

Mining TCGA to Reveal Immunotherapy-related Genes for Soft Tissue Sarcoma

Ruixin Li

First Hospital of China Medical University

Fan Yao (✉ yaofancmu@sina.com)

First Hospital of China Medical University

Yijin Liu

First Hospital of China Medical University

Xiaodan Wu

First Hospital of China Medical University

Peng Su

First Hospital of China Medical University

Tianran Li

the Fourth Medical Center of Chinese PLA General Hospital

Research Article

Keywords: TCGA, immune scores, tumor microenvironment, immune-related gene

Posted Date: December 1st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1107348/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Objective: By mining the TCGA database to look for immunotherapy targets of soft tissue sarcoma, and analyzed their biological behavior.

Methods: The data of 265 samples were downloaded from the TCGA database to analyze the expression profile of soft tissue sarcomas. Research methods include immune and stromal scores, calculating DEGs, volcano maps and differential gene survival curves, gene enrichment analysis.

Results: Kaplan-Meier survival curves showed that in the high immune score group, the total survival time was generally higher than that in the low immune score group. Analysis of the top ten terms resulted in the minimum P values for immune and inflammatory responses, plasma membrane, receptor activity, and chemokine activity. By plotting the K-M curves, we obtained 86 survivals related DEGs. Finally, the genes that can be used as independent risk factors for prognosis of soft tissue sarcoma were obtained by multivariate analysis of the DEGs.

Conclusion: We believe that these genes are expected to be new targets for sarcoma immunotherapy and key genes for the analysis of prognosis of sarcoma.

Introduction

Soft tissue sarcoma is a malignant tumor whose etiology is still unknown. Its classification is variable, and based on its different sources, its histological and biological characteristics also differ. The treatment of soft tissue sarcoma depends on surgery, radiotherapy and chemotherapy, and it is necessary to select high-risk patients for preoperative radiotherapy and chemotherapy in clinical studies. However, the classification of soft tissue sarcomas is complex, and conventional chemotherapy regimens are not effective for all types of tumors (Hall, K.S. et al., 2020). Numerous studies (Francescutti, V. and Skitzki, J. J.,2012; Maki, R. G.,2001) have shown that soft tissue sarcomas are highly correlated with the immune system and can promote tumor growth by inhibiting immune infiltration in the tumor microenvironment. The relationship between the relevant genes and the tumor microenvironment has been extensively studied (Fan, T. *et al.*,2020; Hu, C., Chen, B., *et al.*,2020; Zhu, N. and Hou, J.,2020;Barretina,J.*et al.*,2010), and it has been demonstrated that the tumor microenvironment can affect gene expression in tumor tissues, and the high expression of some genes in tumor tissues can lead to changes in the tumor microenvironment(Zhang, B. *et al.*,2021; Hu, C., Chen, B., *et al.*,2020;Li,F.*et al.*,2020;Toda, Y. *et al.*,2020;Huang,R.*et al.*,2019). For example, undifferentiated pleomorphic sarcoma and leiomyosarcoma have been shown to highly express antigen presenting genes PD-1 and PD-L1, as well as T-cell infiltration-related genes. (Toda, Y. *et al.*, 2020; Toulmonde, M. *et al.*, 2018; Pollack, S. M. *et al.*, 2017) The strong inflammatory expression in the immune microenvironment makes it possible to treat sarcomas by inducing an immune response. (Burgess, M. *et al.*, 2015; Mesiano, G. *et al.*, 2015; Francescutti, V. and Skitzki, J. J., 2012)

The ESTIMATE algorithm based on the TCGA database can be used to evaluate the purity of the tumor. (Fan, T. *et al.*, 2020; Jia, D. *et al.*, 2018) Based on gene expression data, it involves the numbers of immune cells and stromal cells whose expression is the highest in the tumor microenvironment. Thus, we can directly study the genes that have the highest impact with the immune and stromal scores, and the value of these genes in predicting prognosis and survival rate of soft tissue sarcoma can be evaluated. (Li, F. *et al.*, 2020; LAN, H. *et al.*, 2017; Peille, A.-L. *et al.*, 2013)

In this study, the ESTIMATE algorithm was employed to analyze data from the sarcoma group in the TCGA database and collect the immune, stromal, and estimate scores. These three scores were combined to identify genes associated with the prognosis of soft tissue sarcomas. We also confirmed that a part of these differentially expressed genes (DEGs) can be used as independent risk factors for the prediction of prognosis of soft tissue sarcoma. Through a subsequent confirmation based on the Gene Expression Omnibus (GEO) database, we obtained 12 genes that are meaningful both in the TCGA and GEO database.

Methods

Immune and stromal scores: The data of a total of 265 samples were downloaded from the TCGA database to analyze the expression profile of soft tissue sarcomas. The gene expression profile of soft tissue sarcoma was roughly divided into seven subtypes according to Widemann's (Widemann, B. C. and Italiano, A., 2018) reported. The gene expression data of 265 cases downloaded from the TCGA database also. Gene expression data used to calculate immune and stromal scores were derived from TCGA, as were clinical data such as overall survival, age at first diagnosis, and residual tumor data. The independent dataset used to verify differential genes was derived from the GEO database, GSE30929 dataset (SEP 02, 2011).

Methods for calculating DEGs: Data analysis using the "limma" package analysis, set Fold change = 2 and $p < 0.05$.

Graphics: Volcano maps and differential gene survival curves were produced using the tool SangerBox. Kaplan-Meier curves of immunological and stromal scores and sarcoma typing scores were obtained using the tool GraphPad. The Venn diagram was drawn with the Web tool "Venny". Construction of Protein-Protein Interaction Network (PPI): The overall PPI network was derived from the STRING database and analyzed by the tool "Cytoscape". Molecular Complex Detection (MCODE) was used to divide the network into three modules, and only the genes with more than 5 edges were selected as the key genes.

Gene enrichment analysis: Results of gene enrichment analysis were obtained from "David" (The Database for Annotation, Visualization and Integrated Discovery), including GO analysis results of biological processes (BP), molecular functions (MF), and cellular components (CC), and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways, and false discovery rate (FDR) < 0.05 was set as the cut-off.

Statistical analysis methods: Univariate and multivariate regression analyses were conducted using Cox analysis and Kaplan-Meier survival analysis in tool Statistical Product and Service Solutions (SPSS), and only genes with significant < 0.05 in univariate and multivariate analyses were retained.

Results

265 cases soft tissue sarcomas classification: There were 59 cases of dedifferentiated liposarcoma (22.3%), 2 cases of desmoid tumor (0.75%), and 52 cases of undifferentiated pleomorphic sarcoma (UPS)(including pleomorphic, malignant fibrous histiocytoma (MFH) giant cell MFH/undifferentiated pleomorphic sarcoma with giant cells, undifferentiated pleomorphic sarcoma) (19.6%), 107 cases of leiomyosarcoma (LMS) (40.4%), 10 cases of malignant peripheral nerve sheath (MPNST) (3.77%), 25 cases of myxofibrosarcoma (9.43%), and 10 cases of synovial sarcoma (including synovial sarcoma-biphasic, synovial sarcoma- monophasic, synovial sarcoma- poorly differentiated) (3.77%).

Immune and stromal scores

The range of estimate scores based on the ESTIMATE algorithm was from -2445.3107 to 13152.619. Stromal scores ranged from -740.7589 to 6339.22. The range of immune scores was from -1890.07 to 7914.021. The stromal score of dedifferentiated liposarcoma was the highest among the eight subtypes (Among all seven subtypes, only two desmoid tumor samples were included; thus, they were not included for comparison.). This was followed by myxofibrosarcoma, UPS, MPNST, and LMS. The synovial sarcoma subtype had the lowest stromal score ($P < 0.0001$). In addition, the immune scores were ranked from the highest to lowest as follows: UPS>myxofibrosarcoma > dedifferentiated liposarcoma > MPNST > LMS > synovial sarcoma ($P < 0.0001$), indicating that immune and stromal scores are important for the classification of different soft tissue sarcoma subtypes.

The gene expression data were divided into two groups—high expression and low expression groups—based on the median of the immune and stromal score. Kaplan-Meier survival curves were drawn, which showed that in the high immune score group, the total survival time was generally higher than that in the low immune score group (see Fig.1 A). The median of overall survival times in the high immune score group and the low immune score group were 2464 days and 1722 days, respectively (log-rank test $P < 0.05$). Similarly, although the result was not statistically significant (the log-rank test $P > 0.05$), the survival curve obviously indicated that the total survival time of the group with high stromal scores was higher than that of the group with low stromal scores, with a median total survival time of 2448 days for the high score group compared with 1722 days for the low score group (see Fig.1 B).

The sarcoma samples downloaded from the TCGA database were divided into seven types—dedifferentiated liposarcoma, desmoid tumor, UPS, LMS, MPNST, myxofibrosarcoma, and synovial sarcoma—and a box plot (Fig.1 C,D) drawn to compare different types of immunity and stromal scores, and there were statistically significant($P < 0.0001$), indicating that the immune and stromal scores were significantly different for the classification of sarcoma.

Differentially expressed genes results

To compare the difference in gene expression between the high score and low score groups, we used the “limma” algorithm, with the setting, $\log_2(\text{fc}) = 2$ and $\text{FDR} = 0.05$, to calculate the DEGs for the low stromal, immune, and estimate score groups relative to the high group. Volcano plots were drawn to indicate differences in gene expression among cases with different immune, stromal, and estimate scores. In the charts (see Fig.2 A, B, C), green represents downregulated genes, and red represents upregulated genes. In the immune score group, 91 genes were upregulated and 479 genes were downregulated. In the stromal score group, 157 genes were upregulated and 281 genes were downregulated. In the estimate score group, 100 genes were upregulated and 482 were downregulated. A Venn diagram was drawn to obtain 258 genes with common differences in all three groups, among which 66 genes were upregulated and 192 genes were downregulated (see Fig.2 D,E).

Functional enrichment analysis of DEGs

A total of 18 cellular components (CC), 26 molecular functions (MF), and 95 biological processes (BP) were obtained by GO analysis of 258 DEGs. Analysis of the top ten terms resulted in the minimum P values for immune and inflammatory responses, plasma membrane, receptor activity, and chemokine activity. Similarly, the results of KEGG pathway analysis were also mainly related to immunity, among which the highest proportion included *Staphylococcus aureus* infection, accounting for 29.17%, Leishman disease, accounting for 25%, and viral protein interaction with cytokine and cytokine receptor, accounting for 12.5% (see Fig.3).

Relationship between DEGs and overall survival results

To explore the different roles of 258 DEGs in survival, we set the median as the boundary to draw a Kaplan–Meier survival curve for each of the DEGs. Eighty-six genes were statistically significant predictors of overall survival ($\log\text{-rank } P < 0.05$) (see Fig.4).

Construction of a protein co-expression network results

A total of 15 key node genes were selected because they were the most closely related to other genes around. Cluster 1 had 14 nodes and 39 edges; the key nodes were LILRB2, AOA, CCL5, NCF4(Ryan, B. M. *et al.*,2014), CYBB. Cluster2 had a total of 8 nodes and 16 edges; the key nodes were ITGAM, CD2, IL2RG, CD163. Cluster 3 had 16 nodes and 33 edges; the key nodes were VAV1, CXCR3, HCK, MS4A6A, LY86, IDO1. Enrichment analysis was performed on these key genes, and the GO analysis results were as follows:

- positive regulation of T-cell apoptotic process
- negative regulation of T-cell apoptotic process
- inflammatory response

The KEGG results included

- chemokine signaling pathway
- leukocyte transendothelial migration

These results are all associated with the immune system. (Only results of $P < 0.05$ are included) (see Fig.5).

Multivariate regression analysis of DEGs results (see Tab.1)

Table 1. Univariate regression analysis of DEGs

Group	Significant Value	HR	Confidence Interval 95.0%	
age_at_initial_pathologic_diagnosis	0.000	1.082	1.035	1.130
residual_tumor	0.015	3.536	1.279	9.777
ADAP2	0.012	5.874	1.480	23.307
CSF1R	0.001	0.042	0.007	0.266
CYBB	0.003	10.208	2.159	48.262
MPEG1	0.049	0.253	0.064	0.996
NCF4	0.000	69.966	8.793	556.733
TBXAS1	0.019	4.680	1.282	17.084
C13orf33	0.008	0.608	0.421	0.878
GPR34	0.044	3.385	1.034	11.079
ITGAM	0.019	0.165	0.037	0.741
DOK2	0.030	0.219	0.056	0.862
CD163	0.015	0.186	0.048	0.726
TLR8	0.041	2.663	1.043	6.799
LYVE1	0.001	2.269	1.415	3.638
BATF	0.000	0.121	0.045	0.323
P2RY13	0.030	0.434	0.204	0.922
SLAMF6	0.037	3.636	1.080	12.236
S100A9	0.000	0.373	0.220	0.630
CD38	0.023	0.412	0.192	0.884
FCER2	0.033	0.669	0.463	0.968
CADM3	0.013	0.823	0.706	0.959
WISP2	0.049	1.232	1.001	1.516
KLHL23	0.004	2.201	1.279	3.789
MYH11	0.000	0.565	0.448	0.712

Using clinical data from TCGA database, univariate regression analysis was performed to obtain "age at initial pathologic diagnosis "and "residual cancer" as prognostic factors for sarcoma. Multivariate regression analysis of the differential genes was carried out to obtain a gene list that could predict the prognosis of sarcoma, including risk genes (HR>1) and benign genes (HR<1).

Verification of DEGs in the GEO database results

To verify the DEGs obtained are also significant in other datasets, we used the 140 independent liposarcoma samples of GSE30929 in the GEO database to perform survival analysis by calculating the forward relapse-free survival (DFS). We obtained 12 genes that are meaningful both in the TCGA and GEO database: CLEC10A(Kurze, A.-K. *et al.*,2019), IDO1(Nafia, I. *et al.*,2020), FCER2(Sharma, V. *et al.*,2014), CCL13(Mendez-Enriquez, E. and García-Zepeda, E. A.,2013), KLRB1(Zhang, G. *et al.*,2020; Gentles, A. J. *et al.*,2015), P2RY13(Fan, T. *et al.*,2020), PYHIN1(Massa, D. *et al.*,2020; Tong, Y., Song, Y. and Deng, S.,2019), DOK2(Ohsugi, T.,2017), PLA2G2D(Miki, Y. *et al.*,2016; Miki, Y. *et al.*,2013), MARCO(Xiao, Y. *et al.*,2019), MYH11(Hu, C., Chen, B., *et al.*,2020), NCF4(Ryan, B. M. *et al.*,2014). Nine of the genes are important because they have not previously been reported to be associated with the prognosis of sarcoma, including CLEC10A, FCER2, CCL13, KLRB1, P2RY13, PYHIN1, PLA2G2D, MARCO, NCF4.

Discussion

Soft tissue sarcoma is a kind of tumor with complicated classification and high malignancy, and its immunotherapy has been a hot topic. To identify relevant targets for immunotherapy for soft tissue sarcomas, we performed a series of data analyses using samples from the TCGA database. By calculating the immune and stromal scores, we confirmed that the immunological/stromal scores differed among the different types of soft tissue sarcomas. As shown in previous studies, UPS expressed more T-cell permeation-related genes, while the synovial sarcoma expressed significantly less. Therefore, UPS is a kind of soft tissue sarcoma suitable for immunotherapy. (Pollack, *et al.*,2017; Toulmonde, M.,*et al.*,2020) And the overall survival time was longer in the high-immune group than in the low-immune group. Thus, the high expression of some immune-related genes was beneficial to prolong the overall survival time.

To further understand which genes affected immune scores mostly, we performed differential analysis and obtained differential genes. Our results showed that most genes were downregulated when the low score group was compared to the high score group. According to the Venn diagrams, we chose 258 genes which were differential expression in three scores. And these differential genes were analyzed by functional enrichment. Top ten results of GO analysis suggested that these DEGs had immune-related functions. And so it was in the KEGG analysis. As we know, the sarcoma tumor cells inhibit the immune response in the tumor microenvironment to promote its growth and metastasis. Therefore, some treatments use trace amounts of *Staphylococcus aureus* infection to stimulate the immune response of the body to treat sarcoma. (Toulmonde, M. *et al.*,2018; Burgess, M. *et al.* 2015) People with low immunity are susceptible to infection with the *Leishmania* parasite and often develop with Kaposi's sarcoma. Kaposi's sarcoma, which is highly prevalent in AIDS patients, is greatly reduced with the use of cocktails of therapy. It shows that the occurrence and development of sarcoma is directly related to the stability of the patient's immune internal environment. (Burgess, M. *et al.* ,2015) Similarly, as in previous studies, we also identified immune-related pathways such as NK-kappa B signaling and Natural killer cell mediated cytotoxicity.(Hu, C. *et al.*, 2020)

By plotting survival curves for individual genes, we disproved the obtained relationship between immune/stromal scores and overall survival time. AOA, for example, is a downregulated gene; i.e., AOA gene expression was low in the low score group compared with that in the high score group. The higher the expression of AOA, the higher the corresponding immune/stromal score, and the longer the overall survival time. This corresponded to a longer overall survival in the high immune/stromal rating group. There were totally 86 survival related DEGs. Similarly, we performed enrichment analyses on 86 DEGs; the outcomes included 51 biological processes, 9 cellular components, and 11 biological processes that were identified as significant. The top terms were immune and inflammatory responses, NADPH oxidase complex (Martner, A., Aydin, E. and Hellstrand, K.,2019), integral component of plasma membrane, superoxide-generating NADPH oxidase activator activity, and superoxide-generating NADPH oxidase activity.

In further study, we constructed a protein expression network of 86 genes, which was divided into four modules, among which the main three modules included 15 key gene nodes because they are most closely related to the surrounding genes, including LILRB2, AOA, CCL5, NCF4, CYBB, ITGAM, CD2, IL2RG, CD163, VAV1, CXCR3, HCK, MS4A6A, LY86, IDO1. And we found that these genes all had immune-related functions by functional enrichment analysis. We confirmed that these DEGs were significant in immunotherapy. Among them, IDO1 has been proven to be related to the growth and immunosuppression of various tumors (Nafia, I. *et al.*,2020;Toulmonde, M. *et al.*,2020;Toda, Y. *et al.*,2020), and in previous studies, CCL5 has shown the ability to recruit CD 8 T cells, leading to a better prognosis(Morales, E. *et al.*,2020) which verified the result of our study.

To confirm that DEGs can be used as prognostic indicators, Cox regression analysis was performed on 86 DEGs that were significant for survival, and significant genes were obtained ($P < 0.05$), including 10 genes that were indicators of malignancy—ADAP2, CYBB, NCF4(Ryan, B. M. *et al.*,2014), TBXAS1, GPR34, TLR8, LYVE1, SLAMF6, WISP2, and KLHL23 ($HR > 1$)—and 13 genes ($HR < 1$) that were good indicators—CSF1R, MPEG1, C13orf33, ITGAM, DOK2(Ohsumi, T.,2017), CD163, BATF, P2RY13 (Fan, T. *et al.*,2020), S100A9, CD38, FCER2(Sharma, V. *et al.*,2014), CADM3, and MYH11. NCF4, CYBB, ITGAM, CD163 are also key nodes in PPI. A cluster analysis confirmed the importance of the MYH11-ADM regulatory network in STS patients (Hu, C., Chen, B., *et al.*,2020), which further proves the correctness of this study.

Finally, we used an independent group of 140 samples from the GEO database to verify the roles of these genes in survival and confirmed that 12 genes were meaningful in this independent group. nine of the genes are involved in immunity and have been reported to be involved in the development or metastasis of different types of tumors; while, none of them has been reported to be associated with the development and progression of sarcomas. This means that these genes are likely to play a role in the treatment of sarcoma as immunotherapeutic targets. Compared with the sarcoma-related mutant genes previously revealed by gene sequencing, this study identified other genes and related pathways previously unrelated to the disease from the perspective of tumor microenvironment. (Barretina, J. *et al.* ,2010) We believe that these genes are expected to be new targets for sarcoma immunotherapy and key genes for

the analysis of prognosis of sarcoma. This will address a key step in tumor immunotherapy, and effective tumor cell surface antigens can be developed so that immune agents can attack tumor cells directly.

Among them, CLEC10A aroused our interest. CLEC10A is a specific marker of CD1c DC. (Heger, L. *et al.*,2018) When it binds to the receptor, it can enhance the secretion of TNF+, IL-8, and IL-10 induced by TLR 7/8 stimulation (Kurze, A.-K. *et al.*,2019; Heger, L. *et al.*,2018), thereby promoting humoral immunity and inhibiting tumor progression. It has been shown that dendritic cells are independent risk factors for recurrence of soft tissue sarcomas and are directly related to tumor immunity. (Huang, R. *et al.*,2019) In addition, CLEC10A can be used as a marker of damaged and dead cells. After binding with ligands, CLEC10A is preferentially internalized by macrophages to attack tumor cells from the perspective of cellular immunity. In the treatment of breast cancer, we found that female hormone has an inhibitory effect on the CLEC10A ligand, thus demonstrating the induction of the CLEC10A ligand by the estrogen receptor antagonist tamoxifen (Kurze, A.-K. *et al.*,2019), which inspires us to explore endocrine therapy for sarcoma.

In conclusion, we believe that these genes are expected to be new targets for sarcoma immunotherapy and key genes for the analysis of prognosis of sarcoma. To provide gene targets for immunotherapy of soft tissue sarcoma.

Declarations

Acknowledgments

The work was supported by the First Hospital of China Medical University, China. We also acknowledgment TCGA and GEO databases for providing the platform and data to support our research.

Conflict of interest

The authors declare that they have no competing interests.

Author Contributions

Ruixin Li: original draft, formal analysis, visualization, conceptualization; Yijin Liu: original draft, data presentation; Fan Yao: review and editing, funding acquisition; Xiaodan Wu: data curation; Nan Wu: reproducibility of results; Peng Su: supervision, project administration

Data Availability Statement

The data that support the findings of this study are openly available in UCSC Xena (<https://xenabrowser.net/datapages/>) and Gene Expression Omnibus database [GEO GSE30929] <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi> .

Ethics approval and consent to participate and for publication

TCGA database and GEO database are public platforms, and all patient-related information has obtained ethical approval. The databases provide users with free data for relevant research and publication. Our research is entirely based on public data and does not involve ethics or conflicts of interest.

Funding

This study was not funded.

References

1. Burgess, M., Gorantla, V., Weiss, K. *et al.* (2015) 'Immunotherapy in Sarcoma: Future Horizons', *Current Oncology Reports*, 17(11), p. 52. doi: 1007/s11912-015-0476-7.
2. Barretina, J., Taylor, B.S., Banerji, S. *et al.* (2010) 'Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy', *Nature Genetics*, 42(8), pp. 715–721. doi: 1038/ng.619.
3. Francescutti, V. and Skitzki, J. J. (2012) 'Sarcomas and the immune system: implications for therapeutic strategies', *Surgical Oncology Clinics of North America*, 21(2), pp. 341–355. doi: 1016/j.soc.2011.11.002.
4. Fan, T., Zhu, M., Wang, L. *et al.* (2020) 'Immune profile of the tumor microenvironment and the identification of a four-gene signature for lung adenocarcinoma', *Aging*, 12. doi: 18632/aging.202269.
5. Gentles, A.J., Newman, A.M., Liu, C.L. *et al.* (2015) 'The prognostic landscape of genes and infiltrating immune cells across human cancers', *Nature Medicine*, 21(8), pp. 938–945. doi: 1038/nm.3909.
6. Heger, L., Balk, S., Lühr, J.J. *et al.* (2018) 'CLEC10A Is a Specific Marker for Human CD1c+ Dendritic Cells and Enhances Their Toll-Like Receptor 7/8-Induced Cytokine Secretion', *Frontiers in Immunology*, 9, p. 744. doi: 3389/fimmu.2018.00744.
7. Huang, R., Meng, T., Chen, R. *et al.* (2019) 'The construction and analysis of tumor-infiltrating immune cell and ceRNA networks in recurrent soft tissue sarcoma', *Aging (Albany NY)*, 11(22), pp. 10116–10143. doi: 18632/aging.102424.
8. Hall, K.S., Bruland, Ø.S., Bjerkehagen, B. *et al.* (2020) 'Preoperative accelerated radiotherapy combined with chemotherapy in a defined cohort of patients with high risk soft tissue sarcoma: a Scandinavian Sarcoma Group study', *Clinical Sarcoma Research*, 10(1), p. 22. doi: 1186/s13569-020-00145-5.
9. Hu, C., Chen, B., Huang, Z. *et al.* (2020) 'Comprehensive profiling of immune-related genes in soft tissue sarcoma patients', *Journal of Translational Medicine*, 18. doi: 1186/s12967-020-02512-8.
10. Hu, C., Wang, Y., Liu, C. *et al.* (2020) 'Systematic Profiling of Alternative Splicing for Sarcoma Patients Reveals Novel Prognostic Biomarkers Associated with Tumor Microenvironment and Immune Cells', *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, 26, pp. e924126-1-e924126-18. doi: 12659/MSM.924126.

11. Jia, D., Li, S., Li, D. *et al.* (2018) 'Mining TCGA database for genes of prognostic value in glioblastoma microenvironment', *Ageing*, 10(4), pp. 592–605. doi: 18632/aging.101415.
12. Kurze, A.-K., Buhs, S., Eggert, D. *et al.* (2019) 'Immature O-glycans recognized by the macrophage glycoreceptor CLEC10A (MGL) are induced by 4-hydroxy-tamoxifen, oxidative stress and DNA-damage in breast cancer cells', *Cell Communication and Signaling*, 17(1), p. 107. doi: 1186/s12964-019-0420-9.
13. Lan, H., Zeng, J., Chen, G. *et al.* (2017) 'Survival prediction of kidney renal papillary cell carcinoma by comprehensive LncRNA characterization', *Oncotarget*, 8(67), pp. 110811–110829. doi: 18632/oncotarget.22732.
14. Li, F., Hu, W., Zhang, W. *et al.* (2020) 'A 17-Gene Signature Predicted Prognosis in Renal Cell Carcinoma', *Disease Markers*, 2020. doi: 1155/2020/8352809.
15. Maki, R. G. (2001) 'Soft tissue sarcoma as a model disease to examine cancer immunotherapy', *Current Opinion in Oncology*, 13(4), pp. 270–274. doi: 1097/00001622-200107000-00010.
16. Mendez-Enriquez, E. and García-Zepeda, E. A. (2013) 'The multiple faces of CCL13 in immunity and inflammation', *Inflammopharmacology*, 21(6), pp. 397–406. doi: 1007/s10787-013-0177-5.
17. Miki, Y., Yamamoto, K., Taketomi, Y. *et al.* (2013) 'Lymphoid tissue phospholipase A2 group IID resolves contact hypersensitivity by driving antiinflammatory lipid mediators', *The Journal of Experimental Medicine*, 210(6), pp. 1217–1234. doi: 1084/jem.20121887.
18. Mesiano, G., Leuci, V., Giraud, L. *et al.* (2015) 'Adoptive immunotherapy against sarcomas', *Expert Opinion on Biological Therapy*, 15(4), pp. 517–528. doi: 1517/14712598.2015.987121.
19. Miki, Y., Kidoguchi, Y., Sato, M. *et al.* (2016) 'Dual Roles of Group IID Phospholipase A2 in Inflammation and Cancer', *The Journal of Biological Chemistry*, 291(30), pp. 15588–15601. doi: 1074/jbc.M116.734624.
20. Martner, A., Aydin, E. and Hellstrand, K. (2019) 'NOX2 in autoimmunity, tumor growth and metastasis', *The Journal of Pathology*, 247(2), pp. 151–154. doi: 1002/path.5175.
21. Massa, D., Baran, M., Bengoechea, J.A. *et al.* (2020) 'PYHIN1 regulates pro-inflammatory cytokine induction rather than innate immune DNA sensing in airway epithelial cells', *The Journal of Biological Chemistry*, 295(14), pp. 4438–4450. doi: 1074/jbc.RA119.011400.
22. Morales, E., Olson, M., Iglesias, F. *et al.* (2020) 'Role of immunotherapy in Ewing sarcoma', *Journal for ImmunoTherapy of Cancer*, 8(2), p. e000653. doi: 1136/jitc-2020-000653.
23. Nakayama, R., Nemoto, T., Takahashi, H. *et al.* (2007) 'Gene expression analysis of soft tissue sarcomas: characterization and reclassification of malignant fibrous histiocytoma', *Modern Pathology: An Official Journal of the United States and Canadian Academy of Pathology, Inc*, 20(7), pp. 749–759. doi: 1038/modpathol.3800794.
24. Nafia, I., Toulmonde, M., Bortolotto, D. *et al.* (2020) 'IDO Targeting in Sarcoma: Biological and Clinical Implications', *Frontiers in Immunology*, 11, p. 274. doi: 3389/fimmu.2020.00274.
25. Ohsugi, T. (2017) 'Effects of expressing human T-cell leukemia virus type 1 (HTLV-I) oncoprotein Tax on *DOK1*, *DOK2* and *DOK3* gene expression in mice', *Journal of Veterinary Medical Science*, 79(5), pp.

- 935–938. doi: 1292/jyms.17-0034.
26. Peille, A.-L., Brouste, V., Kauffmann, A. *et al.* (2013) 'Prognostic value of PLAGL1-specific CpG site methylation in soft-tissue sarcomas', *PLoS One*, 8(11), p. e80741. doi: 1371/journal.pone.0080741.
 27. Pollack, S.M., He, Q., Yearley, J.H. *et al.* (2017) 'T-cell infiltration and clonality correlate with programmed cell death protein 1 and programmed death-ligand 1 expression in patients with soft tissue sarcomas', *Cancer*, 123(17), pp. 3291–3304. doi: 1002/cncr.30726.
 28. Ryan, B.M., Zanetti, K.A., Robles, A.I. *et al.* (2014) 'Germline variation in NCF4, an innate immunity gene, is associated with an increased risk of colorectal cancer', *International Journal of Cancer*, 134(6), pp. 1399–1407. doi: 1002/ijc.28457.
 29. Sharma, V., Michel, S., Gaertner, V. *et al.* (2014) 'A role of FCER1A and FCER2 polymorphisms in IgE regulation', *Allergy*, 69(2), pp. 231–236. doi: 1111/all.12336.
 30. Toulmonde, M., Penel, N., Adam, J. *et al.* (2018) 'Use of PD-1 Targeting, Macrophage Infiltration, and IDO Pathway Activation in Sarcomas', *JAMA Oncology*, 4(1), pp. 93–97. doi: 1001/jamaoncol.2017.1617.
 31. Tong, Y., Song, Y. and Deng, S. (2019) 'Combined analysis and validation for DNA methylation and gene expression profiles associated with prostate cancer', *Cancer Cell International*, 19, p. 50. doi: 1186/s12935-019-0753-x.
 32. Toda, Y., Kohashi, K., Yamada, Y. *et al.* (2020) 'PD-L1 and IDO1 expression and tumor-infiltrating lymphocytes in osteosarcoma patients: comparative study of primary and metastatic lesions', *Journal of Cancer Research and Clinical Oncology*, 146(10), pp. 2607–2620. doi: 1007/s00432-020-03242-6.
 33. Toulmonde, M., Lucchesi, C., Verbeke, S. *et al.* (2020) 'High throughput profiling of undifferentiated pleomorphic sarcomas identifies two main subgroups with distinct immune profile, clinical outcome and sensitivity to targeted therapies', *EBioMedicine*, 62, p. 103131. doi: 1016/j.ebiom.2020.103131.
 34. Widemann, B. C. and Italiano, A. (2018) 'Biology and Management of Undifferentiated Pleomorphic Sarcoma, Myxofibrosarcoma, and Malignant Peripheral Nerve Sheath Tumors: State of the Art and Perspectives', *Journal of Clinical Oncology*, 36(2), pp. 160–167. doi: 1200/JCO.2017.75.3467.
 35. Xiao, Y., Chen, B., Yang, K. *et al.* (2019) 'Down-regulation of MARCO associates with tumor progression in hepatocellular carcinoma', *Experimental Cell Research*, 383(2), p. 111542. doi: 1016/j.yexcr.2019.111542.
 36. Zhang, G., Liu, Y., Dong, F. *et al.* (2020) 'Transcription/Expression of KLRB1 Gene as A Prognostic Indicator in Human Esophageal Squamous Cell Carcinoma', *Combinatorial Chemistry & High Throughput Screening*, 23(7), pp. 667–674. doi: 2174/1386207323666200517114154.
 37. Zhu, N. and Hou, J. (2020) 'Assessing immune infiltration and the tumor microenvironment for the diagnosis and prognosis of sarcoma', *Cancer Cell International*, 20(1), p. 577. doi: 1186/s12935-020-01672-3.
 38. Zhang, B., Yang, L., Wang, X. *et al.* (2021) 'Identification of a survival-related signature for sarcoma patients through integrated transcriptomic and proteomic profiling analyses', *Gene*, 764, p. 145105.

Figures

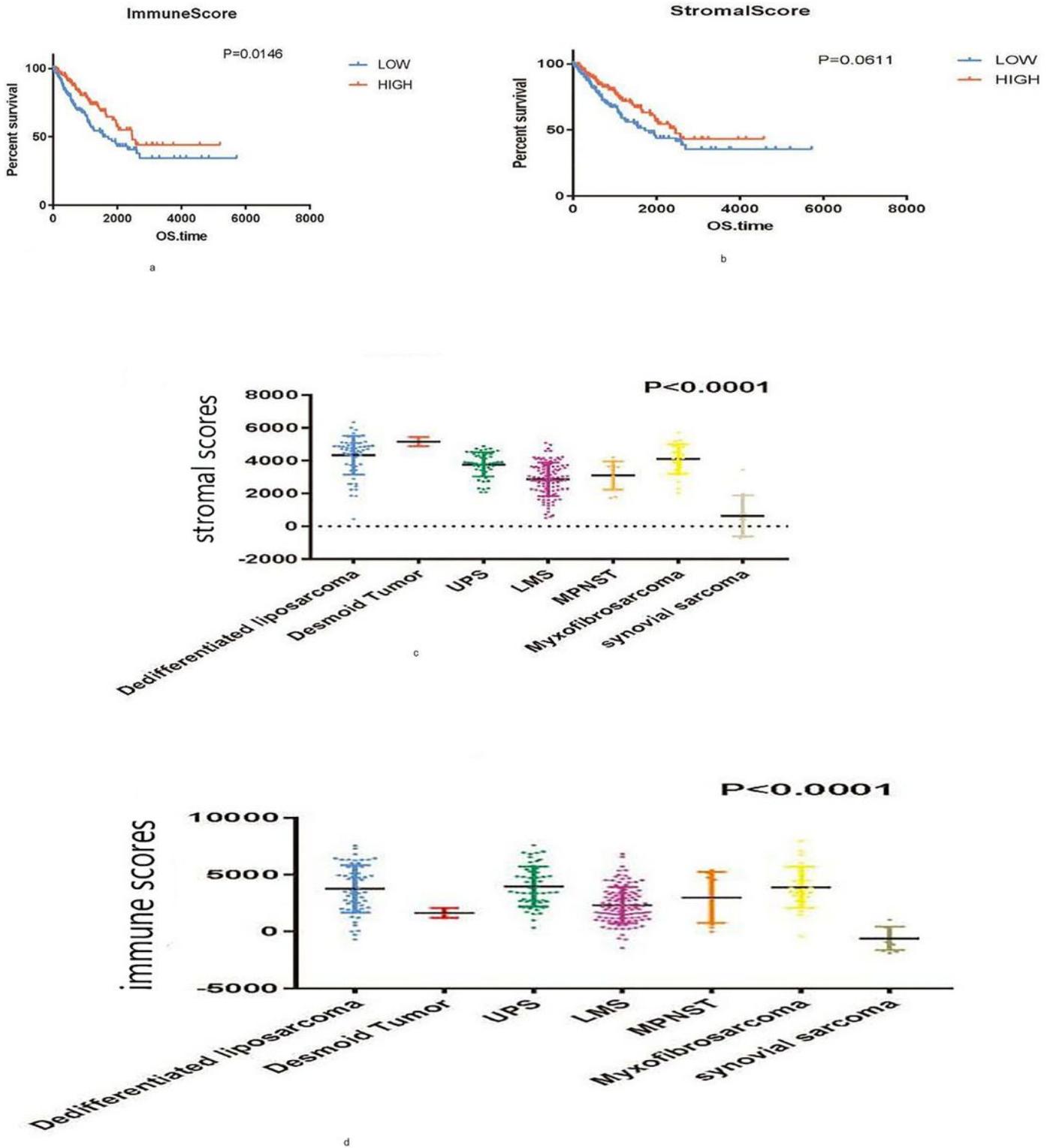


Figure 1

Relationship between immune and stromal scores and the overall survival time. (A) People with a high immune score had a longer overall survival time, which was statistically significant. Similarly, as shown in the figure B, a high stromal score indicated a longer overall survival time, although this was not statistically significant.

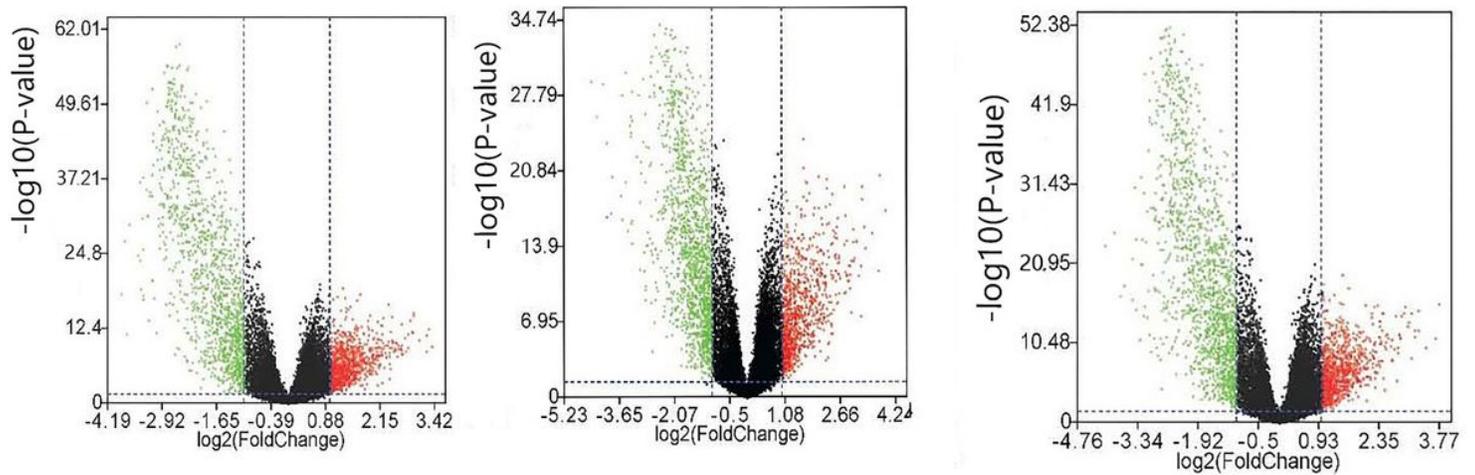


Figure 2

In volcano map(A B C), green represents downregulated genes, red represents upregulated genes, and black represents genes with the same expression level in the high and low score groups. Venn diagrams (D E) show genes that are generally up-regulated or down-regulated in all three scores.

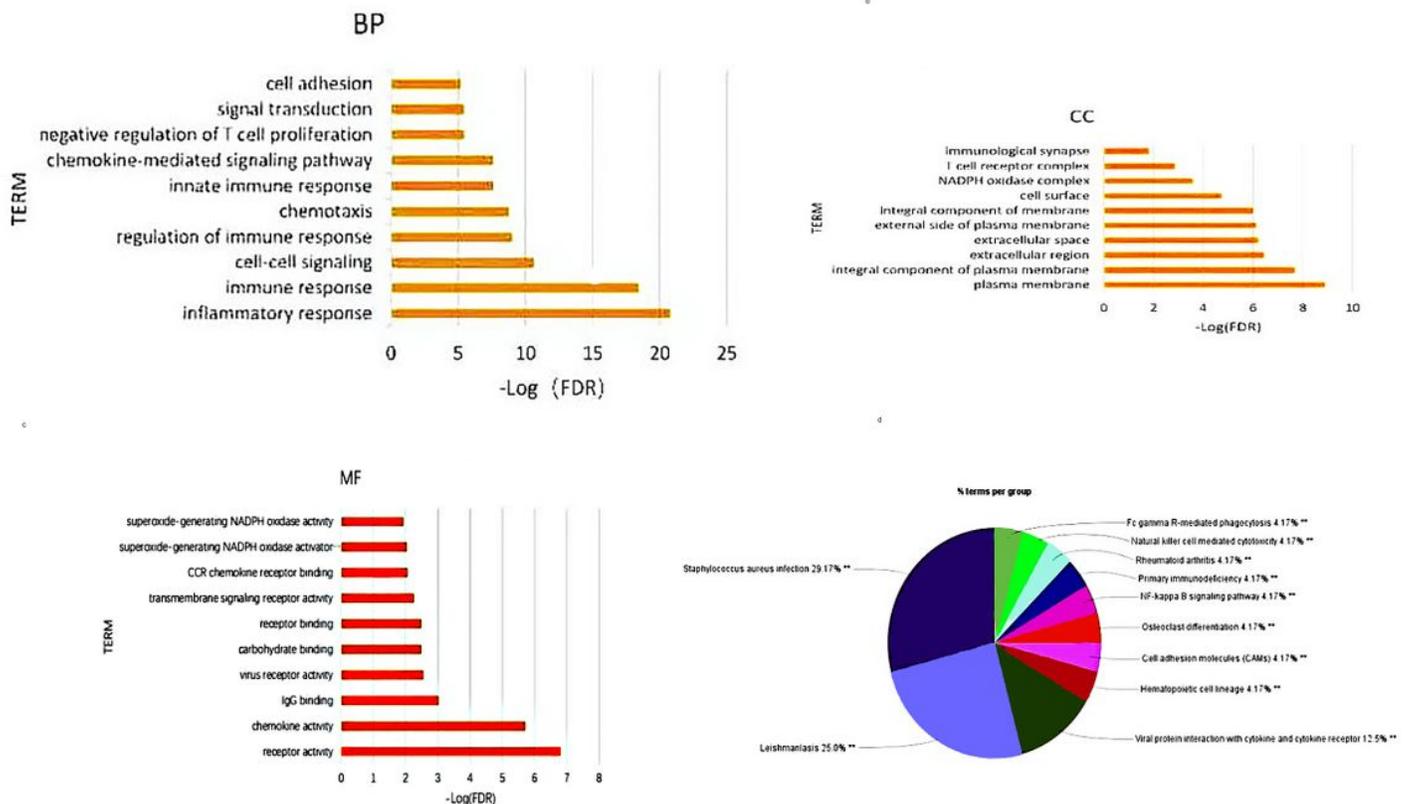


Figure 3

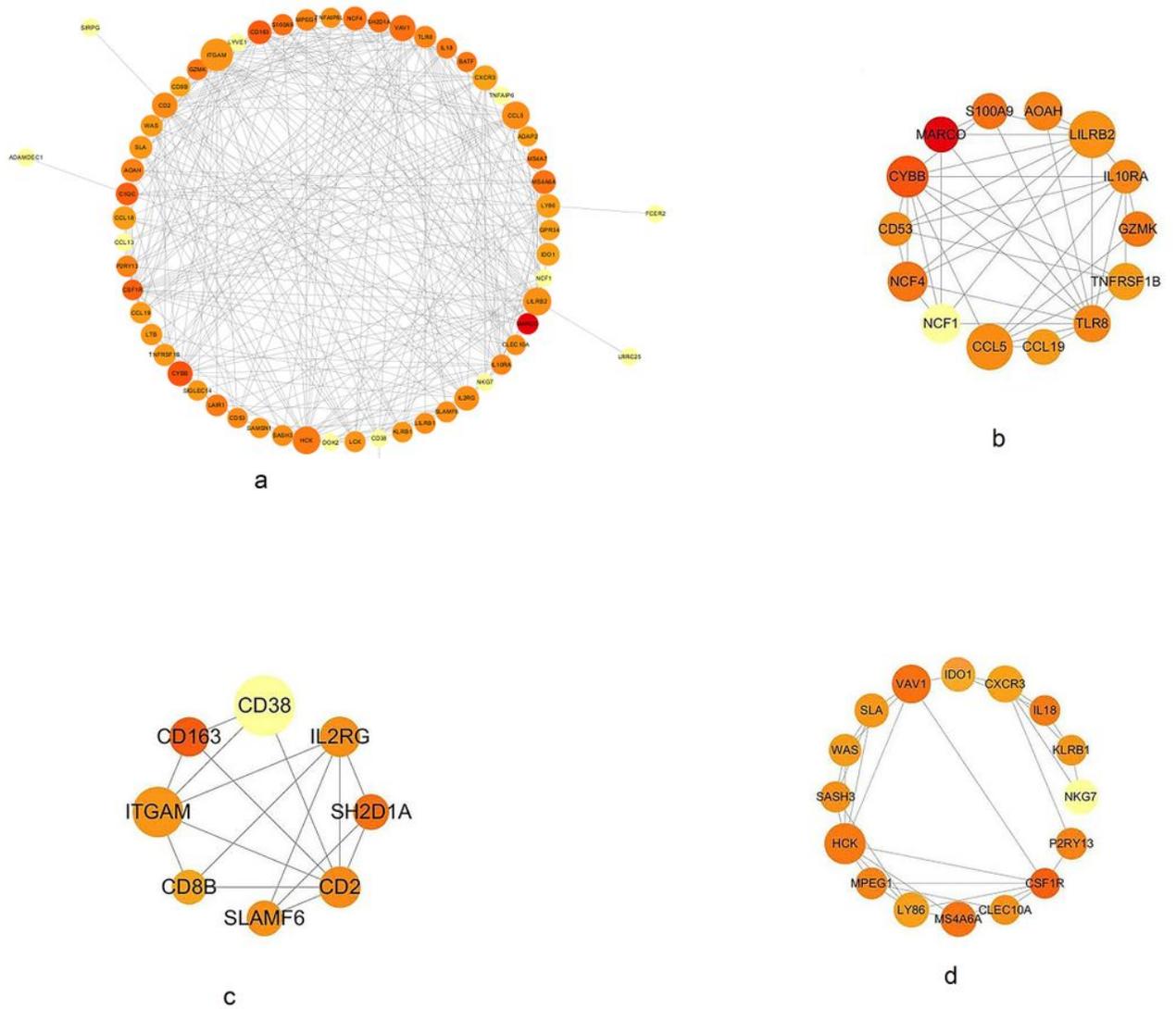


Figure 5

The PPI network of 86 different genes(A) was created to discover the relation; B C and D represent cluster 1, 2 and 3, respectively. Protein expression network of a total of 86 DEGs obtained using the web String tool divided into 3 clusters with 64 nodes with 326 edges by cytoscope analysis. The darker the color is, the greater the log (FC) value of the score of gene expression is, and the larger the circle is, the more associated it is with other proteins.