPG differences were more pronounced, but no significant differences were seen in SARC. In the clinical correlation analysis, group comparisons with statistical differences existed only in the gender comparison, 81-100 years old group compared with 41-60 years old group and 61-80 years old group, respectively.FGF7 expression analysis in pan-cancer showed greater differences in BLCA, CESC, UCEC.FGF7 expression and clinical correlation analysis showed that in tumor tissues, TP53 mutated The analysis of FLT1 expression in pan-cancer showed a significant difference in CESC, KIRC, but not in SARC. the analysis of FGF7 expression and clinical correlation showed a significant difference in tumor tissues when comparing gender, and a statistically significant difference when comparing TP53 mutated group with the unmutated group. hub gene promoter methylation and clinical correlation, as shown in Figures 14, 15, and 16. analysis of promoter methylation levels of EGFR in SARC showed a statistically significant difference in the 61- to 80-year-old group compared with the normal group. 21- to 40-year-old group compared with the 61- to 80-year-old group. analysis of promoter methylation levels of FGF7 in SARC showed a statistically significant difference in the 21- to 40-year-old group compared with the 41- to 60-year-old group, respectively. Analysis of promoter methylation levels of FGF7 in SARC showed statistically significant differences in the 21-40 years group compared to the

41-60 years group, 61-80 years group, and 81-100 years group, respectively. analysis of promoter methylation levels of FLT1 in SARC showed significant differences in the normal group compared to the tumor group. Similarly, differences were shown in the age grouping.

Immuno-infiltration analysis

Figure 17-b Results show a positive correlation between regulatory T cells and naive CD4+ cells (value = 0.80). Neutrophils have a positive correlation with naive CD4+ T cells (value = 0.65). Regulatory T cells and γδ T cells had a positive correlation (value=0.74). Neutrophils had a positive correlation with regulatory T cells (value=0.72). Neutrophils had a positive correlation with γδ T cells (value=0.72). And naive B cells had a significant negative correlation with B cell memory cells (value = - 0.64). bar chart of 20 samples summarizes the relative percentages of 22 immune cells as shown in Figure 17-a. and Figure 18-a heatmap demonstrates the visualization of each immune infiltrating cell within each sample. The violin map of each immune infiltrating cell compared to normal tissue (Figure 18-b) shows that in OS tissue, infiltration of M2 macrophages, unactivated dendritic cells, and neutrophils is statistically greater, while infiltration of naive B cells and unactivated CD4+ memory T cells are statistically less. The correlation analysis of core genes with immune infiltrating cells is shown in Figures 19, 20, 21. The correlation analysis of core genes with each immune checkpoint is shown in Figures 22, 23.

Discussion

Most osteosarcomas develop from mesenchymal stem cells with osteogenic potential and are highly aggressive and distantly metastatic.[7]. The incidence in the general population is 2–3/million/year, but is higher in adolescents, with a peak annual incidence of 8–11/million/year at the age of 15–19 years[8]. It is histologically subdivided into conventional, low-grade central, periosteal, parosteal, capillary dilated, chondroblast and small cell types[9]. Current treatment of osteosarcoma is mostly based on chemotherapy regimens with adriamycin or surgery. Before multidrug chemotherapy, 90% of patients with osteosarcoma died from pulmonary metastases[10]. Data from next-generation sequencing studies have gradually revealed new patterns of genomic alterations in osteosarcoma, and may even identify new potential therapeutic targets. There is no doubt that epigenetic modifications also play an invaluable role in the development of osteosarcoma, with hypermethylation and hypomethylation observed in tissues. Genes that have been identified to have a role include TP53, RB1 and members of the RecQ DNA decapping enzyme family (BOX 1), which are mostly associated with drug resistance and metastasis in osteosarcoma. However, to date, research on common molecular therapeutic targets in osteosarcoma remains disappointing[11].

While current treatment for osteosarcoma is based on chemotherapy and surgery, a growing number of potential gene-targeted therapies are in development[12]. Therefore, it is crucial to observe and study the mechanisms of osteosarcoma development at the molecular level. Based on this, differentially expressed genes (DEGs) have been effectively used to predict OS phenotypes such as invasion, metastasis, angiogenesis, and drug resistance, which help us to be able to intervene in target genes from multiple aspects, and thus can play a role in inhibiting the development of osteosarcoma or improving the survival rate.

In this study, we performed a specific collation of gene expression profiles from the GEO database for GSE14359 and used R (version 4.1.1) to perform a specific code to analyze this dataset. We identified a total of 625 DEGs using the limma package, including 314 up-regulated genes and 311 down-regulated genes. Given that the number of screened DEGs was too large to facilitate our subsequent merit-based analysis, we immediately performed a WGCNA-weighted gene co-expression network analysis to identify 32 gene modules and selected the MEgrey60 module containing 594 genes based on absolute values as a way to find genes with the higher association. the DEGs and WGCNA-identified modules were taken to intersect to obtain candidate differential genes A total of 248 were obtained. A PPI network of differential candidate genes encoding proteins were constructed from the STRING database (minimum required interaction score = 0.990) and the 50 most important associated genes were screened. Among these genes, EGFR, FLT1, VAV2, and PXN had the highest modality. From this, the core gene network map was obtained by CYTOSCAPE software based on the PPI network interactions and the corresponding node files. Finally, the core ranking histogram of key genes was obtained from the above files, and the top 23 genes as shown in the figure, in order, were EGFR, FLT1, PXN, VAV2, MMP3, SPP1, ADAMTS2, AOX1, CHI3L1, CNDP2, COX7B, CTSB, CTSL, DDX18, EFNA1, ELL ENTPD1, FGF7, FGFR3, FHL2, GATM, HEY1, HEY2. GO enrichment analysis of candidate differential genes in OS using R software and correlation analysis confirmed that the candidate differential genes are mainly involved in cellular processes such as osteogenesis, epithelial cell proliferation, osteoblast differentiation and molecular functions such as growth factors and fibronectin binding. This finding is also consistent with biological processes that play a key role in OS development and progression. Li et al. found that the degree of ossification within soft tissue masses in common type osteosarcoma correlated with tumor surgical stage, chemotherapy sensitivity, and prognosis, and that regulation of PLK2 enriched for TAp73 affects osteogenic differentiation and prognosis in human osteosarcoma, also confirming the role of the ossification process in osteosarcoma[13]. Our data suggest that this biological process involves the largest number of genes. Luo et al. found that CDKN2B-AS1 can play an oncogenic role in osteosarcoma by promoting cell proliferation and epithelial-to-mesenchymal transition, suggesting that epithelial cell proliferation is an important part of osteosarcoma development[14]. Our enrichment analysis also confirmed exactly this, and Baumann's team demonstrated that collagen can be post-translationally modified by prolyl and lysyl hydroxylation followed by glycosylation of hydroxylysine, and the experimental conclusion illustrates that complete loss of collagen glycosylation decreases osteosarcoma cell proliferation and viability[15]. In other words, the invasion of bone tissue by tumor cells affects the dynamic balance between bone resorption and bone formation[16]. The bone remodelling process is significant in osteosarcoma development. The role of fibronectin binding in osteosarcoma inhibition is also significant, and some studies have also confirmed that it enhances the apoptotic response of U2-OS and contributes to the fight against osteosarcoma[17]. It is well known that vascular endothelial growth factor A (VEGF) is one of the most important growth factors that regulate vascular development and angiogenesis. VEGF is involved in different stages of bone repair, including the inflammatory phase, endochondral ossification, healing tissue formation and intramembranous ossification during bone remodelling.[18] These biological processes are all involved in our work. These biological processes were confirmed in our analysis, and it is reasonable to speculate that phenotypic studies on osteosarcoma could focus on these aspects.

In addition, candidate differential genes were enriched in the Kyoto Encyclopedia of Genes (KEGG) pathways mainly in adherent spots, Rap1, PI3K-Akt, MAPK, and insulin resistance, and only a few pathways with relatively more enriched genes are listed in this paper for subsequent studies. It has been shown that focal adhesion kinase (FAK) overexpression and phosphorylation may predict more aggressive biological behavior in osteosarcoma and may be an independent predictor of poor prognosis[19] Feng et al. also demonstrated the involvement of FAK in the migration of human osteosarcoma cells[20]. We know that cancer cachexia is easily driven by inflammation, and metabolic alterations (e.g. increased energy expenditure, elevated plasma glucose, insulin resistance and excess catabolism). Insulin insensitivity decreases glucose uptake in organs and leads to loss of skeletal muscle and adipose tissue.[21] Imerb also found in his study of insulin resistance that it leads to impaired skeletal regulation and imbalances in bone homeostasis and pathological bone.[22]. Insulin resistance, a common metabolic feature and a risk factor for many diseases are caused by the inability of insulin to perform its normal metabolic role, as well as by nutritional imbalances and abnormal lipid accumulation in skeletal muscle, liver and adipose tissue.[23] The PI3K/Akt pathway is considered to be one of the most important oncogenic pathways in human cancers. There is increasing evidence that this pathway is frequently over-activated in OS and contributes to disease onset and progression including tumorigenesis, proliferation, invasion, etc.[24]. Certainly, the role of MAPK signaling pathway in the development of osteosarcoma has also been long established[25–27] The role of MAPK signaling pathway in the development of osteosarcoma has also been well established. In addition, the expression levels of HIF-1, Rap1, PI3K and Akt were also found to be elevated in OS cells, suggesting that the Rap1/PI3K-Akt pathway could be a therapeutic target for OS.[28]. In conclusion, our KEGG enrichment results were consistent with previous studies, and most genes were enriched in two pathways, PI3K/Akt and MAPK, suggesting that the future direction of research on osteosarcoma may have to focus on these two pathways.

We constructed a PPI network of candidate differential genes encoding proteins, and then based on the core ranking histogram of key genes, we selected the top-ranked genes such as EGFR, FLT1, VAV2, PXN, and FGF7 for analysis. As shown by the survival curves, the survival curves of the key genes did not show statistical significance, except for FGF7. As seen by the FGF7 survival curve, the FGF7 high expression group had a higher survival rate than the FGF7 low expression group until 100 months of survival. The rest of the Hub genes failed to show differences. Thus, we speculate that FGF7 may serve as a marker to predict survival outcomes in OS patients, which can be focused on in later validation experiments. However, it has also been shown that FGF7 induces the proliferation of osteosarcoma cells and accelerates the secretion of EMT and inflammatory mediators[29]. This is contrary to our conclusion, so the follow-up study should make FGF7 a focus and increase the sample size appropriately for validation again. As for Epidermal growth factor receptor (EGFR), it can activate various signaling pathways such as PI3K/Akt, Jak/STAT, etc. Targeted regulation of EGFR expression can inhibit the development of osteosarcoma.[30]. Therefore, combined with survival analysis, we believe that although EGFR cannot predict the survival of patients, it can be a potential therapeutic target. Tsuchida, while studying the molecular mechanism of tumor progression after chemotherapy, demonstrated that cisplatin (CDDP) upregulates FLT1 expression, leading to the survival and proliferation of highly oncogenic side population (SP) cells in cell lines such as osteosarcoma[31] Yin et al. slowed the growth of osteosarcoma tumors in this mouse model by FLT1 gene modification.[32]. Therefore, FLT1 could also be an intervention target for osteosarcoma. It has been shown that VAV2 is associated with the migration and invasion of sarcoma cells[33]. PXN, a focal adhesion-associated protein, significantly promotes signaling between protein networks[34]. There is evidence that PXN may be associated with cell proliferation and migration[35]. However, PXN is currently poorly studied in osteosarcoma, suggesting that we should subsequently deepen our research on PXN as a therapeutic target in osteosarcoma. Also based on the analysis of gene expression and clinical correlation, we should also pay sufficient attention to the gender and age of tumor tissue origin. The comparison of promoter methylation level analysis suggests that we should also take TP53 mutation typing into full consideration.

Finally, we performed immune cell infiltration analysis of the tissues and found that infiltration of M2 macrophages, unactivated dendritic cells, and neutrophils accounted for a greater proportion of OS tissues, while infiltration of naive B cells and unactivated CD4 + memory T cells was statistically less. This suggests that we should focus on M2 macrophages in osteosarcoma research to inhibit tumor cell development by modulating immune infiltrating cells in the osteosarcoma microenvironment. In the correlation analysis of core genes with immune and infiltrating cells, key genes were found to be generally more correlated with naive B cells. On the correlation analysis between core genes and immune checkpoints, three immune checkpoints, ADAMTS2, AOX1, and CHI3L1, were found to be the most correlated with core genes. These two correlation analyses also provided us with the next direction in osteosarcoma research.

The identification of differential genes and further bioinformatics analysis have been carried out in the past few years, and these findings may provide new ideas for the study of the mechanisms of OS onset, development, metastasis, invasion, and drug resistance. However, a limitation of this study is that it was not validated with clinical specimens, which may need to be added in future studies.

Conclusion

Studying this dataset helps us to further understand the interactions between OS-related candidate DEGs. These findings may help us to deepen our overall understanding of the molecular mechanisms of osteosarcoma and the factors influencing osteosarcoma. However, further relevant molecular biology experiments are needed to confirm the functions of the identified OS-related genes.

Declarations

Acknowledgements

Not applicable.

Funding

This study was supported by the Natural Science Foundation of Ningxia Province (Grant No.: 2022AAC03594), Natural Science Foundation of Ningxia Province (Grant No.: 2019AAC03193).

Availability of data and materials

The GSE14359 dataset's expression profiles were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/).

Authors' contributions

Jiangdang Shi were involved in the conception and design of this study. Ruiqing Xu collected the data and performed the bioinformatics analyses. Ruiqing Xu and Ningjun Wan prepared the figures and interpreted the data. Ruiqing Xu drafted the manuscript. Jiangdang Shi and Ningjun Wan revised the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Kumar R, Kumar M, Malhotra K, et al. Primary Osteosarcoma in the Elderly Revisited: Current Concepts in Diagnosis and Treatment[J]. Curr Oncol Rep, 2018,20(2):13.
2. Singh K, Mukherjee A B, De Vouge M W, et al. Differential processing of osteopontin transcripts in rat kidney- and osteoblast-derived cell lines[J]. J Biol Chem, 1992,267(33):23847-23851.
3. Ottaviani G, Jaffe N. The epidemiology of osteosarcoma [J]. Cancer Treat Res, 2009,152:3-13.
4. Simpson S, Dunning M D, de Brot S, et al. Comparative review of human and canine osteosarcoma: morphology, epidemiology, prognosis, treatment and genetics[J]. Acta Vet Scand, 2017,59(1):71.
5. Selmic L E, Burton J H, Thamm D H, et al. Comparison of carboplatin and doxorubicin-based chemotherapy protocols in 470 dogs after amputation for treatment of appendicular osteosarcoma[J]. J Vet Intern Med, 2014,28(2):554-563.
6. Siegel R L, Miller K D, Fuchs H E, et al. Cancer Statistics, 2021[J]. CA Cancer J Clin, 2021,71(1):7-33.
7. Liu Q, Wang K. The induction of ferroptosis by impairing STAT3/Nrf2/GPx4 signaling enhances the sensitivity of osteosarcoma cells to cisplatin[J]. Cell Biol Int, 2019,43(11):1245-1256.
8. Stiller C A, Bielack S S, Jundt G, et al. Bone tumours in European children and adolescents, 1978-1997. report from the Automated Childhood Cancer Information System project[J]. Eur J Cancer, 2006,42(13):2124-2135.
9. Kansara M, Teng M W, Smyth M J, et al. Translational biology of osteosarcoma [J]. Nat Rev Cancer, 2014,14(11):722-735.
10. Ritter J, Bielack S S. Osteosarcoma [J]. Ann Oncol, 2010,21 Suppl 7:i320-i325.
11. Chen X, Bahrami A, Pappo A, et al. Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma[J]. Cell Rep, 2014,7(1):104-112.
12. Czarnecka A M, Synoradzki K, Firlej W, et al. Molecular Biology of Osteosarcoma [J]. Cancers (Basel), 2020,12(8).
13. Li W, Zhang X, Xi X, et al. PLK2 modulation of enriched TAp73 affects osteogenic differentiation and prognosis in human osteosarcoma[J]. Cancer Med, 2020,9(12):4371-4385.
14. Luo Y, Tao H, Jin L, et al. CDKN2B-AS1 Exerts Oncogenic Role in Osteosarcoma by Promoting Cell Proliferation and Epithelial to Mesenchymal Transition[J]. . Cancer Biother Radiopharm, 2020,35(1):58-65.
15. Baumann S, Hennet T. Collagen Accumulation in Osteosarcoma Cells lacking GLT25D1 Collagen Galactosyltransferase[J]. J Biol Chem, 2016,291(35):18514-18524.
16. Lamora A, Talbot J, Mullard M, et al. TGF-beta Signaling in Bone Remodeling and Osteosarcoma Progression[J]. J Clin Med, 2016,5(11).
17. Incesu Z, Hatipoglu I, Sivas H, et al. Effects of fibronectin and type IV collagen on osteosarcoma cell apoptosis[J]. Indian J Exp Biol, 2013,51(10):789-796.
18. Hu K, Olsen B R. The roles of vascular endothelial growth factor in bone repair and regeneration[J]. Bone, 2016,91:30-38.
19. Ren K, Lu X, Yao N, et al. Focal adhesion kinase overexpression and its impact on human osteosarcoma[J]. Oncotarget, 2015,6(31):31085-31103.
20. Feng S, Shi X, Ren K E, et al. Focal adhesion kinase is involved in the migration of human osteosarcoma cells[J]. Oncol Lett, 2015,9(6):2670-2674.

21. Masi T, Patel B M. Altered glucose metabolism and insulin resistance in cancer-induced cachexia: a sweet poison[J]. Pharmacol Rep, 2021,73(1):17-30.
22. Imerb N, Thonusin C, Chattipakorn N, et al. Aging, obese-insulin resistance, and bone remodeling [J]. Mech Ageing Dev, 2020,191:111335.
23. Li A, Qiu M, Zhou H, et al. PTEN, Insulin Resistance and Cancer[J]. Curr Pharm Des, 2017,23(25):3667-3676.
24. Zhang J, Yu X H, Yan Y G, et al. PI3K/Akt signaling in osteosarcoma[J]. Clin Chim Acta, 2015,444:182-192.
25. Tsai H C, Cheng S P, Han C K, et al. Resistin enhances angiogenesis in osteosarcoma via the MAPK signaling pathway[J]. Aging (Albany NY), 2019,11(21):9767-9777.
26. Gui Z L, Wu T L, Zhao G C, et al. MicroRNA-497 suppress osteosarcoma by targeting MAPK/Erk pathway[J]. Bratisl Lek Listy, 2017,118(8):449-452.
27. Lv H, Zhen C, Liu J, et al. beta-Phenethyl Isothiocyanate Induces Cell Death in Human Osteosarcoma through Altering Iron Metabolism, Disturbing the Redox Balance, and Activating the MAPK Signaling Pathway[J]. Oxid Med Cell Longev, 2020,2020:5021983.
28. Li Z Q, Wang Z, Zhang Y, et al. CircRNA_103801 accelerates proliferation of osteosarcoma cells by sponging miR-338-3p and regulating HIF-1/Rap1/PI3K- Akt pathway[J]. J Biol Regul Homeost Agents, 2021,35(3):1021-1028.
29. Han G, Guo Q, Ma N, et al. LncRNA BCRT1 facilitates osteosarcoma progression via regulating miR-1303/FGF7 axis[J]. Aging (Albany NY), 2021,13(11):15501-15510.
30. Wang J, Wang G, Li B, et al. miR-141-3p is a key negative regulator of the EGFR pathway in osteosarcoma[J]. Onco Targets Ther, 2018,11:4461-4478.
31. Tsuchida R, Das B, Yeger H, et al. Cisplatin treatment increases survival and expansion of a highly tumorigenic side-population fraction by upregulating VEGF/Flt1 autocrine signaling[J]. Oncogene, 2008,27(28):3923-3934.
32. Yin D, Jia T, Gong W, et al. VEGF blockade decelerates the growth of a murine experimental osteosarcoma[J]. Int J Oncol, 2008,33(2):253-259.
33. Tie S R, McCarthy D J, Kendrick T S, et al. Regulation of sarcoma cell migration, invasion and invadopodia formation by AFAP1L1 through a phosphotyrosine- dependent pathway[J]. Oncogene, 2016,35(16):2098-2111.
34. Turner C E, West K A, Brown M C. Paxillin-ARF GAP signaling and the cytoskeleton[J]. Curr Opin Cell Biol, 2001,13(5):593-599.
35. Huang C, Jacobson K, Schaller M D. A role for JNK-paxillin signaling in cell migration[J]. Cell Cycle, 2004,3(1):4-6.

Figures

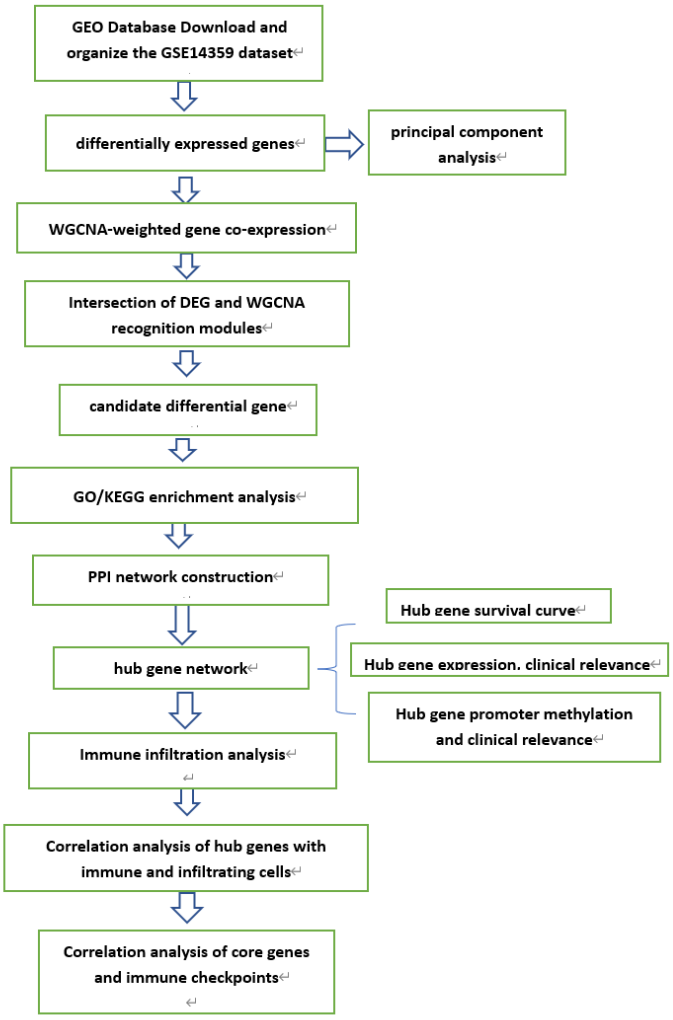

Figure 1

flow chart

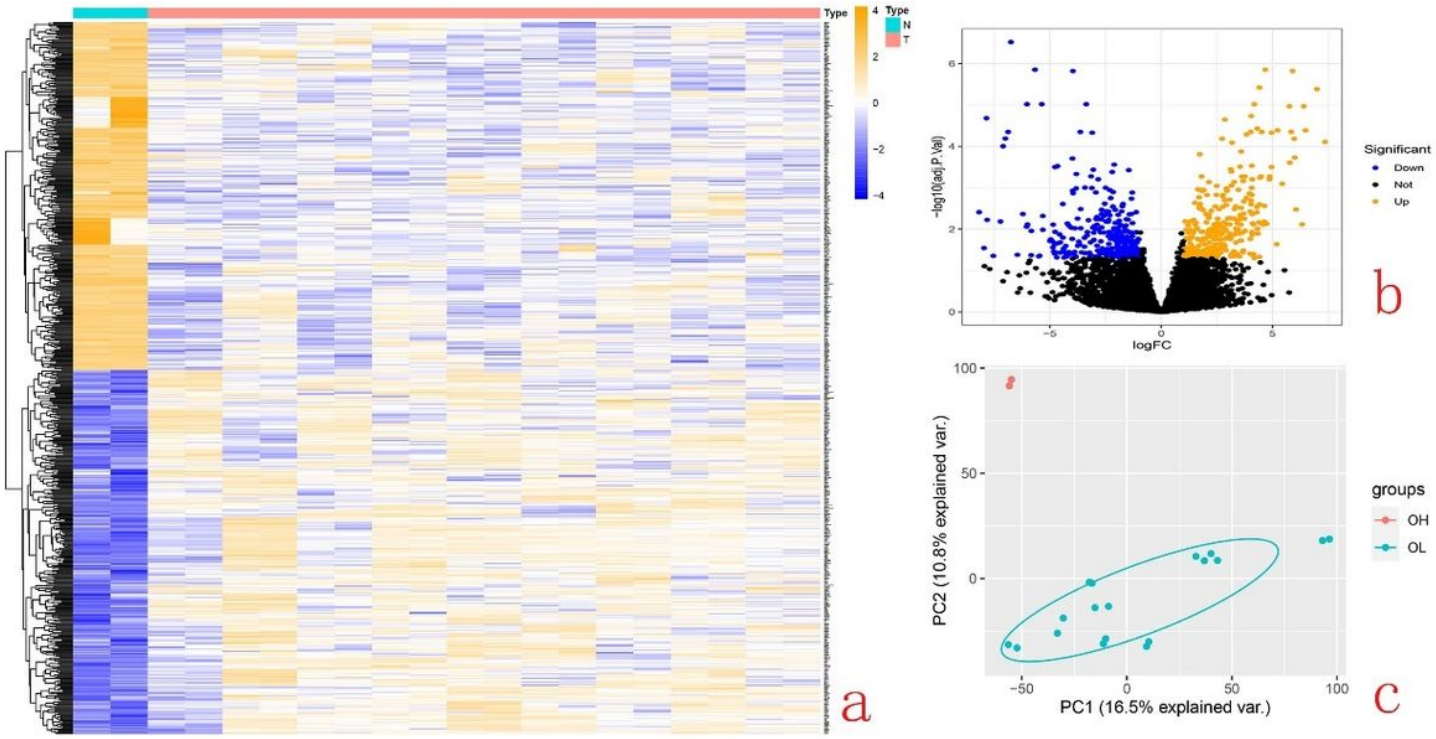

Figure 2

a) Heat map of differential gene expression between OS and healthy groups; b) Volcano map of differentially expressed genes between OS and healthy groups. Black dots indicate insignificant expression differences, adjusted P>0.05. Green dots represent down-regulated genes, adjusted P<0.05. Red dots represent up-regulated genes, adjusted P<0.05. c) Principal component analysis between two groups; OH indicates normal group; OL indicates tumor group.

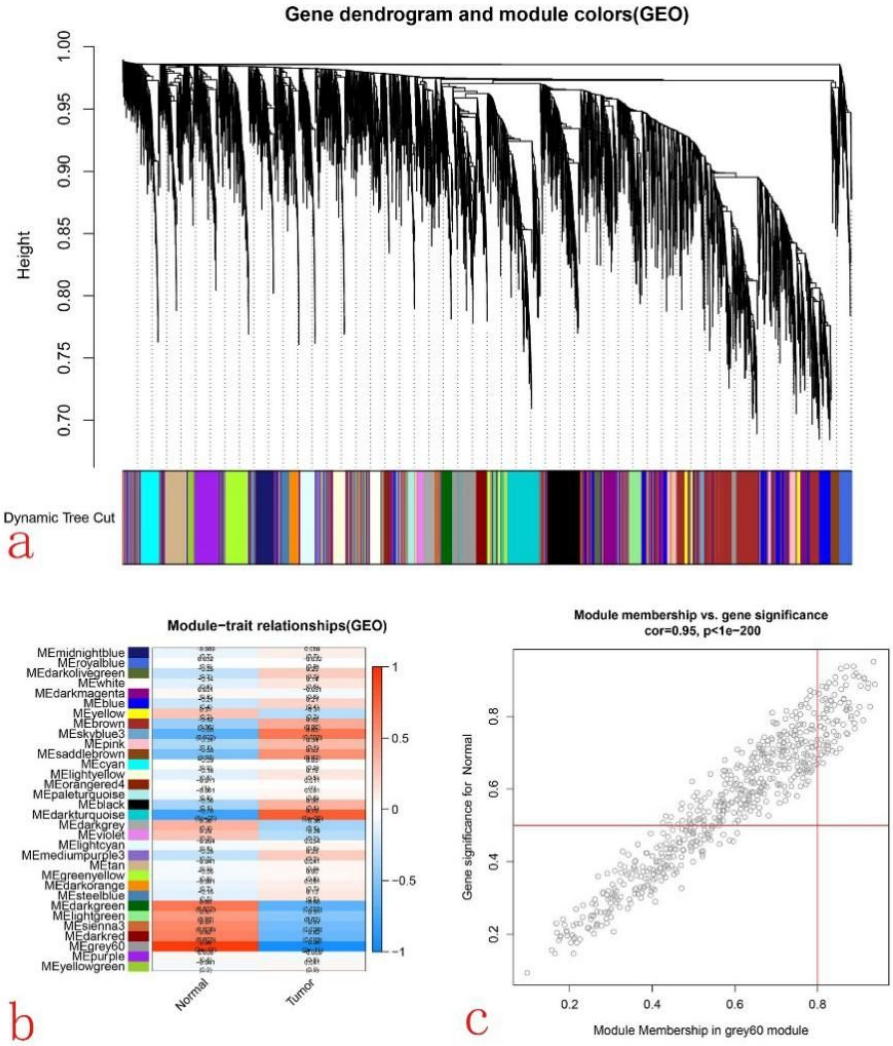

Figure 3

WGCNA-weighted gene co-expression network analysis. a,b,c) The 32 modules identified

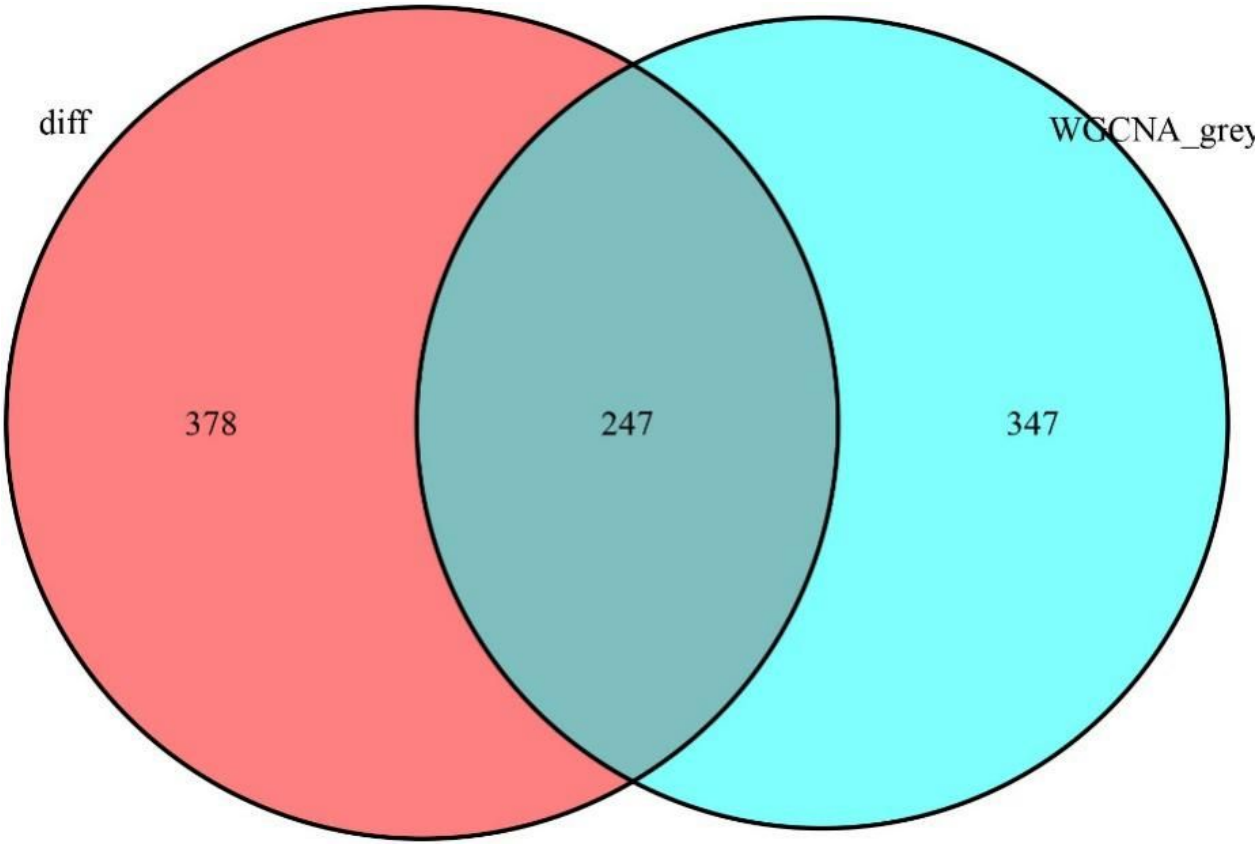

Figure 4

Venn diagram of DEGs and MEgrey60 modules taking the intersection

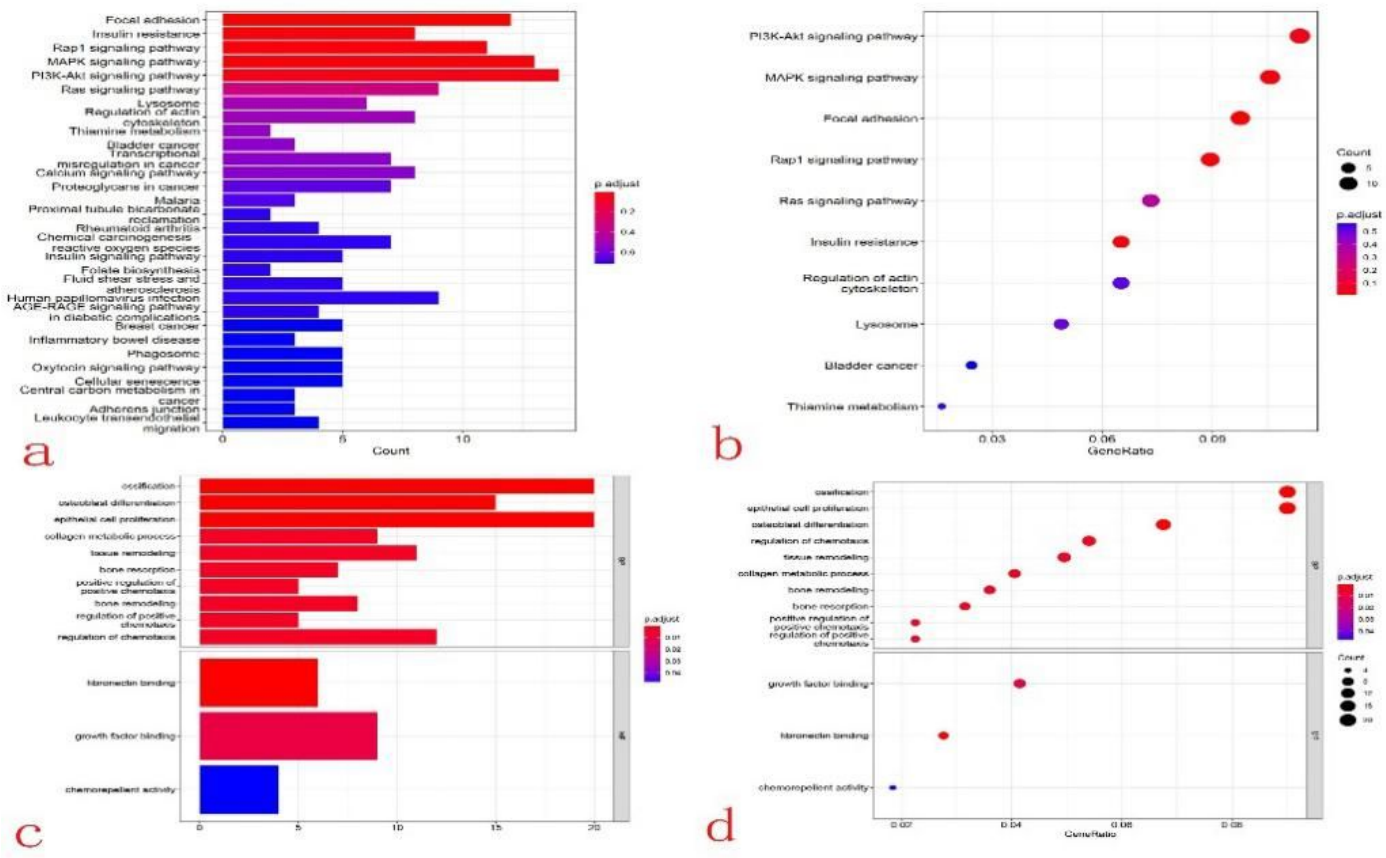

Figure 5

a,b) KEGG enrichment results of the candidate differential genes; c,d) GO enrichment results of the candidate differential genes.

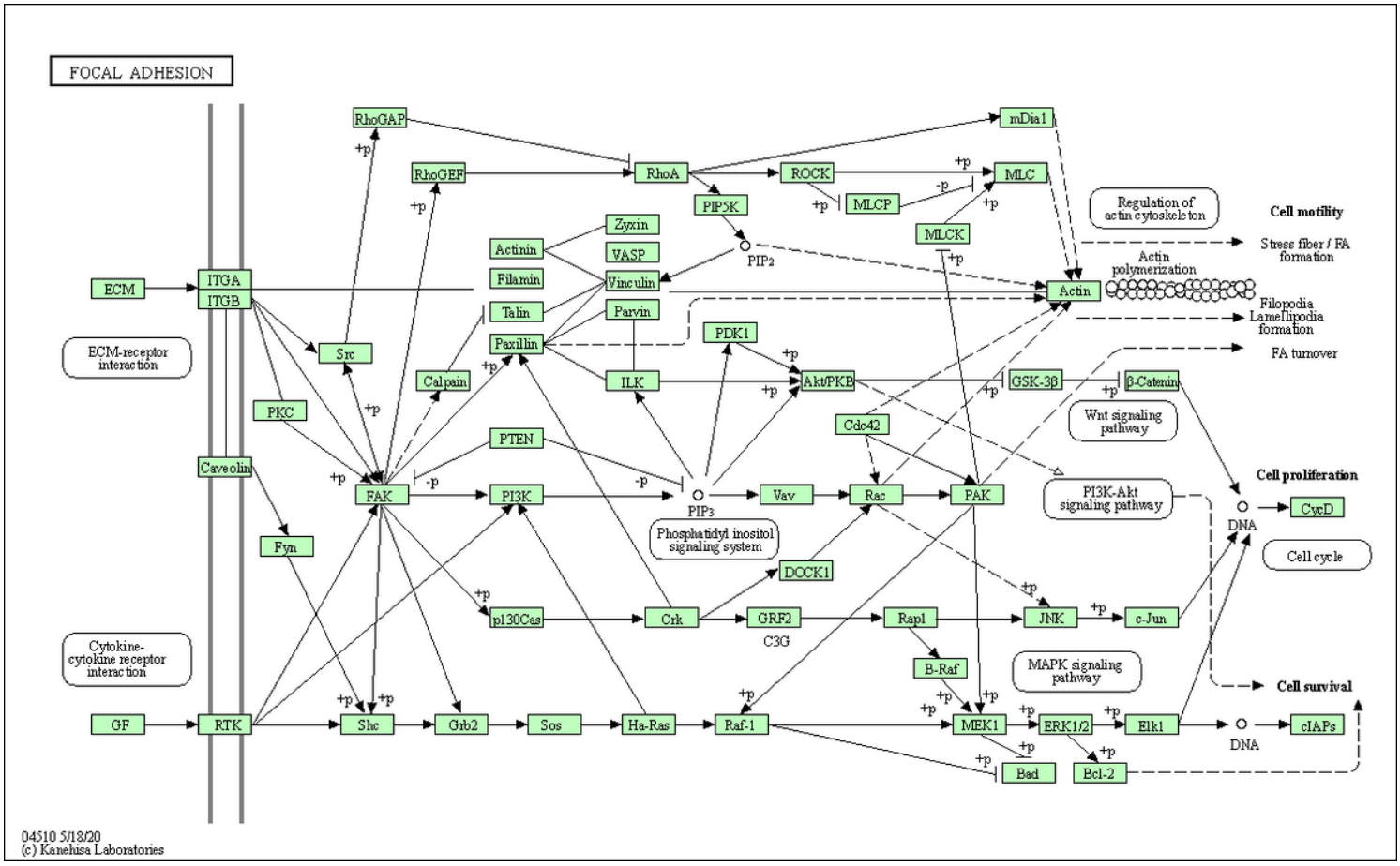

Figure 6

FOCAL ADHESION pathway diagram

Figure 7

MAPK pathway diagram

Figure 8

PI3K-Akt pathway diagram

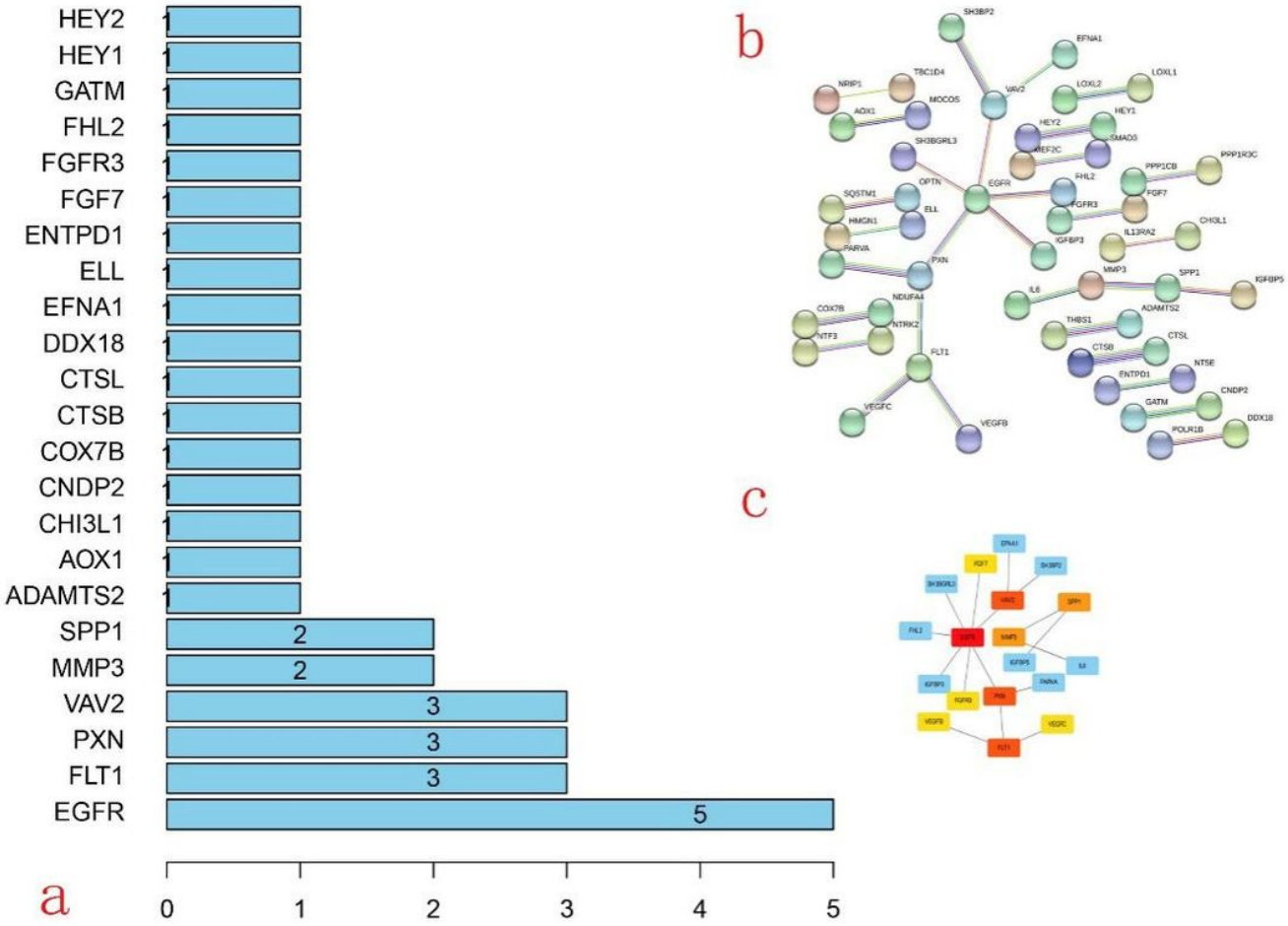

Figure 9

a) Key gene ranking histogram (CYTOSCAPE software); b) PPI network; c) hub gene network map

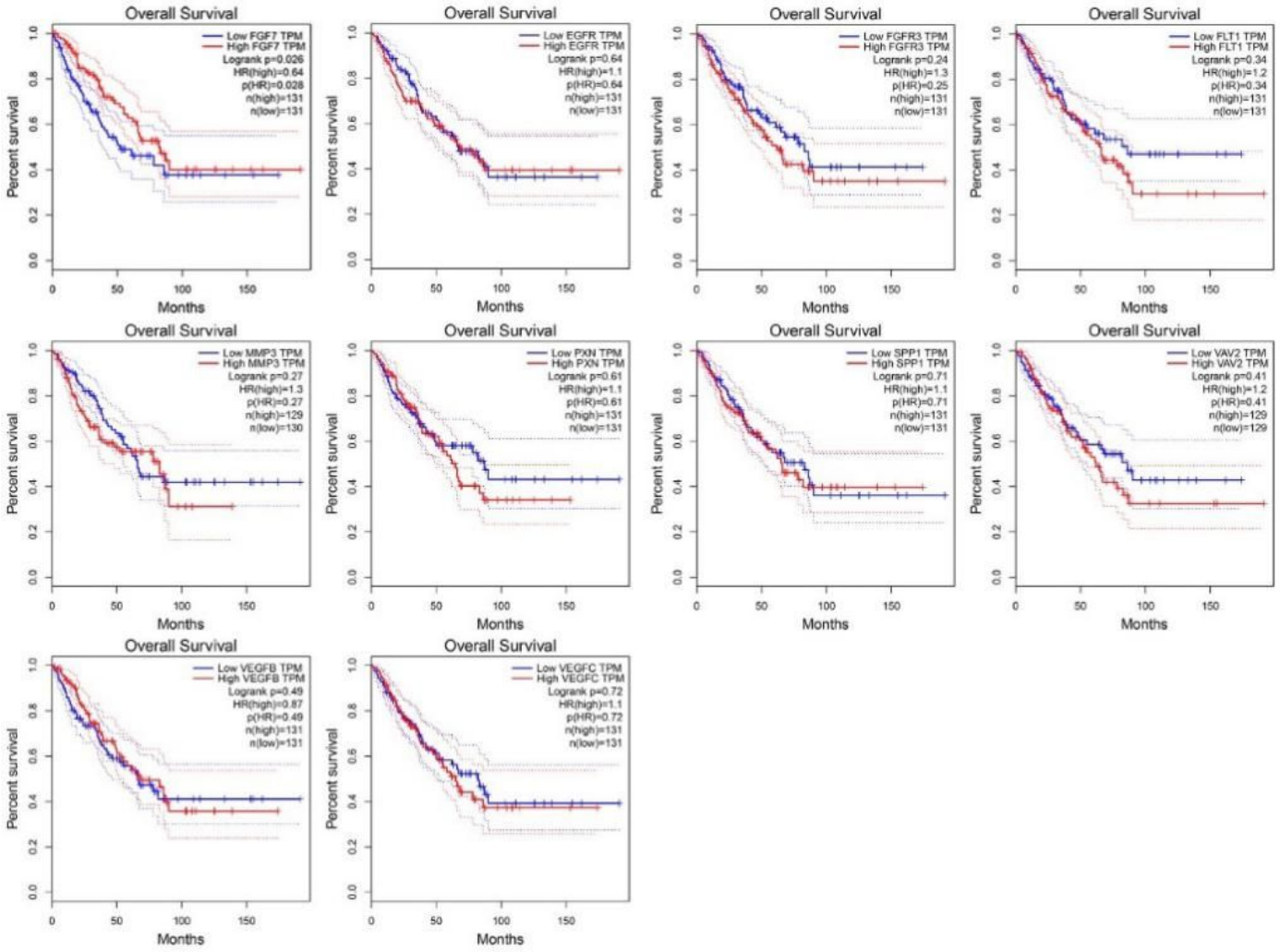

Figure 10

Survival curves of the 10 hub genes (no statistical difference except for FGF7).

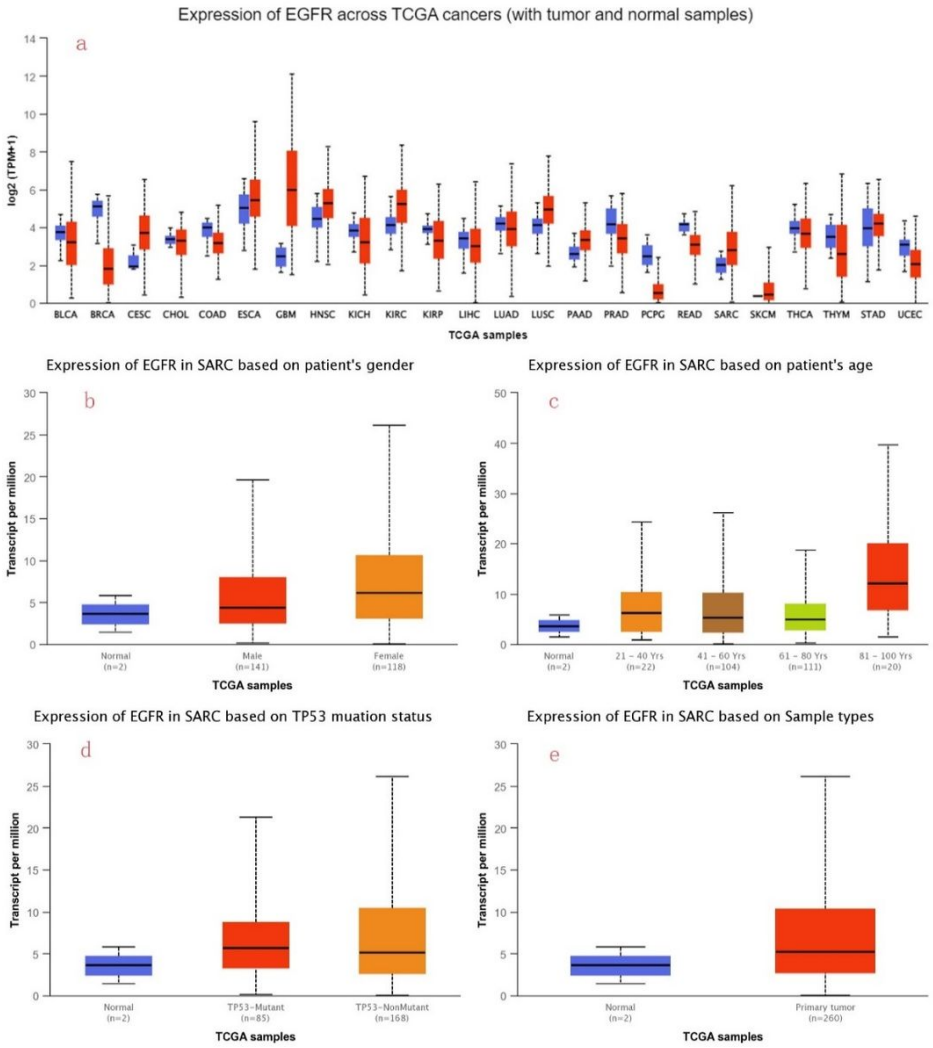

Figure 11

a) EGFR expression differences in pan-cancer; b) EGFR expression in sarcoma with gender correlation; c) EGFR expression in sarcoma with age correlation; d) EGFR expression in sarcoma with TP53 mutation correlation; e) EGFR expression in sarcoma with primary nature correlation.

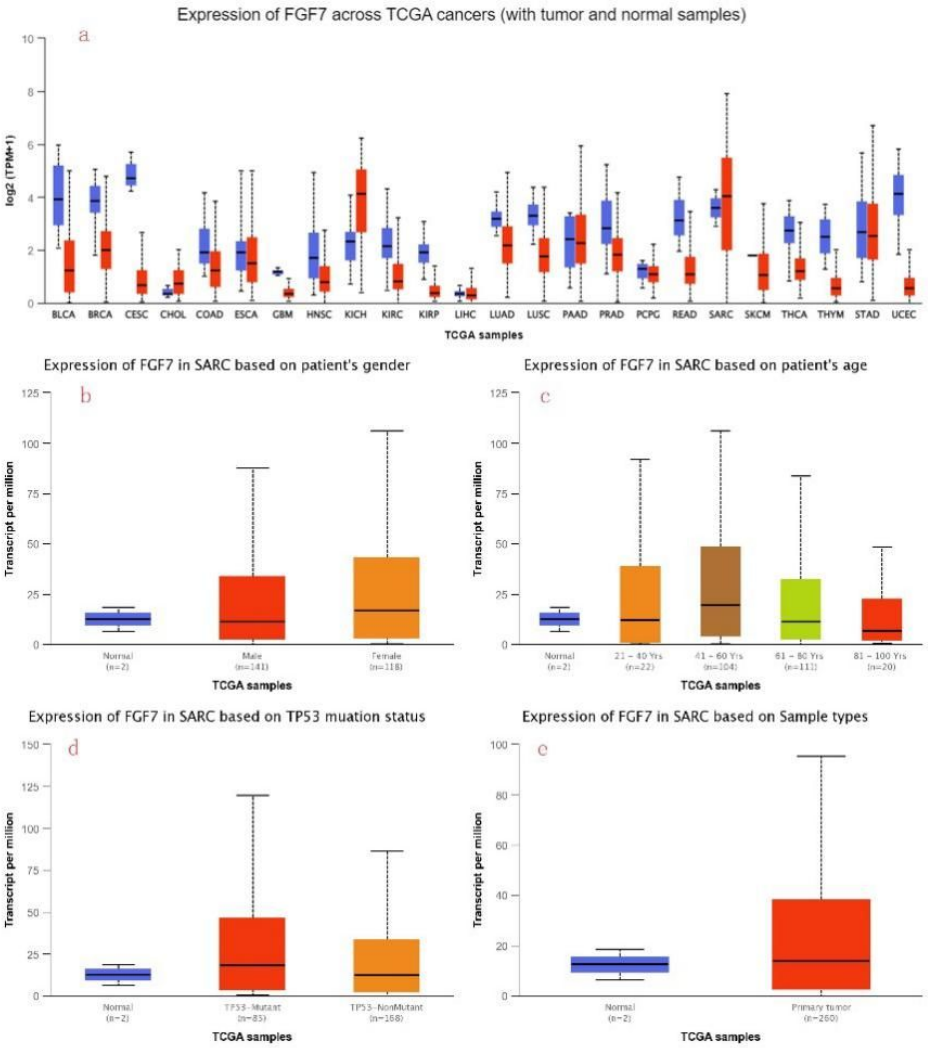

Figure 12

a) Differential expression of FGF7 in pan carcinoma; b) Correlation of FGF7 expression in sarcoma with gender; c) Correlation of FGF7 expression in sarcoma with age; d) Correlation of FGF7 expression in sarcoma with TP53 mutation; e) Correlation of FGF7 expression in sarcoma with primary nature.

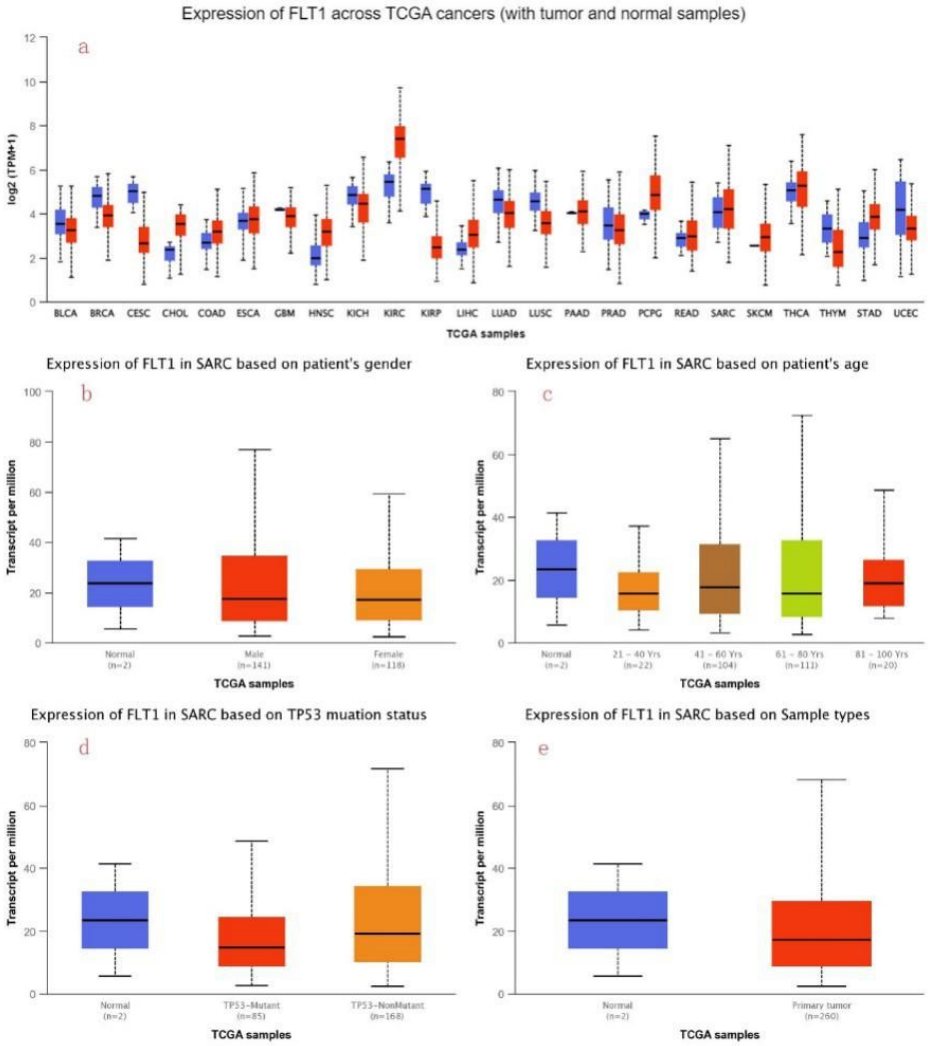

Figure 13

a) Differential expression of FLT1 in pan carcinoma; b) Correlation of FLT1 expression in sarcoma with sex; c) Correlation of FLT1 expression in sarcoma with age; d) Correlation of FLT1 expression in sarcoma with TP53 mutation; e) Correlation of FLT1 expression in sarcoma with primary nature.

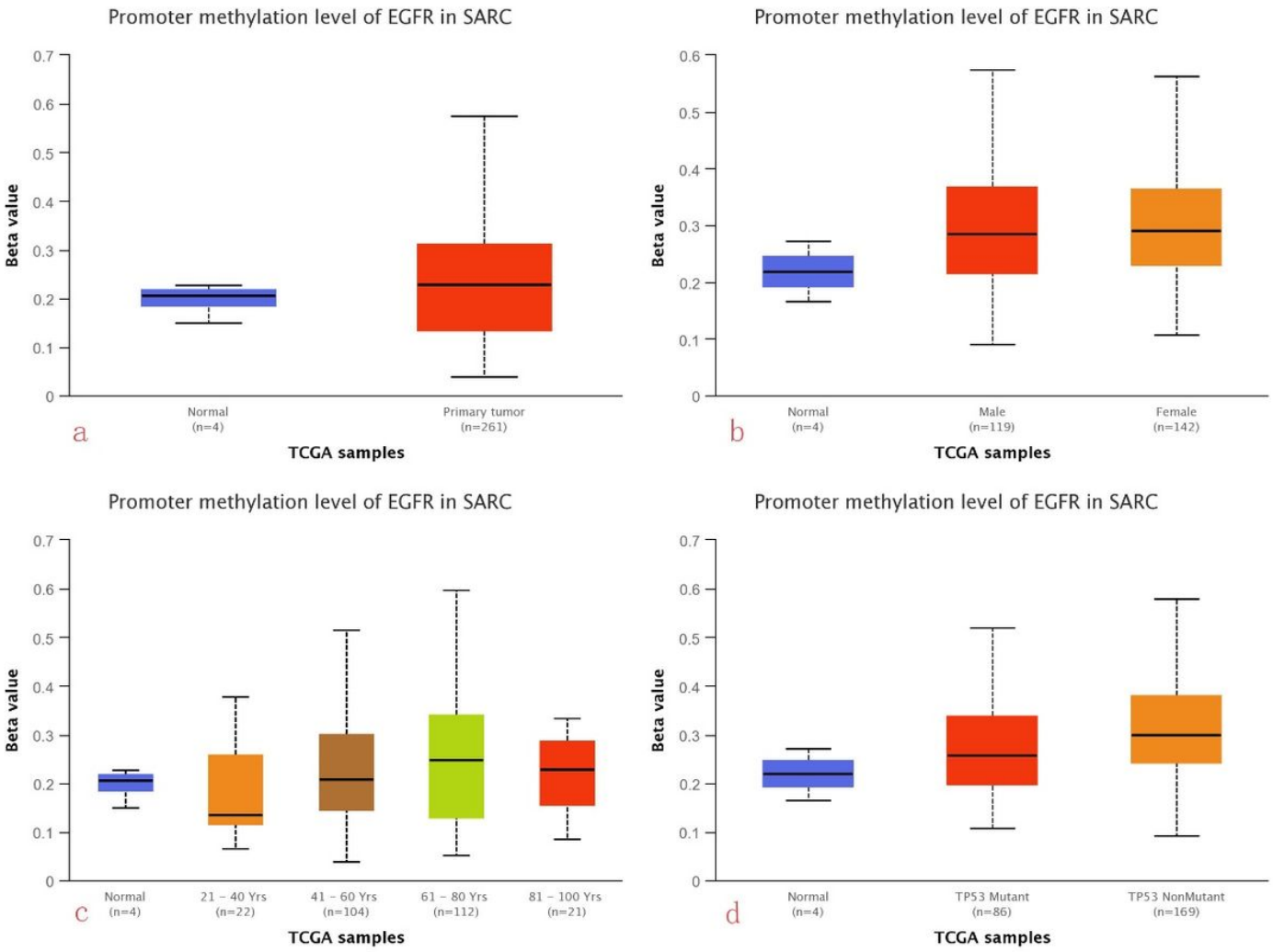

Figure 14

a) Correlation of EGFR promoter methylation levels in sarcoma with primary cancer; b) Correlation of EGFR promoter methylation levels in sarcoma with gender; c) Correlation of EGFR promoter methylation levels in sarcoma with age; d) Correlation of EGFR promoter methylation levels in sarcoma with TP53 mutations.

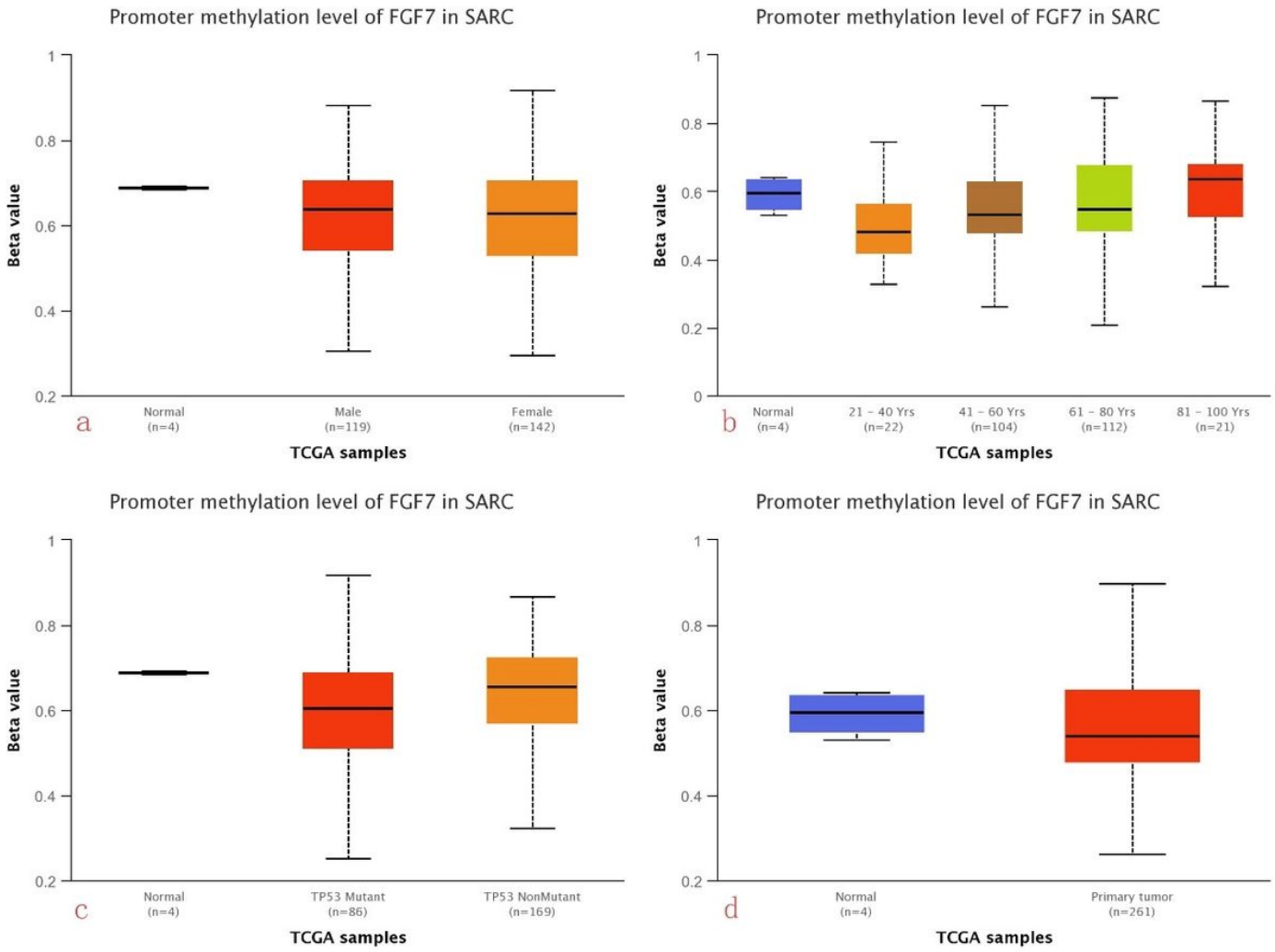

Figure 15

a) Correlation of promoter methylation levels of FGF7 in sarcoma with sex; b) Correlation of promoter methylation levels of FGF7 in sarcoma with age; c) Correlation of promoter methylation levels of FGF7 in sarcoma with TP53 mutations; d) Correlation of promoter methylation levels of FGF7 in sarcoma with primary cancer.

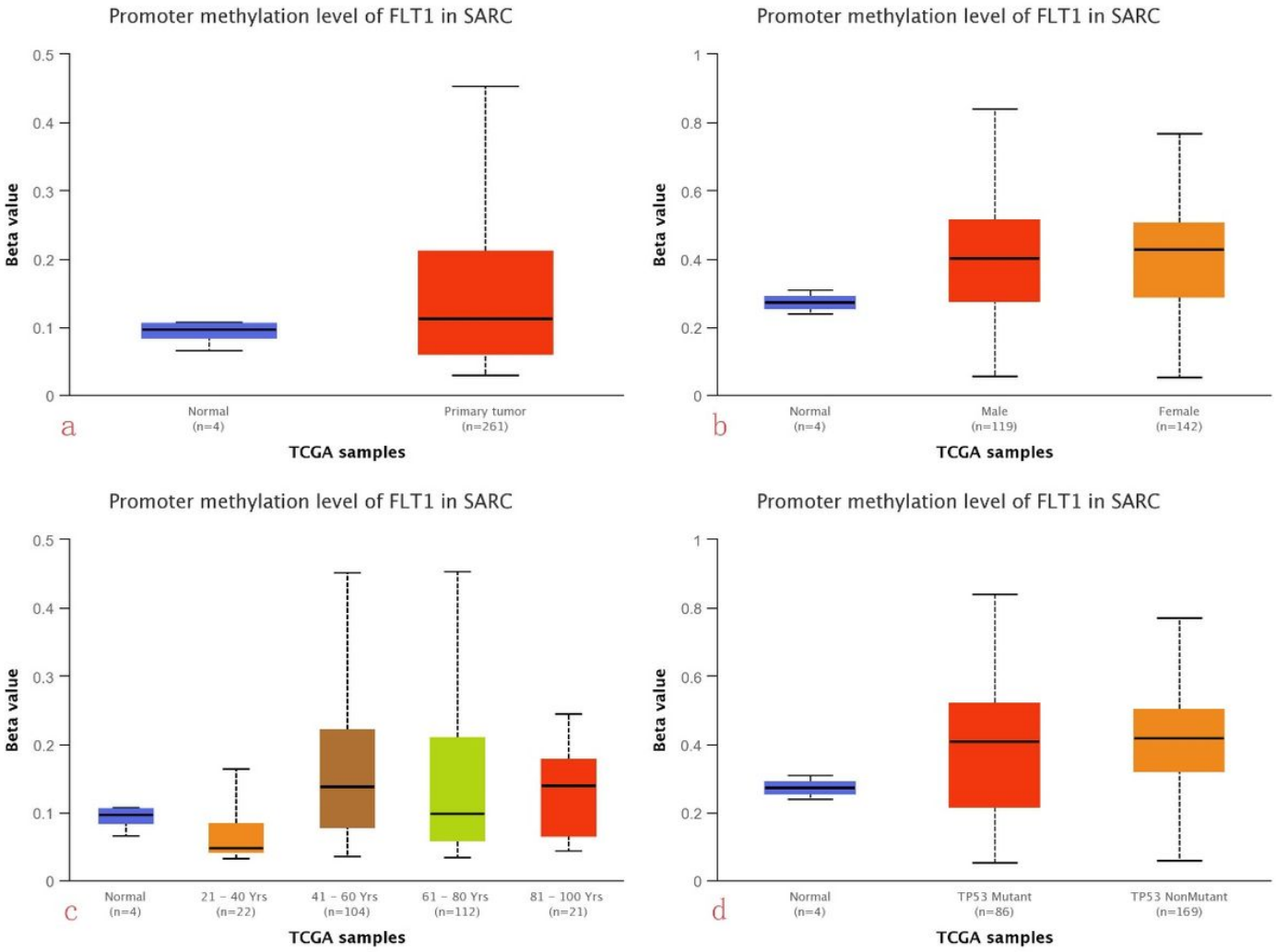

Figure 16

a) Correlation of FLT1 promoter methylation levels in sarcoma with primary cancer; b) Correlation of FLT1 promoter methylation levels in sarcoma with sex; c) Correlation of FLT1 promoter methylation levels in sarcoma with age; d) Correlation of FLT1 promoter methylation levels in sarcoma with TP53 mutations.

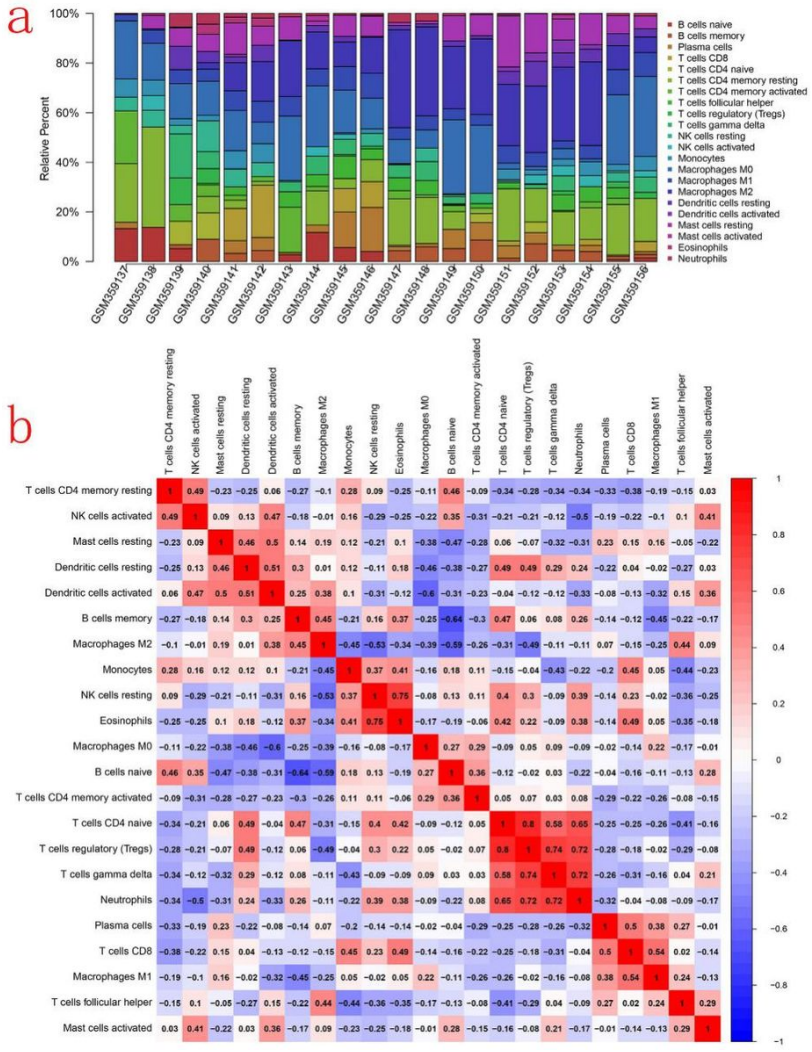

Figure 17

) Bar graph of the percentage of 22 immune infiltrating cells in 20 samples (18 osteosarcoma samples, 2 normal tissues); b) correlation between the 22 immune infiltrating cells.

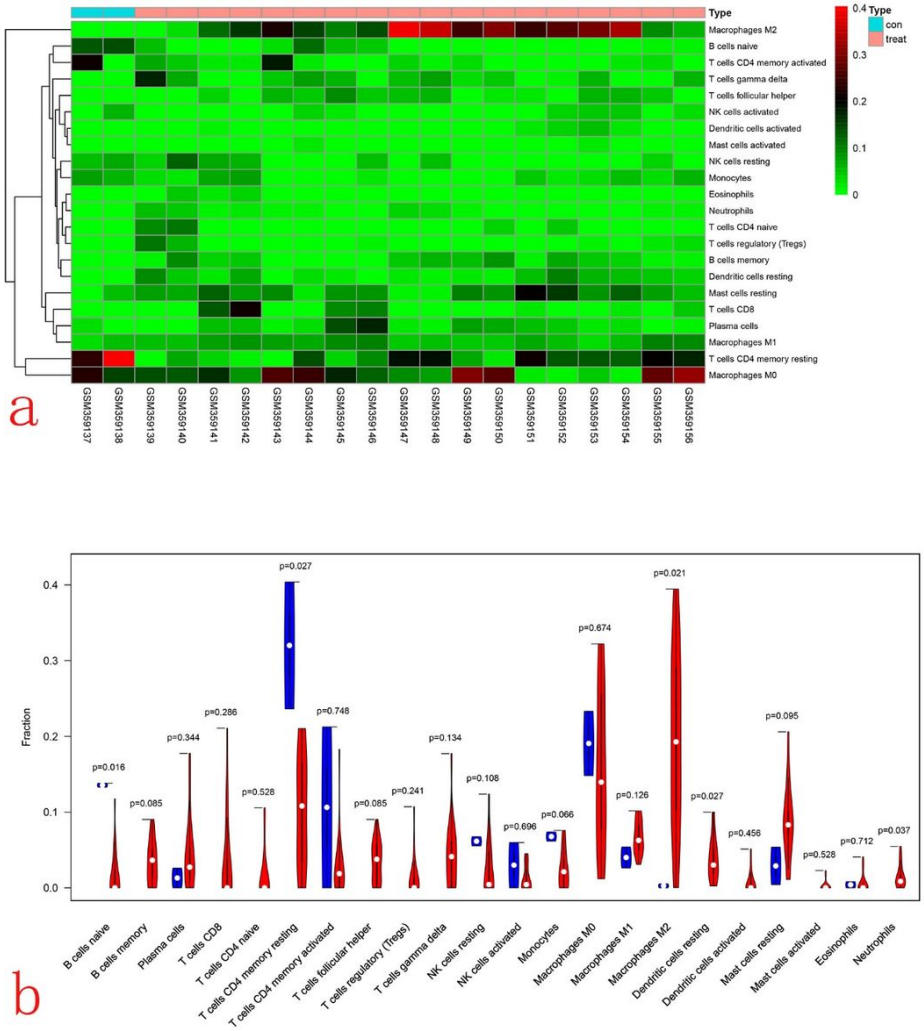

Figure 18

a) Heat map visualizing the level of infiltration of each immune cell in each sample; con indicates normal group; treat indicates tumor group; b) This violin plot shows the level of each immune infiltrating cell in the normal and OS groups.

Figure 19

correlation of hub genes with each immune infiltrating cell.

Figure 20

correlation of hub genes with each immune infiltrating cell.

Figure 21

correlation of hub genes with each immune infiltrating cell.

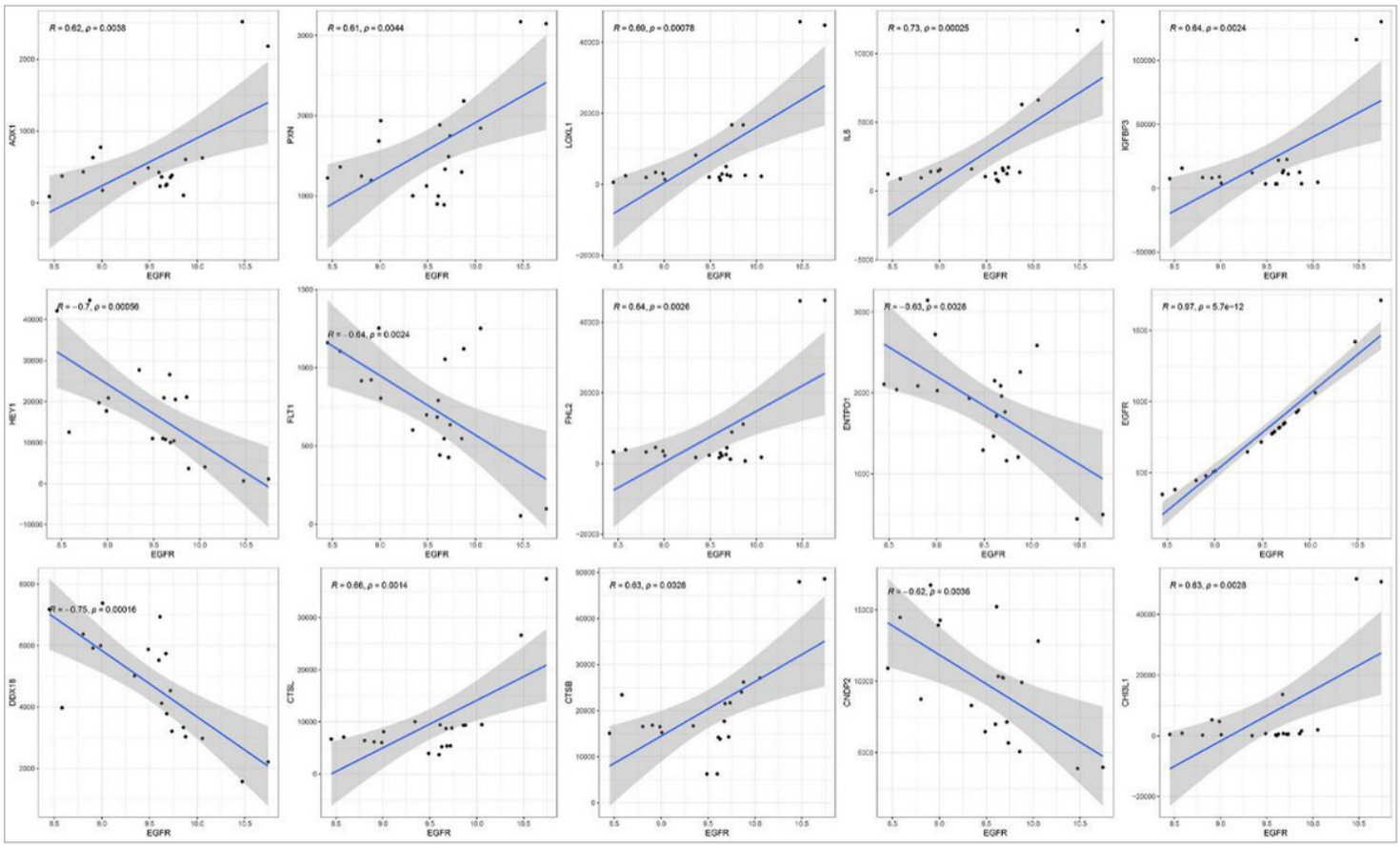

Figure 22

correlation of hub genes with each immune checkpoint.

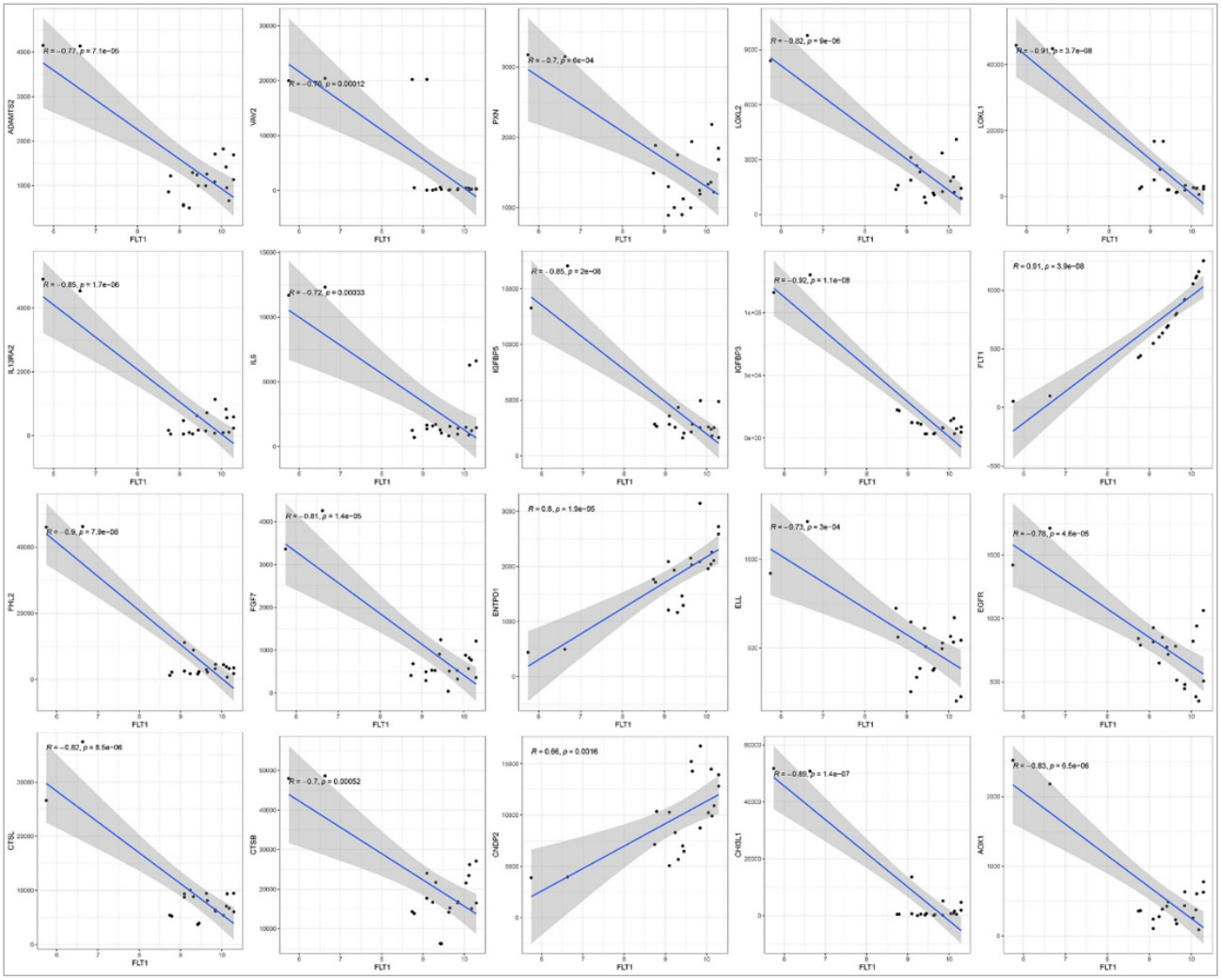

Figure 23

correlation of hub genes with each immune checkpoint.