

Article

# Cell Differentiation Trajectory-Associated Molecular Classification of Osteosarcoma

Ankai Xu <sup>1,2,†</sup> , Chao Qian <sup>1,2,†</sup>, Jinti Lin <sup>1,2</sup>, Wei Yu <sup>1,2</sup>, Jiakang Jin <sup>1,2</sup>, Bing Liu <sup>1,2</sup> and Huimin Tao <sup>1,2,\*</sup>

<sup>1</sup> Department of Orthopedics Surgery, The Second Affiliated Hospital, Zhejiang University School of Medicine, No. 88, Jiefang Road, Hangzhou 310009, China; 21718246@zju.edu.cn (A.X.); med\_qianchao@zju.edu.cn (C.Q.); 11918336@zju.edu.cn (J.L.); 21518152@zju.edu.cn (W.Y.); 21918384@zju.edu.cn (J.J.); liubingzju@zju.edu.cn (B.L.)

<sup>2</sup> Orthopedics Research Institute of Zhejiang University, No. 88, Jiefang Road, Hangzhou 310009, China

\* Correspondence: huimintao@zju.edu.cn

† These authors contributed equally to this work.

**Abstract:** This study aims to investigate the differentiation trajectory of osteosarcoma cells and to construct molecular subtypes with their respective characteristics and generate a multi-gene signature for predicting prognosis. Integrated single-cell RNA-sequencing (scRNA-seq) data, bulk RNA-seq data and microarray data from osteosarcoma samples were used for analysis. Via scRNA-seq data, time-related as well as differentiation-related genes were recognized as osteosarcoma tumor stem cell-related genes (OSCGs). In Gene Expression Omnibus (GEO) cohort, osteosarcoma patients were classified into two subtypes based on prognostic OSCGs and it was found that molecular typing successfully predicted overall survival, tumor microenvironment and immune infiltration status. Further, available drugs for influencing osteosarcoma via prognostic OSCGs were revealed. A 3-OSCG-based prognostic risk score signature was generated and by combining other clinic-pathological independent prognostic factor, stage at diagnosis, a nomogram was established to predict individual survival probability. In external independent TARGET cohort, the molecular types, the 3-gene signature as well as nomogram were validated. In conclusion, osteosarcoma cell differentiation occupies a crucial position in many facets, such as tumor prognosis and microenvironment, suggesting promising therapeutic targets for this disease.

**Keywords:** tumor/cancer stem cell; differentiation trajectory; molecular typing



**Citation:** Xu, A.; Qian, C.; Lin, J.; Yu, W.; Jin, J.; Liu, B.; Tao, H. Cell Differentiation Trajectory-Associated Molecular Classification of Osteosarcoma. *Genes* **2021**, *12*, 1685. <https://doi.org/10.3390/genes12111685>

Academic Editors: Stefano Lonardi and David Gómez Matallanas

Received: 9 September 2021

Accepted: 21 October 2021

Published: 23 October 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

As the most common malignant primary bone tumor, it is reported that osteosarcoma often presented resistance to the methods of checkpoint blockade, such as inhibitors for PD-1/PD-L1 [1,2]. In addition, as the mainstay of therapy, chemotherapy is not always resultful. It is obvious that other efficient treatment thoughts are urgently needed. Naturally, tumor stem cell, as one of the reasons for immunologic escape and resistance of treatments, have emerged into researchers' vision.

Tumor heterogeneity took part in many facets in the progression of cancer including therapy resistance, recurrence and metastasis [3]. The same as other many kinds of malignant tumors, osteosarcoma has been reported extensive intratumorally heterogeneity [4,5]. Tumor stem cell, or cancer stem cell, was defined those clones drive tumor initiation, also as a crucial contributor to tumor heterogeneity [6]. Although the conception of cancer stem cell is still controversial and often been confused with cancer initiating cells, it has been reported that there are some unique subsets of tumor clones with high-grade stemness traits as well as oncogenicity [7,8].

In osteosarcoma, researchers never ceased to explore the tumor heterogeneity and tumor stem cells with various methods. Using the suspension culture, Gibbs with his group obtained a small subset of self-renewing bone sarcoma cells with high expression

of Oct 3/4 and Nanog [9]. Bulk RNA sequencing (RNA-seq) technology only provides us increasing knowledge of the osteosarcoma genome and the understanding of intra-tumor heterogeneity [7,10]. In previous study, multiregional sequencing was also an appropriate to explore heterogeneity in osteosarcoma [11]. Due to bulk RNA sequencing provides the average expression level of tissues, as the technology advances, single-cell RNA-seq (scRNA-seq) afford an opportunity to us that learn about tumor heterogeneity in the cellular level and explore tumor stem cells in greater depth [5].

In our study, using scRNA-seq data of primary osteosarcoma as well as trajectory analysis, time-related genes and cell differentiation-related genes were identified. Mutual genes were recognized as osteosarcoma tumor stem cell-related genes (OSCGs) for the next analysis. Next, we included osteosarcoma patients from Gene Expression Omnibus (GEO) database and obtained prognostic genes. Moreover, based on the expression patterns of prognostic OSCGs, in the GEO cohort, two OSCG-based molecular subtypes were identified and this cell differentiation state-based classification was proved prognostic correlative as well as tumor immunomicroenvironment related. Additionally, we explored the biological functions of prognostic OSCGs and as expected found that they are related to ossification, osteoblast differentiation and collagen metabolism. Then, latent available drugs were also revealed. Further, *CKLF*, *DKK1* and *MYC* were identified as the 3 key OSCGs which also effectual in bulk RNA sequencing, and the nomogram consisting of these 3 OSCGs and tumor stage at diagnosis was constructed. Finally, the above findings were validated using Therapeutically Applicable Research to Generate Effective Treatments (TARGET) osteosarcoma patient cohort. Brief intratumorally osteosarcoma cell differentiation states were discovered and the OSCGs played crucial roles in predicting the clinical outcomes as well as tumor immunomicroenvironment.

## 2. Materials and Methods

### 2.1. Data Sources and Data Preprocessing

The scRNA-seq data were obtained from the GSE152048 dataset in the GEO (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE152048>, accessed on 9 August 2021) database. 7 samples of primary osteosarcoma (BC2, BC3, BC5, BC6, BC16, BC21, BC22) were collected. Next, scRNA-seq data were processed with the Seurat package (version 3.1.5; <http://satijalab.org/seurat/>, accessed on 9 August 2021) in R software (version 3.6.1) for each individual sample [12]. Low-quality cells were excluded based on the following quality control standards: (1) genes detected in <200 cells were excluded; (2) cells with <1000 total detected genes were excluded; and (3) cells with  $\geq 20\%$  of mitochondria-expressed genes were excluded. After data filtering, using LogNormalize method as well as Harmony package (version 1.0), the data were normalized and removed the batch effects [13].

Further, we searched all the open samples with OS rate for osteosarcoma. Total three human datasets (GSE21257, GSE16091, GSE39055) in GEO (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE21257>; <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE16091>; <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE39055>, accessed on 9 August 2021) were collected to make GEO merged cohort. And the necessary documents were also respectively downloaded to annotate the respective probe sets into gene symbol sets. Finally, the “ComBat” algorithm was applied to reduce the likelihood of batch effects from non-biological technical biases [14].

And then samples including the prognosis information of osteosarcoma patients from TARGET (<https://ocg.cancer.gov/programs/target/projects/osteosarcoma>, accessed on 9 August 2021) were also gathered, and FPKM values were transformed into transcripts per kilobase million (TPM) values and then  $\log_2(\text{TPM}+1)$  values, which are more comparable between samples and more similar to data from microarrays [15,16]. All the gene matrices were standardized by quantiles normalization. Finally, 121 patients in GEO merged cohort and 95 patients in TARGET cohort were selected.

## 2.2. Dimensionality Reduction and Cell Annotation

Based on the integrated joint embedding produced by Harmony with the Louvain algorithm as well as the *t*-SNE algorithm, the dimension reduction of the scRNA-seq data was accomplished and the major clusters was obtained. Additionally, sub-clustering analysis was applied a similar procedure.

Using FindAllMarkers function in Seurat, differentially expressed genes (DEGs) between the cell clusters were gathered. The cell groups were annotated based on the previous study [5].

## 2.3. Pseudotime and Trajectory Analysis

Pseudotime and trajectory analyses of osteosarcoma cells were carried out by the 'Monocle' package (version 2.18.0) [17]. DEGs with *q*-value < 0.01 between the sub-cell groups were applied for dimension reduction with the reduceDimension function using the parameters reduction\_method = "DDRTree" and max\_components = 2. Based on plot\_cell\_trajectory function, osteosarcoma cells were ordered and visualized. Time-related genes were calculated (*q*-value < 0.01) and exhibited via plot\_cell\_trajectory and plot\_pseudotime\_heatmap function. BEAM analysis was performed (*q*-value < 0.01) to find cell differentiation-related genes visualized by plot\_genes\_branched\_heatmap function.

## 2.4. OSCG-Based Molecular Subtypes of Osteosarcoma Patients

In GEO cohort, the Kaplan-Meier survival analysis based on the median of gene expression and univariate Cox hazard analysis were applied to screen OSCGs significantly associated with OS rate (*p* < 0.05).

In the discovery GEO and validation TARGET cohorts respectively, unsupervised consensus clustering was used to discover OSCG-based osteosarcoma subtypes [18]. In detail, a bootstrap procedure was performed with 80% item resampling 1000 times using the agglomerative kmeans clustering algorithm with the Euclidean distance metric. In the number of clusters from 2 to 9, the optimal number of clusters was selected that corresponds to both stable consensus matrices and unambiguous cluster assignments. Moreover, to validate the result, IGP analysis was used to evaluate the reproducibility of the clusters derived from consensus clustering in the two independent cohorts [19]. Due to the different types of expression values in the two datasets, we normalized the expression data by Z-score prior to the IGP statistical analysis.

## 2.5. Tumor Microenvironment Evaluation and Latent Drugs Related to Molecular Subtypes

Via 'ESTIMATE' package, the immune/stromal scores and tumor purity of each sample were calculated [20]. Tumor immune cell infiltration situation of 22 immune cells was identified by CIBERSORT algorithm [21]. Based on 'limma' package, the discrepancy of immune cells' infiltration degree was evaluated, and a violin plot of immune cells with significant differences was presented. Using the Connectivity Map database (Previous version; <https://portals.broadinstitute.org/cmap/>, accessed on 9 August 2021), available drugs for influencing osteosarcoma OSCGs were revealed.

Additionally, GO analyses were realized via "clusterProfiler" package, and *p* adjust < 0.05 was believed to be statistically significant [22].

## 2.6. Risk Model Construction and Evaluation

In GEO cohort, we used the randomSurvivalForest algorithm to rank the importance of OSCGs [23]. We identified genes with a relative importance >0.50 as the final signature. Next, the multivariate Cox regression analysis was performed to build the risk score model and the model formula was showed above-mentioned.

Via timeROC package in R, ROC and AUC was performed to evaluate the classification efficiency of the prognosis predictions for 1 year, 2 years, 3 years and 5 years [24]. Kaplan-Meier survival analysis was performed to estimate the OS of two groups, and survival differences were evaluated by a two-sided log-rank test.

### 2.7. The Construction of Nomogram

Univariate and multivariate Cox regression analyses were performed in both osteosarcoma cohorts to determine whether the predictive performance of risk scores could be independent of the clinicopathological variables. These variables included gender, age, tumor site, stage at diagnosis, stromal scores, immune scores, ESTIMATE scores, tumor purity and risk scores.

Via the “rms” package in R, independent predictors, metastasis stage at diagnosis and risk score were used to construct the nomogram. The predictive accuracy of the OSCG-based prognostic model was evaluated by Harrell’s concordance index (C-index). In addition, calibration curves were drawn to assess the consistency between actual and predicted survival.

## 3. Results

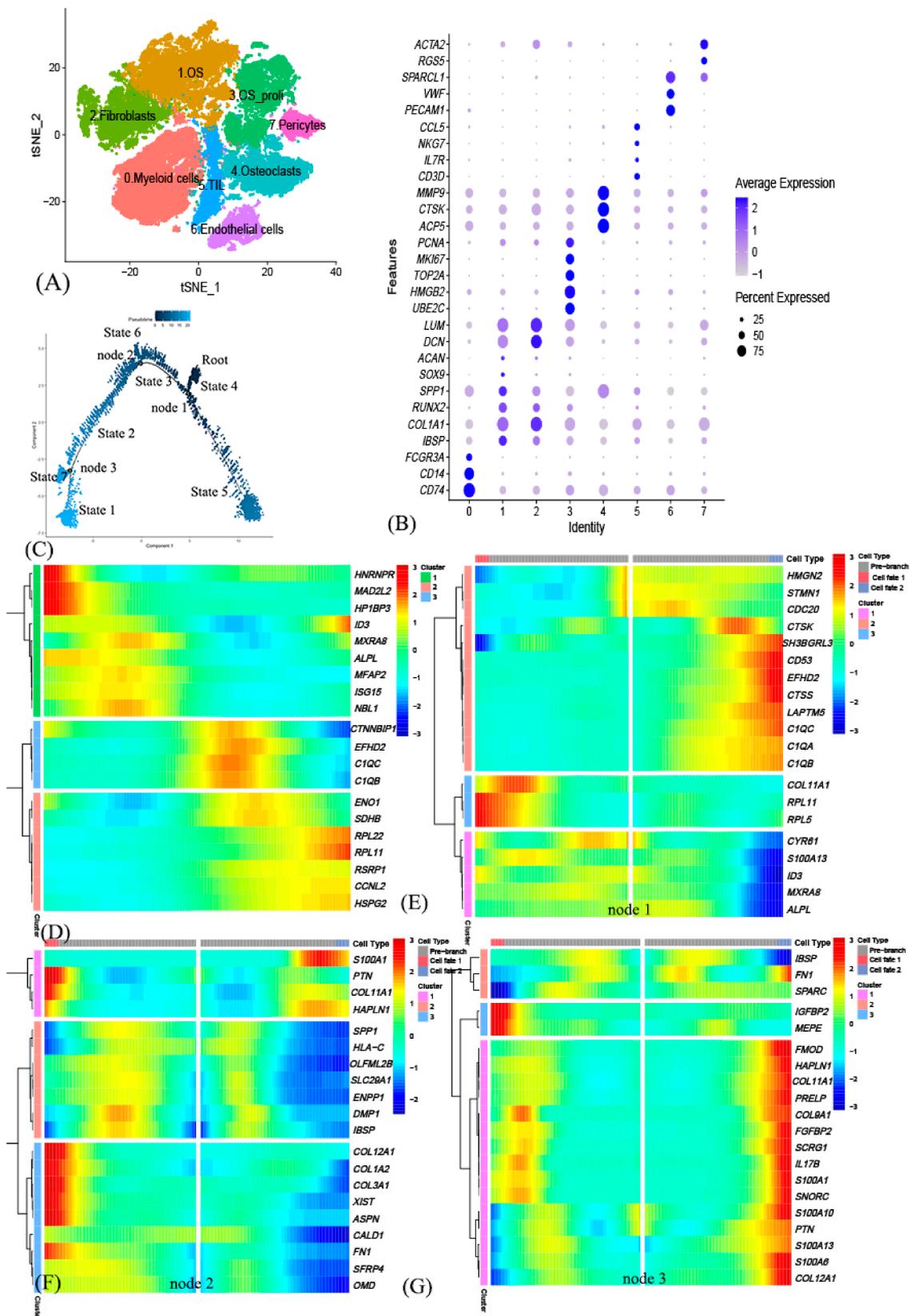
### 3.1. Identification of 8 Cell Clusters in Primary Osteosarcoma Using scRNA-seq Data

The schematic diagram of this study is illuminated in Supplementary Figure S1. From GEO database, seven raw output data of scRNA-seq for primary osteosarcoma were obtained.

After initial quality control assessment and batch effects removal, we obtained single-cell transcriptomes from 67,146 cells for the further analysis. Based on the *t*-distributed stochastic neighbor embedding (*t*-SNE) algorithm, the cells in human osteosarcoma were successfully classified into 8 clusters (Figure 1A). And the typical canonical markers for major cell types were illuminated in Figure 1B (Supplementary Figures S2 and S3, Figure S4A), according to the original previous research [5]. Cluster 0, containing 15,420 cells, was annotated as myeloid cells; Cluster 1, containing 14,882 cells, was annotated as osteosarcoma cells; Cluster 2, containing 9528 cells, was annotated as fibroblasts; Cluster 3, containing 8876 cells, was annotated as osteosarcoma cells with high proliferation; Cluster 4, containing 8125 cells, was annotated as osteoclasts; Cluster 5, containing 4640 cells, was annotated as tumor infiltrating lymphocytes (TILs); Cluster 6, containing 3191 cells, was annotated as endothelial cells; Cluster 7, containing 2484 cells, was annotated as pericytes.

### 3.2. Cell Trajectory Analysis Identified OSCGs

Then, osteosarcoma subsets, Cluster 1 and Cluster 3, including 11,780 cells were extracted for further analysis (Supplementary Figure S4B–D). Trajectory analysis of osteosarcoma cells was performed based on the Monocle 2 algorithm, finding three differentiation nodes (Figure 1C) [17]. And then 859 pseudo time-related genes were obtained (Figure 1D). Via the branch expression analysis modeling (BEAM) analysis, three cell differentiation nodes-related genes were received respectively, and a total of 1042 genes were received. (Figure 1E–G). Combining with pseudotime-related genes and cell differentiation nodes-related genes, there were total 542 genes recognized as OSCGs (Supplementary Table S1).



**Figure 1.** Single-cell transcriptomic analysis of osteosarcoma. **(A)** The t-distributed stochastic neighbor embedding (t-SNE) plot of the 8 identified main cell types. **(B)** Typical 28 signature gene expressions across cellular clusters. **(C)** The Monocle 2 trajectory plot showing the pseudotime curve of osteosarcoma subclusters. **(D)** Heatmap of the top 20 genes that were differentially expressed along the pseudotime. **(E–G)** Heatmap of the top 20 genes that were differentially expressed in each cell fate branch.

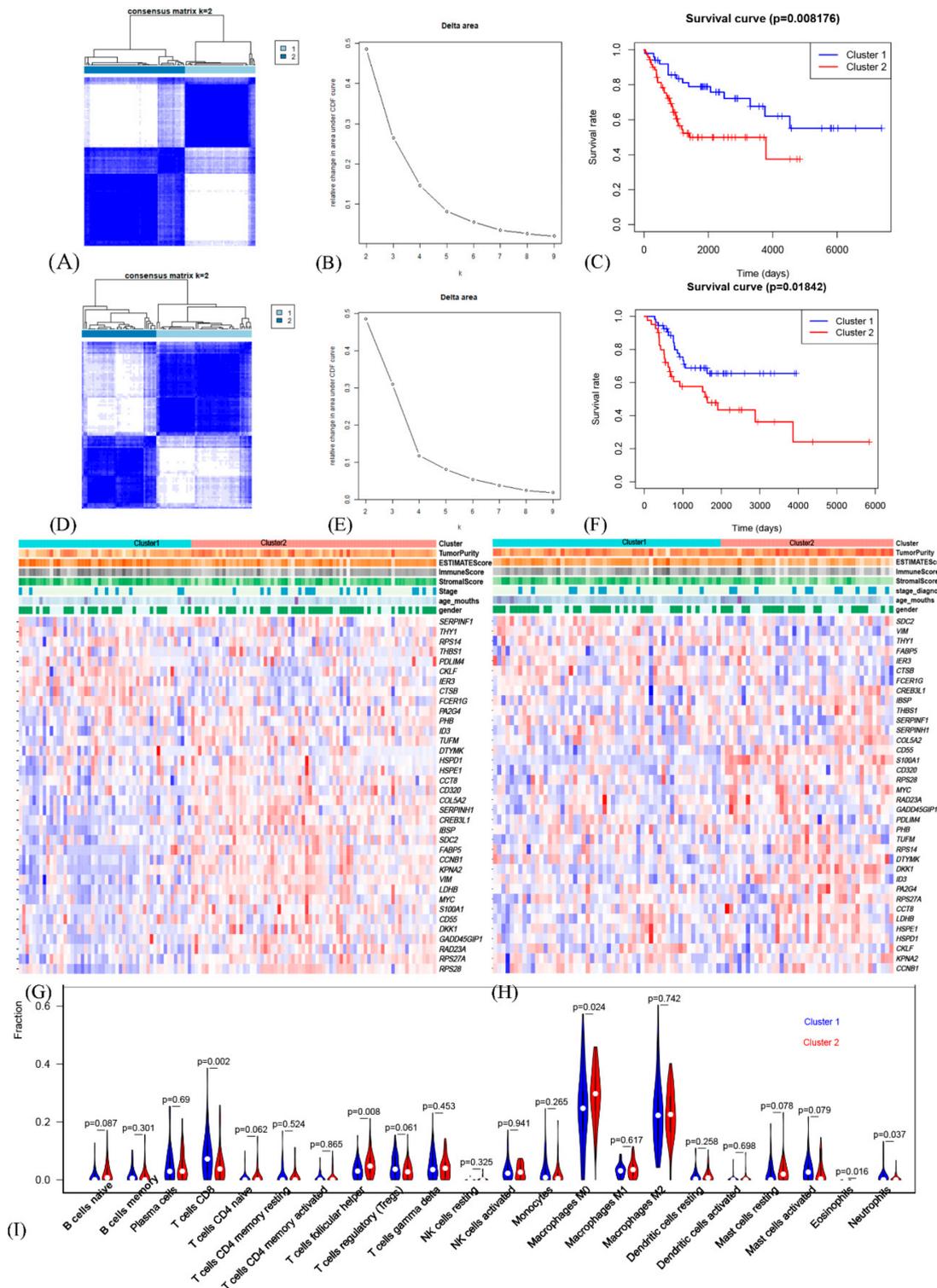
### 3.3. The Prognostic OSCG-Based Molecular Subtypes of Osteosarcoma Patients

First, respectively, Kaplan-Meier survival analysis based on the median of gene expression and univariate Cox hazard analysis were applied to the GEO cohort to screen OSCGs significantly associated with overall survival (OS) rate. Subsequently, 36 prognostic OSCGs were obtained (Supplementary Table S2). Based on prognostic OSCGs and unsupervised consensus clustering analysis in the GEO cohort, two molecular subtypes of osteosarcoma samples were identified (Figure 2A) [18]. The 2-cluster solution corresponded to the largest cluster number that induced the least incremental change in the area under the cumulative distribution function (CDF) curves while keeping the maximal consensus within clusters and the minimal rate of ambiguity in cluster assignments (Figure 2B). Then, the Kaplan-Meier analysis determined the statistical significance of consensus clustering result of osteosarcoma that subtype\_1 having better OS than subtype\_2 (Figure 2C,  $p = 0.008176$ ). Further, the same results were obtained in an external independent TARGET cohort (Figure 2D–F,  $p = 0.01842$ ). Finally, the reproducibility of the two osteosarcoma subtypes across the discovery and validation cohorts was evaluated by in-group proportion (IGP) statistic [19]. The IGP values are 77.3 and 71.2% for subtype\_1 and subtype\_2, respectively, indicating that both subtypes had high consistency between the two cohorts.

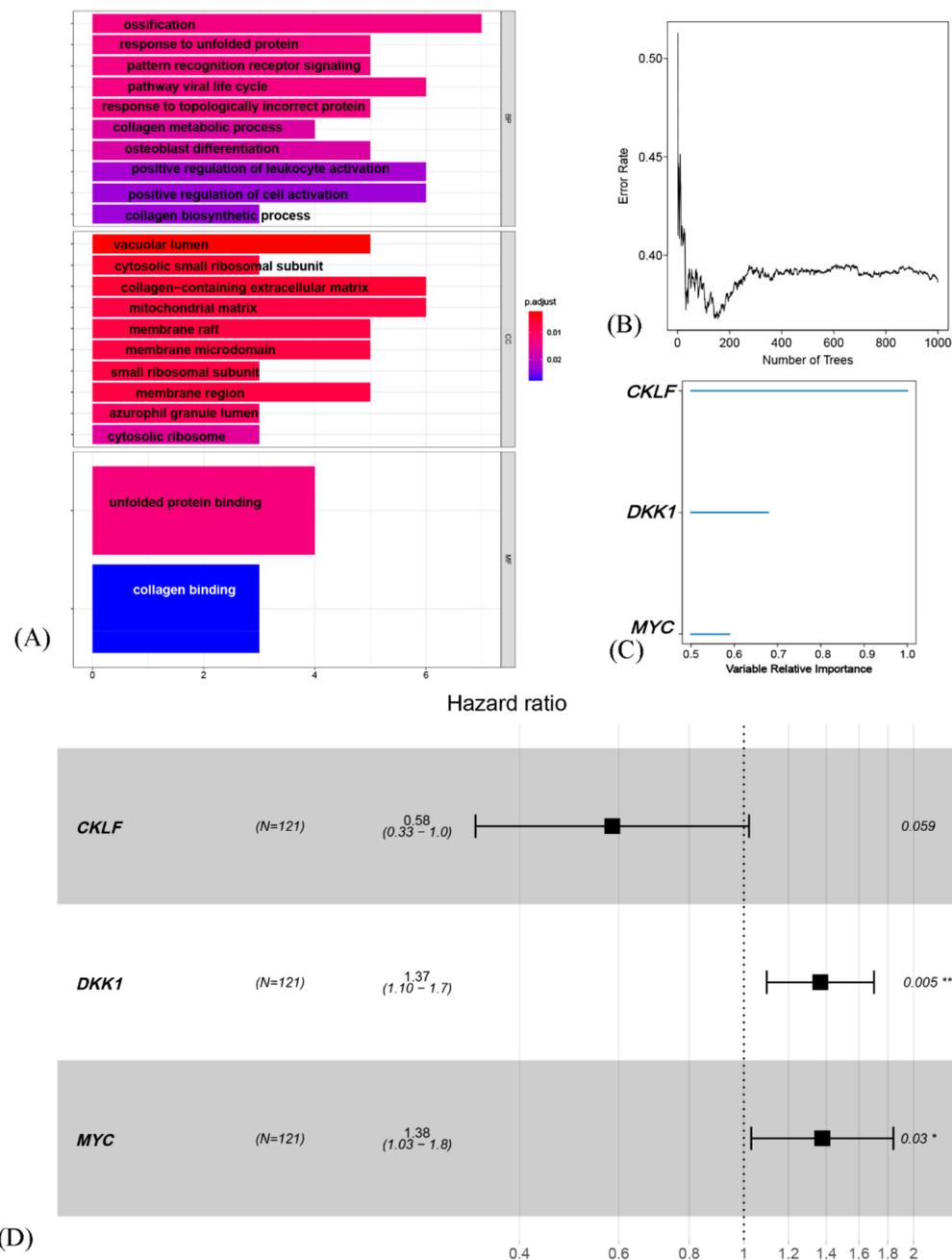
### 3.4. Comprehensive Analysis of Tumor Microenvironment Scores and Immune Cell Infiltration across Molecular Subtypes

According to tumor microenvironment evaluation, in GEO cohort, our study found that cluster\_2 had significantly higher tumor purity ( $p = 0.00026$ ) but lower ESTIMATE score ( $p = 0.00026$ ), immune score ( $p = 0.000016$ ) and stromal score ( $p = 0.012$ ) than cluster\_1 (Figure 2G, Supplementary Table S3) [20]. Additionally, there were no significant differences in age, gender and stage at diagnosis between two subtypes. Also, the same situation could be found in TARGET cohort, that cluster\_2 had significantly higher tumor purity ( $p = 0.0012$ ) but lower ESTIMATE score ( $p = 0.0012$ ), immune score ( $p = 0.019$ ) and stromal score ( $p = 0.00056$ ) than cluster\_1 and there were also no significant differences in age, gender and stage at diagnosis between two subtypes. (Figure 2H, Supplementary Table S3). Further, via the CIBERSORT algorithm, the tumor immunocytes infiltration microenvironment was evaluated (Figure 2I) [25]. Differential analysis found that cluster\_2 contained more follicular helper T cells ( $p = 0.008$ ) and M0 macrophages ( $p = 0.024$ ), but less CD8 T cells ( $p = 0.002$ ), implying the important function of immunotherapy as well as tumor-associated macrophages and suggesting the potential relationship between immunity and tumor stem cell.

Additionally, using Gene Ontology analysis for prognostic OSCGs, the process of ossification, response to unfolded protein, response to topologically incorrect protein, osteoblast differentiation and collagen metabolism could have close relationship with cell differentiation trajectory of tumor stem cell in osteosarcoma (Figure 3A). These results supported that osteosarcoma are derived from progenitor cells in the osteoblast lineage [26,27]. Moreover, consistent with previous studies, osteoblast differentiation and the formation of osteospheres could reflect the stemness of osteosarcoma [28,29]. Although collagen metabolism is significant for mesenchyma stem cells, few studies have explored the relationship between collagen metabolism and osteosarcoma stem cells. Using the Connectivity Map database, available drugs for influencing osteosarcoma OSCGs were revealed, including proglumide, midecamycin, naproxen, ethambutol, denatonium benzoate, tiapride, salsolidin and pyrimethamine (Supplementary Table S4) [30].



**Figure 2.** The prognostic OSGC-based molecular subtypes of osteosarcoma. (A) Consensus matrix heatmaps for the chosen optimal cluster number ( $k = 2$ ) for the GEO cohort. (B) Cumulative distribution function (CDF) curves for the GEO cohort. (C) Kaplan-Meier analysis of OSGC-based molecular subtypes of osteosarcoma for GEO cohort. (D–F) The similar analyses in TARGET cohort. (G) Heatmap of prognostic OSGC-based genes expression, molecular subclusters, tumor microenvironment evaluation and clinicopathological characteristics in GEO cohort. (H) The similar heatmap in TARGET cohort. (I) The tumor immunocytes infiltration microenvironment across the two subtypes.



**Figure 3.** Generation a prognostic risk scoring signature (A) GO terms analysis for prognostic OSCGs. (B) Error rate for the data as a function of the classification tree. (C) out-of-bag importance values for the predictors. (D) Forest plot showing hazard ratio of the three genes via multivariate cox regression analysis in GEO. \*  $p < 0.05$ , \*\*  $p < 0.01$ .

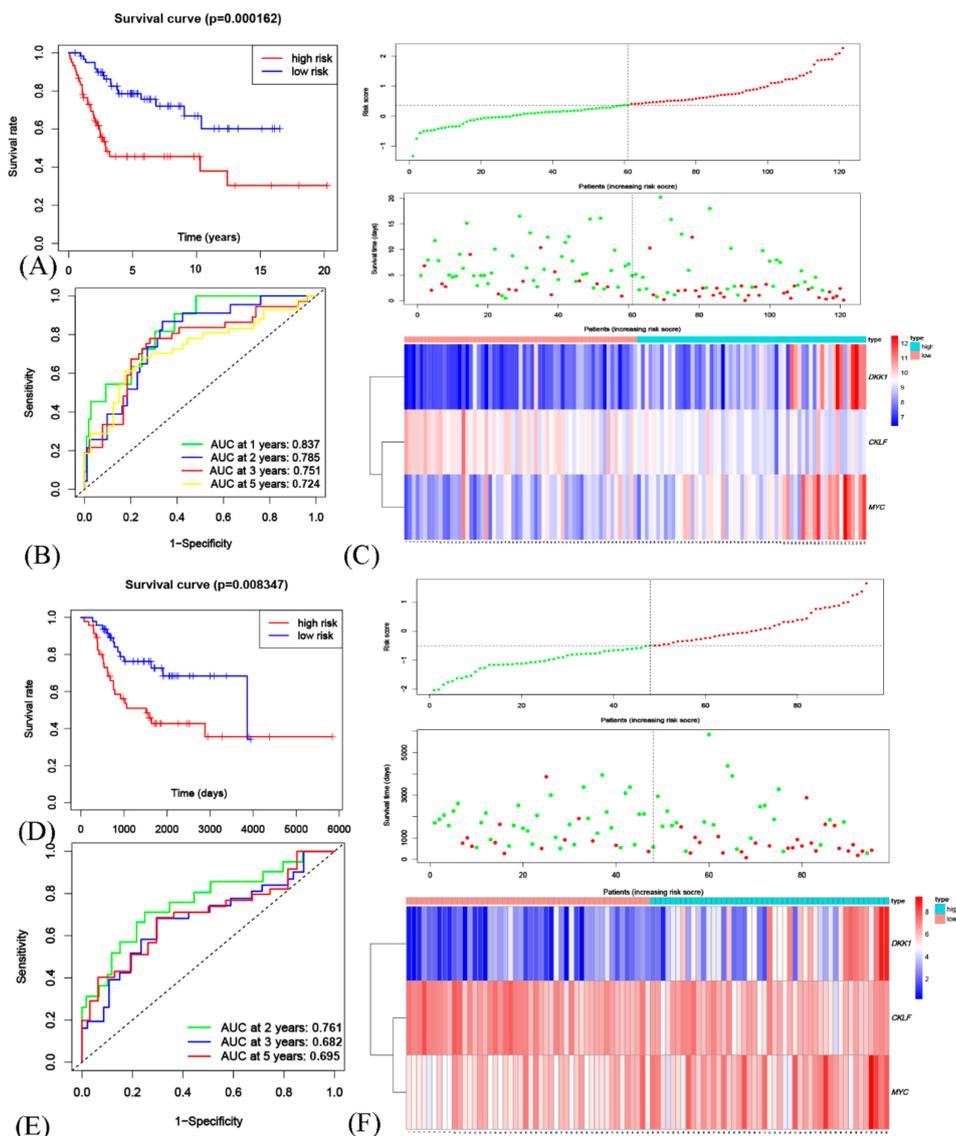
### 3.5. Generation and Validation of a Prognostic Risk Scoring Signature

In GEO cohort, univariate Cox hazard analysis was performed with  $p < 0.05$  as the threshold. Then, using the randomForestSRC software package, we ranked the importance of prognostic-OSCGs. The final signature was identified as genes with a relative importance  $> 0.50$  (Supplementary Table S5). The relationship between the error rate and the number of classification trees as well as the out-of-bag importance of the first three genes were then exhibited (Figure 3B,C).

Three genes (*CKLF*, *DKK1*, *MYC*) were identified from the random forest algorithm and subsequently multivariate Cox regression analysis were exerted to establish the 3-gene

signature (Figure 3D). The model was as follows:  $\text{RiskScore} = -0.537343156 * \exp\text{CKLF} + 0.312512732 * \exp\text{DKK1} + 0.320496576 * \exp\text{MYC}$ .

In GEO cohort, the Kaplan-Meier curves were drawn (Figure 4A), and a markedly significant difference of  $p < 0.001$  was observed between the groups by median cut-off value (Median cut-off value = 0.37). By the time-dependent receiver operating characteristic (ROC) curves, the area under the ROC curve (AUC) for 1-year, 2-year, 3-year, and 5-year OS were gathered (Figure 4B). Additionally, the distribution of risk scores in patients and the relationship between risk scores and survival time were displayed (Figure 4C). By heatmap, gene expression profiles in high-risk and low-risk were exhibited (Figure 4C). The genes with  $\text{HR} > 1$  (*DKK1*, *MYC*) were considered dangerous, while the gene with  $\text{HR} < 1$  (*CKLF*) protective. Additionally, the high-risk group has the higher expression dangerous genes with the lower protective one.



**Figure 4.** Validation of the OSCG-based risk score model in osteosarcoma patients. (A) Kaplan-Meier survival analysis to estimate the OS of high-risk and low-risk patients in GEO cohort. (B) Time-dependent ROC curve analysis was performed to evaluate the prognostic performance of the OSCG signature for predicting the 1-, 2-, 3- and 5-year OS rates in GEO cohort. (C) Risk score in the GEO cohort were calculated, and the patients were divided into either a high-risk group or a low-risk group using the median value. Risk scores, patient survival status and the signatures expression across the two groups were displayed in GEO cohort. (D–F) Similar analyses were performed in TARGET cohort using the median value as cutoff value.

Further, the 3-gene signature was validated in the external independent TARGET cohort. In TARGET cohort, by the median value as cut-off value, the Kaplan-Meier curves were drawn (Figure 4D), and a markedly significant difference of  $p < 0.01$  was observed. By ROC curves, the AUC for 2-year, 3-year, and 5-year OS were gathered (Figure 4E). The gene expression heatmap, the distribution of risk scores in patients and the relationship between risk scores and survival time were also displayed (Figure 4F).

### 3.6. The Risk Model Based on OSCGs as an Independent Prognostic Factor

In GEO merged cohort, both univariate and multivariate analysis showed that the risk scores an independent prognostic indicator (HR = 2.718, 95%CI = 1.730–4.271,  $p < 0.001$ ; HR = 3.148, 95%CI = 1.722–5.756,  $p < 0.001$ ) (Supplementary Table S6). Further, we validated this model in the external independent TARGET cohort with the same way. And the risk score was still proved an independent prognostic factor the same as tumor stage at diagnosis (HR = 2.308, 95%CI = 1.498–3.554,  $p < 0.001$ ; HR = 2.344, 95%CI = 1.419–3.870,  $p < 0.001$ ).

### 3.7. The Construction of Nomogram for Predicting Patient 3-Year and 5-Year OS

In the clinical environment, a nomogram is a highly legible illustration of a mathematical model and a convenient tool for predicting the outcome of individual patients. The above analysis showed that the risk score as well as tumor stage at diagnosis were independent prognostic indicators. And in this study, the nomogram consisting of above significant risk factors was established, with C-index of 0.757 in GEO merged cohort and 0.773 in TARGET cohort (Figure 5A). The calibration curves showed non-significant deviations between predicted and actual probability in both GEO merged cohort and TARGET cohort (Figure 5B,C).

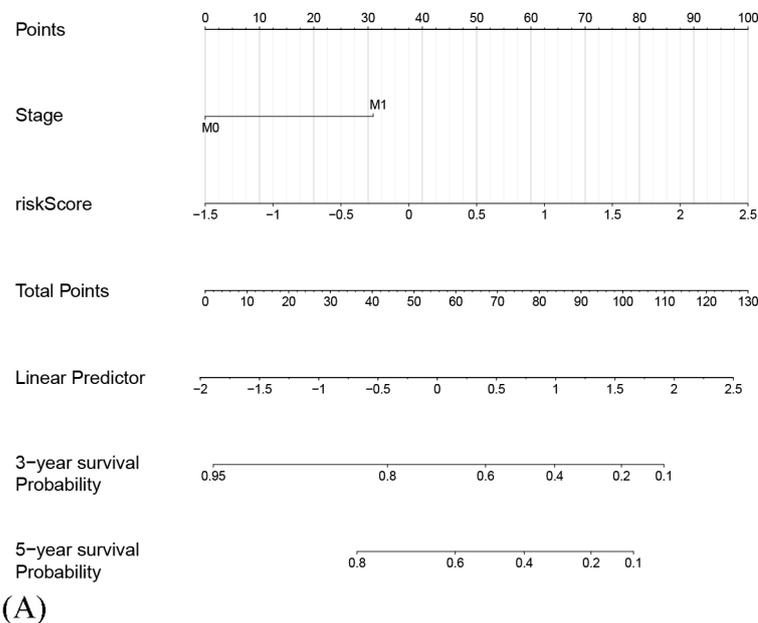
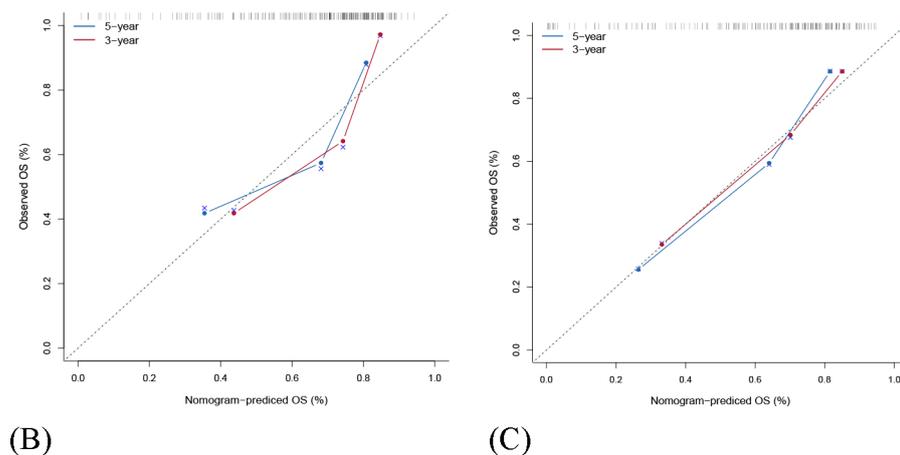


Figure 5. Cont.



**Figure 5.** Construction and evaluation of a nomogram. (A) A nomogram for predicting 3-year and 5-year OS. (B) The calibration curves for predicting 3-year and 5-year OS in GEO cohort. (C) The calibration curves for predicting 3-year and 5-year OS in TARGET cohort.

#### 4. Discussion

Whereas the researches of the earlier years had noticed the existence of tumor stem cells (TSCs, also cancer stem cells—CSCs), the precise definition and the standard of perfection for TSCs are still deficient. Both tumor-initiating cells and drug-resistant cells could be deemed to TSCs. Similarly, cells with highly proliferation ability or cells in the dormancy stage may also be thought to TSCs [31].

In terms of TSCs in osteosarcoma, the scientists never ceased from exploration. There are many methods for obtaining TSCs: dye exclusion assays, utilizing aldehyde dehydrogenase, utilizing cell surface markers (such as *CD133*, *CD117*, *IBSP* and *CD105*) and sphere formation assays [8]. However, it is hard to identify quiescent TSCs. Moreover, the alleged TSC phenotypes may vary with diverse isolation methods [8]. With continuing development of scRNA-seq technology as well as machine learning, new methods as well as perspectives have been provided to us for seeking the OSCGs. In our study, 36 prognostic OSCGs were obtained to generate molecular subtypes of osteosarcoma. Among them, many genes have been proved related to cancer stemness and tumorigenesis in osteosarcoma. It has been reported that osteosarcoma may be origin from the mesenchymal stem cell (MSC) or the osteoblast [32]. In addition, transformed human MSCs or osteoblasts were observed acquiring malignant osteosarcoma-like properties [33–35]. From MSCs to osteosarcoma, *CDKN2A*, *ALPL*, *SPARC* (osteonectin) and *MYC* were considered as origin-related factors [32,35,36]. From osteoblasts to osteosarcoma, *SOX9*, *SPP1* (osteopontin) and *RUNX2* were also considered as origin-related factors [32,37]. Moreover, in patients with Li-Fraumeni syndrome, *DCN* (decorin) mediated Li-Fraumeni syndrome related osteosarcoma [33]. By the way, *CD117*, *IBSP* (Stro-1), *ID1*, *SOX9*, *CD24*, *CD44* and *THBS-1* were also reported to affect osteosarcoma growth and stemness [37–42].

Previous studies have indicated that some TSCs-related genes correlate with the prognosis for many neoplasms [43,44]. Moreover, due to the TSCs-related genes are not universal, specific researches should be implemented respectively for various neoplasms. Currently, aiming at TSCs-related hedgehog signaling pathway, vismodegib and sonidegib are approved by the US Food and Drug Administration (FDA) for use in advanced basal cell carcinoma [45]. In addition, *CD47*, was not only considered as TSCs-related gene in leukemia and bladder tumor, but also an important marker to evade phagocytosis [46,47]. Further, in clinical trials, *CD47* blockade combining rituximab in non-Hodgkin's lymphoma showed promising activity [48]. In osteosarcoma, *Dkk1* was considered as molecule with pro-tumor function [49]. It is interesting that the clinical trials have already been in progress for *DKK1* inhibited drugs BHQ880 and DKN-01 in other tumors [50,51].

Although earlier studies mentioned above indicated the importance of TSCs in osteosarcoma, little attention was paid to using scRNA-seq technology for exploring TSC

related genes neither related molecular subtyping. On the one hand, the heterogeneity of tumor makes it hard to target all the tumor cells. On the other hand, similar to traditional pathology types, molecular phenotype identification can also be used to optimize and simplify treatment strategies. Based on TSCs to construct molecular subtypes can obviously an appropriate strategy. Considering the TSCs may only occupy a small proportion of the total tumor cells, directly targeting well-known stem-cell markers like *CD133* may serve a restricted role. In our study, combining scRNA-seq data and bulk RNA-seq data develops better strategy. TSC related genes as well as molecular subtypes were obtained. Moreover, the molecular subtypes successfully predicted overall survival, tumor microenvironment and immune infiltration status, no matter the metastasis stage at diagnosis. Additionally, related latent drugs were exploited. Further, for the osteosarcoma population, our 3-gene risk score model based on OSCGs was shown to be a reliable predictor of prognosis. Subsequently, the nomogram consisting of significant risk factors was established.

However, the present study still has some limitations. Firstly, the race and district of patient were heterogeneous because of the retrospective multiple-platform study. Secondly, although gene symbols with changed official name were converted into latest edition as far as possible, some old gene symbols have been officially abandoned. Thirdly, integrated clinical characteristic was incomplete for osteosarcoma cohort, such as, the incomplete EFS or PFS curves, needing preferable clinical follow-up and data collection. Additionally, drug-resistant TSCs or dormancy stage TSCs were still hard to identify, needing more information involved. Finally, experimental studies as well as further prospective clinical studies are required to further validate and polish our results.

## 5. Conclusions

Via scRNA-seq data, bulk RNA-seq data and microarray data, we found that osteosarcoma samples could be briefly divided into two molecular subtypes associated with prognostic OSCGs. Moreover, the classification of osteosarcoma patients based on aforesaid patterns can to some extent predict patient overall survival, tumor microenvironment and immune infiltration status. Further, the key prognosis-predicting OSCGs were identified and a related nomogram was established. In summary, this study highlights not only the promise of researches about osteosarcoma cell differentiation and osteosarcoma stem cells, but the essential roles of corresponding molecular subtypes or OSCGs in predicting the clinical outcome and immune-micro-environment.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/genes12111685/s1>, Table S1: 542 tumor stem cell-related genes, Table S2: 36 prognostic tumor stem cell-related genes, Table S3: The differences between two subtypes via unsupervised clustering, Table S4: The latent drugs via prognostic tumor stem cell-related genes in osteosarcoma, Table S5: The top 10 outcome by random forest analysis, Table S6: To find independent predictors by univariate and multivariate cox regression analyses, Figure S1: Schematic diagram showing the study design and principal findings, Figures S2 and S3: Other marker gene expressions across cellular clusters, Figure S4: Single-cell transcriptomic analyses of all cells and the subclusters of osteosarcoma cells.

**Author Contributions:** A.X. and C.Q. contributed to the planning of the study. A.X., W.Y., J.L. and J.J. carried out the study and collected important background information. A.X. and C.Q. drafted the manuscript and revised the manuscript. B.L. and H.T. are in charge of the whole research conduction and paper writing. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by grant from the National Natural Science Foundation of China (NO.81872181, NO. 82002836 and NO. 81702662), Natural Science Foundation of Zhejiang Province (NO.LY20H160025 and LY18H160015) and the Science and Technology Planning Project of Zhejiang Province (NO. 2019KY079).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Publicly available datasets were analyzed in this study. The data are accessible at the following repositories: <https://www.ncbi.nlm.nih.gov/gds>, <https://ocg.cancer.gov/programs/target/data-matrix>; accessed on 9 August 2021.

**Acknowledgments:** A tool by Welsh, E.A. et al. in Excel was used for us to prevent gene symbols' conversion errors [52] And via a convenient website portal, we generated Venn diagrams in our study (<http://bioinformatics.psb.ugent.be/webtools/Venn/> (accessed on 9 August 2021)).

**Conflicts of Interest:** The authors declare that they have no competing interests.

## References

1. Tawbi, H.A.; Burgess, M.; Bolejack, V.; Van Tine, B.A.; Schuetze, S.M.; Hu, J.; D'Angelo, S.; Attia, S.; Riedel, R.F.; Priebat, D.A.; et al. Pembrolizumab in advanced soft-tissue sarcoma and bone sarcoma (SARC028): A multicentre, two-cohort, single-arm, open-label, phase 2 trial. *Lancet Oncol.* **2017**, *18*, 1493–1501. [[CrossRef](#)]
2. Duerinck, J.; Du Four, S.; Vandervorst, F.; D'Haene, N.; Le Mercier, M.; Michotte, A.; Van Binst, A.M.; Everaert, H.; Salmon, I.; Bouttens, F.; et al. Randomized phase II study of axitinib versus physicians best alternative choice of therapy in patients with recurrent glioblastoma. *J. Neurooncol.* **2016**, *128*, 147–155. [[CrossRef](#)]
3. Turajlic, S.; Sottoriva, A.; Graham, T.; Swanton, C. Resolving genetic heterogeneity in cancer. *Nat. Rev. Genet.* **2019**, *20*, 404–416. [[CrossRef](#)] [[PubMed](#)]
4. Wang, D.; Niu, X.; Wang, Z.; Song, C.L.; Huang, Z.; Chen, K.N.; Duan, J.; Bai, H.; Xu, J.; Zhao, J.; et al. Multiregion Sequencing Reveals the Genetic Heterogeneity and Evolutionary History of Osteosarcoma and Matched Pulmonary Metastases. *Cancer Res.* **2019**, *79*, 7–20. [[CrossRef](#)]
5. Gianferante, D.M.; Mirabello, L.; Savage, S.A. Germline and somatic genetics of osteosarcoma—Connecting aetiology, biology and therapy. *Nat. Rev. Endocrinol.* **2017**, *13*, 480–491. [[CrossRef](#)] [[PubMed](#)]
6. Zhou, Y.; Yang, D.; Yang, Q.; Lv, X.; Huang, W.; Zhou, Z.; Wang, Y.; Zhang, Z.; Yuan, T.; Ding, X.; et al. Single-cell RNA landscape of intratumoral heterogeneity and immunosuppressive microenvironment in advanced osteosarcoma. *Nat. Commun.* **2020**, *11*, 6322. [[CrossRef](#)] [[PubMed](#)]
7. Reya, T.; Morrison, S.J.; Clarke, M.F.; Weissman, I.L. Stem cells, cancer, and cancer stem cells. *Nature* **2001**, *414*, 105–111. [[CrossRef](#)]
8. Brown, H.K.; Tellez-Gabriel, M.; Heymann, D. Cancer stem cells in osteosarcoma. *Cancer Lett.* **2017**, *386*, 189–195. [[CrossRef](#)]
9. Najafi, M.; Farhood, B.; Mortezaee, K. Cancer stem cells (CSCs) in cancer progression and therapy. *J. Cell. Physiol.* **2019**, *234*, 8381–8395. [[CrossRef](#)]
10. Gibbs, C.P.; Kukekov, V.G.; Reith, J.D.; Tchigrinova, O.; Suslov, O.N.; Scott, E.W.; Ghivizzani, S.C.; Ignatova, T.N.; Steindler, D.A. Stem-like cells in bone sarcomas: Implications for tumorigenesis. *Neoplasia* **2005**, *7*, 967–976. [[CrossRef](#)]
11. Kovac, M.; Blattmann, C.; Ribí, S.; Smida, J.; Mueller, N.S.; Engert, F.; Castro-Giner, F.; Weischenfeldt, J.; Kovacova, M.; Krieg, A.; et al. Exome sequencing of osteosarcoma reveals mutation signatures reminiscent of BRCA deficiency. *Nat. Commun.* **2015**, *6*, 8940. [[CrossRef](#)] [[PubMed](#)]
12. Butler, A.; Hoffman, P.; Smibert, P.; Papalexi, E.; Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **2018**, *36*, 411–420. [[CrossRef](#)] [[PubMed](#)]
13. Korsunsky, I.; Millard, N.; Fan, J.; Slowikowski, K.; Zhang, F.; Wei, K.; Baglaenko, Y.; Brenner, M.; Loh, P.R.; Raychaudhuri, S. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **2019**, *16*, 1289–1296. [[CrossRef](#)] [[PubMed](#)]
14. Johnson, W.E.; Li, C.; Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **2007**, *8*, 118–127. [[CrossRef](#)] [[PubMed](#)]
15. Zeng, D.; Li, M.; Zhou, R.; Zhang, J.; Sun, H.; Shi, M.; Bin, J.; Liao, Y.; Rao, J.; Liao, W. Tumor Microenvironment Characterization in Gastric Cancer Identifies Prognostic and Immunotherapeutically Relevant Gene Signatures. *Cancer Immunol. Res.* **2019**, *7*, 737–750. [[CrossRef](#)] [[PubMed](#)]
16. Wagner, G.P.; Kin, K.; Lynch, V.J. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* **2012**, *131*, 281–285. [[CrossRef](#)] [[PubMed](#)]
17. Qiu, X.; Mao, Q.; Tang, Y.; Wang, L.; Chawla, R.; Pliner, H.A.; Trapnell, C. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **2017**, *14*, 979–982. [[CrossRef](#)] [[PubMed](#)]
18. Monti, S.; Tamayo, P.; Mesirov, J.; Golub, T. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. *Mach. Learn.* **2003**, *52*, 91–118. [[CrossRef](#)]
19. Kapp, A.V.; Tibshirani, R. Are clusters found in one dataset present in another dataset? *Biostatistics* **2006**, *8*, 9–31. [[CrossRef](#)]
20. Yoshihara, K.; Shahmoradgoli, M.; Martinez, E.; Vegesna, R.; Kim, H.; Torres-Garcia, W.; Trevino, V.; Shen, H.; Laird, P.W.; Levine, D.A.; et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **2013**, *4*, 2612. [[CrossRef](#)]
21. Newman, A.M.; Steen, C.B.; Liu, C.L.; Gentles, A.J.; Chaudhuri, A.A.; Scherer, F.; Khodadoust, M.S.; Esfahani, M.S.; Luca, B.A.; Steiner, D.; et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **2019**, *37*, 773–782. [[CrossRef](#)]
22. Yu, G.; Wang, L.G.; Han, Y.; He, Q.Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *Omics* **2012**, *16*, 284–287. [[CrossRef](#)] [[PubMed](#)]

23. Wang, H.; Zhou, L. Random survival forest with space extensions for censored data. *Artif. Intell. Med.* **2017**, *79*, 52–61. [[CrossRef](#)] [[PubMed](#)]
24. Blanche, P.; Dartigues, J.F.; Jacqmin-Gadda, H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat. Med.* **2013**, *32*, 5381–5397. [[CrossRef](#)] [[PubMed](#)]
25. Newman, A.M.; Liu, C.L.; Green, M.R.; Gentles, A.J.; Feng, W.; Xu, Y.; Hoang, C.D.; Diehn, M.; Alizadeh, A.A. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **2015**, *12*, 453–457. [[CrossRef](#)]
26. Basu-Roy, U.; Basilico, C.; Mansukhani, A. Perspectives on cancer stem cells in osteosarcoma. *Cancer Lett.* **2013**, *338*, 158–167. [[CrossRef](#)]
27. Maurizi, G.; Verma, N.; Gadi, A.; Mansukhani, A.; Basilico, C. Sox2 is required for tumor development and cancer cell proliferation in osteosarcoma. *Oncogene* **2018**, *37*, 4626–4632. [[CrossRef](#)]
28. Lillo, O.M.; Garcia-Lopez, J.; El, A.I.; Fatima, I.; Khalid, A.B.; Kumpati, J.; Slayden, A.V.; Seagroves, T.N.; Miranda-Carboni, G.A.; Krum, S.A. Activation of Estrogen Receptor Alpha by Decitabine Inhibits Osteosarcoma Growth and Metastasis. *Cancer Res.* **2019**, *79*, 1054–1068. [[CrossRef](#)]
29. Basu-Roy, U.; Seo, E.; Ramanathapuram, L.; Rapp, T.B.; Perry, J.A.; Orkin, S.H.; Mansukhani, A.; Basilico, C. Sox2 maintains self renewal of tumor-initiating cells in osteosarcomas. *Oncogene* **2012**, *31*, 2270–2282. [[CrossRef](#)]
30. Lamb, J.; Crawford, E.D.; Peck, D.; Modell, J.W.; Blat, I.C.; Wrobel, M.J.; Lerner, J.; Brunet, J.P.; Subramanian, A.; Ross, K.N.; et al. The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* **2006**, *313*, 1929–1935. [[CrossRef](#)]
31. Gyurko, D.M.; Veres, D.V.; Modos, D.; Lenti, K.; Korcsmaros, T.; Csermely, P. Adaptation and learning of molecular networks as a description of cancer development at the systems-level: Potential use in anti-cancer therapies. *Semin. Cancer Biol.* **2013**, *23*, 262–269. [[CrossRef](#)]
32. Lin, Y.H.; Jewell, B.E.; Gingold, J.; Lu, L.; Zhao, R.; Wang, L.L.; Lee, D.F. Osteosarcoma: Molecular Pathogenesis and iPSC Modeling. *Trends Mol. Med.* **2017**, *23*, 737–755. [[CrossRef](#)] [[PubMed](#)]
33. Lee, D.F.; Su, J.; Kim, H.S.; Chang, B.; Papatsenko, D.; Zhao, R.; Yuan, Y.; Gingold, J.; Xia, W.; Darr, H.; et al. Modeling familial cancer with induced pluripotent stem cells. *Cell* **2015**, *161*, 240–254. [[CrossRef](#)] [[PubMed](#)]
34. Gambera, S.; Abarrategi, A.; Rodriguez-Milla, M.A.; Mulero, F.; Menendez, S.T.; Rodriguez, R.; Navarro, S.; Garcia-Castro, J. Role of Activator Protein-1 Complex on the Phenotype of Human Osteosarcomas Generated from Mesenchymal Stem Cells. *Stem Cells* **2018**, *36*, 1487–1500. [[CrossRef](#)] [[PubMed](#)]
35. Wang, J.Y.; Wu, P.K.; Chen, P.C.; Lee, C.W.; Chen, W.M.; Hung, S.C. Generation of Osteosarcomas from a Combination of Rb Silencing and c-Myc Overexpression in Human Mesenchymal Stem Cells. *Stem Cells Transl. Med.* **2017**, *6*, 512–526. [[CrossRef](#)] [[PubMed](#)]
36. Mohseny, A.B.; Szuhai, K.; Romeo, S.; Buddingh, E.P.; Briaire-De, B.I.; de Jong, D.; van Pel, M.; Cleton-Jansen, A.M.; Hogendoorn, P.C. Osteosarcoma originates from mesenchymal stem cells in consequence of aneuploidization and genomic loss of Cdkn2. *J. Pathol.* **2009**, *219*, 294–305. [[CrossRef](#)] [[PubMed](#)]
37. Chen, Y.; Wang, T.; Huang, M.; Liu, Q.; Hu, C.; Wang, B.; Han, D.; Chen, C.; Zhang, J.; Li, Z.; et al. MAFB Promotes Cancer Stemness and Tumorigenesis in Osteosarcoma through a Sox9-Mediated Positive Feedback Loop. *Cancer Res.* **2020**, *80*, 2472–2483. [[CrossRef](#)]
38. Sharma, A.; Cinti, C.; Capobianco, E. Multitype Network-Guided Target Controllability in Phenotypically Characterized Osteosarcoma: Role of Tumor Microenvironment. *Front. Immunol.* **2017**, *8*, 918. [[CrossRef](#)]
39. Williams, S.A.; Maecker, H.L.; French, D.M.; Liu, J.; Gregg, A.; Silverstein, L.B.; Cao, T.C.; Carano, R.A.; Dixit, V.M. USP1 deubiquitinates ID proteins to preserve a mesenchymal stem cell program in osteosarcoma. *Cell* **2011**, *146*, 918–930. [[CrossRef](#)]
40. Adhikari, A.S.; Agarwal, N.; Wood, B.M.; Porretta, C.; Ruiz, B.; Pochampally, R.R.; Iwakuma, T. CD117 and Stro-1 identify osteosarcoma tumor-initiating cells associated with metastasis and drug resistance. *Cancer Res.* **2010**, *70*, 4602–4612. [[CrossRef](#)]
41. Subramaniam, D.; Angulo, P.; Ponnurangam, S.; Dandawate, P.; Ramamoorthy, P.; Srinivasan, P.; Iwakuma, T.; Weir, S.J.; Chastain, K.; Anant, S. Suppressing STAT5 signaling affects osteosarcoma growth and stemness. *Cell Death Dis.* **2020**, *11*, 149. [[CrossRef](#)] [[PubMed](#)]
42. Zhou, Z.; Li, Y.; Kuang, M.; Wang, X.; Jia, Q.; Cao, J.; Hu, J.; Wu, S.; Wang, Z.; Xiao, J. The CD24(+) cell subset promotes invasion and metastasis in human osteosarcoma. *EBioMedicine* **2020**, *51*, 102598. [[CrossRef](#)] [[PubMed](#)]
43. Wei, H.; Fu, P.; Yao, M.; Chen, Y.; Du, L. Breast cancer stem cells phenotype and plasma cell-predominant breast cancer independently indicate poor survival. *Pathol. Res. Pract.* **2016**, *212*, 294–301. [[CrossRef](#)] [[PubMed](#)]
44. Wang, H.; Wang, L.; Song, Y.; Wang, S.; Huang, X.; Xuan, Q.; Kang, X.; Zhang, Q. CD44(+)/CD24(-) phenotype predicts a poor prognosis in triple-negative breast cancer. *Oncol. Lett.* **2017**, *14*, 5890–5898. [[CrossRef](#)]
45. Cortes, J.E.; Gutzmer, R.; Kieran, M.W.; Solomon, J.A. Hedgehog signaling inhibitors in solid and hematological cancers. *Cancer Treat. Rev.* **2019**, *76*, 41–50. [[CrossRef](#)]
46. Jaiswal, S.; Jamieson, C.H.; Pang, W.W.; Park, C.Y.; Chao, M.P.; Majeti, R.; Traver, D.; van Rooijen, N.; Weissman, I.L. CD47 is upregulated on circulating hematopoietic stem cells and leukemia cells to avoid phagocytosis. *Cell* **2009**, *138*, 271–285. [[CrossRef](#)]
47. Chan, K.S.; Espinosa, I.; Chao, M.; Wong, D.; Ailles, L.; Diehn, M.; Gill, H.; Presti, J.J.; Chang, H.Y.; van de Rijn, M.; et al. Identification, molecular characterization, clinical prognosis, and therapeutic targeting of human bladder tumor-initiating cells. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 14016–14021. [[CrossRef](#)]

48. Advani, R.; Flinn, I.; Popplewell, L.; Forero, A.; Bartlett, N.L.; Ghosh, N.; Kline, J.; Roschewski, M.; Lacasce, A.; Collins, G.P.; et al. CD47 Blockade by Hu5F9-G4 and Rituximab in Non-Hodgkin's Lymphoma. *N. Engl. J. Med.* **2018**, *379*, 1711–1721. [[CrossRef](#)]
49. Lee, N.; Smolarz, A.J.; Olson, S.; David, O.; Reiser, J.; Kutner, R.; Daw, N.C.; Prockop, D.J.; Horwitz, E.M.; Gregory, C.A. A potential role for Dkk-1 in the pathogenesis of osteosarcoma predicts novel diagnostic and treatment strategies. *Br. J. Cancer* **2007**, *97*, 1552–1559. [[CrossRef](#)]
50. Klemptner, S.J.; Bendell, J.C.; Villaflor, V.M.; Tenner, L.L.; Stein, S.M.; Rottman, J.B.; Naik, G.S.; Sirard, C.A.; Kagey, M.H.; Chaney, M.F.; et al. Safety, Efficacy, and Biomarker Results from a Phase Ib Study of the Anti-DKK1 Antibody DKN-01 in Combination with Pembrolizumab in Advanced Esophagogastric Cancers. *Mol. Cancer Ther.* **2021**. [[CrossRef](#)]
51. Iyer, S.P.; Beck, J.T.; Stewart, A.K.; Shah, J.; Kelly, K.R.; Isaacs, R.; Bilic, S.; Sen, S.; Munshi, N.C. A Phase IB multicentre dose-determination study of BHQ880 in combination with anti-myeloma therapy and zoledronic acid in patients with relapsed or refractory multiple myeloma and prior skeletal-related events. *Br. J. Haematol.* **2014**, *167*, 366–375. [[CrossRef](#)] [[PubMed](#)]
52. Welsh, E.A.; Stewart, P.A.; Kuenzi, B.M.; Eschrich, J.A. Escape Excel: A tool for preventing gene symbol and accession conversion errors. *PLoS ONE* **2017**, *12*, e185207. [[CrossRef](#)] [[PubMed](#)]