



Application of Immune Infiltration Signature and Machine Learning Model in the Differential Diagnosis and Prognosis of Bone-Related Malignancies

Guo-Qi Li^{1,2,3†}, Yi-Kai Wang^{1,2†}, Hao Zhou^{1,2†}, Lin-Guang Jin^{1,2}, Chun-Yu Wang^{1,2}, Mugahed Albahde⁴, Yan Wu^{1,2}, Heng-Yuan Li^{1,2}, Wen-Kan Zhang^{1,2}, Bing-Hao Li^{1,2,3*} and Zhao-Ming Ye^{1,2,3*}

OPEN ACCESS

Edited by:

Lei Deng,
Central South University, China

Reviewed by:

Emenike K. Onyido,
Swansea University Medical School,
United Kingdom
Yu Wu,
Sichuan University, China

*Correspondence:

Bing-Hao Li
libinghao@zju.edu.cn
Zhao-Ming Ye
yezhaoming@zju.edu.cn

† These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Molecular Medicine,
a section of the journal
Frontiers in Cell and Developmental
Biology

Received: 17 November 2020

Accepted: 15 March 2021

Published: 15 April 2021

Citation:

Li G-Q, Wang Y-K, Zhou H,
Jin L-G, Wang C-Y, Albahde M, Wu Y,
Li H-Y, Zhang W-K, Li B-H and
Ye Z-M (2021) Application of Immune
Infiltration Signature and Machine
Learning Model in the Differential
Diagnosis and Prognosis
of Bone-Related Malignancies.
Front. Cell Dev. Biol. 9:630355.
doi: 10.3389/fcell.2021.630355

¹ Department of Orthopedics, Musculoskeletal Tumor Center, The Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China, ² Orthopedics Research Institute of Zhejiang University, Hangzhou, China, ³ Key Laboratory of Motor System Disease Research and Precision Therapy of Zhejiang Province, Hangzhou, China, ⁴ Department of Hepatobiliary and Pancreatic Surgery, School of Medicine, The Second Affiliated Hospital, Zhejiang University, Hangzhou, China

Bone-related malignancies, such as osteosarcoma, Ewing's sarcoma, multiple myeloma, and cancer bone metastases have similar histological context, but they are distinct in origin and biological behavior. We hypothesize that a distinct immune infiltrative microenvironment exists in these four most common malignant bone-associated tumors and can be used for tumor diagnosis and patient prognosis. After sample cleaning, data integration, and batch effect removal, we used 22 publicly available datasets to draw out the tumor immune microenvironment using the ssGSEA algorithm. The diagnostic model was developed using the random forest. Further statistical analysis of the immune microenvironment and clinical data of patients with osteosarcoma and Ewing's sarcoma was carried out. The results suggested significant differences in the microenvironment of bone-related tumors, and the diagnostic accuracy of the model was higher than 97%. Also, high infiltration of multiple immune cells in Ewing's sarcoma was suggestive of poor patient prognosis. Meanwhile, increased infiltration of macrophages and B cells suggested a better prognosis for patients with osteosarcoma, and effector memory CD8 T cells and type 2 T helper cells correlated with patients' chemotherapy responsiveness and tumor metastasis. Our study revealed that the random forest diagnostic model based on immune infiltration can accurately perform the differential diagnosis of bone-related malignancies. The immune microenvironment of osteosarcoma and Ewing's sarcoma has an important impact on patient prognosis. Suppressing the highly inflammatory environment of Ewing's sarcoma and promoting macrophage and B cell infiltration may have good potential to be a novel adjuvant treatment option for osteosarcoma and Ewing's sarcoma.

Keywords: osteosarcoma, Ewing's sarcoma, multiple myeloma, bone metastases, immune microenvironment, ssGSEA, random forest

INTRODUCTION

The tumor immune microenvironment shapes tumorigenesis and development (Wu and Dai, 2017) and the diagnosis, treatment, and prognosis of tumor patients (Chen and Mellman, 2013; Clara et al., 2020). Tumor immune infiltration characteristics can be used to classify tumors into immune subtypes and potentially influence the choice of treatment options for patients (Angell et al., 2020). Bone tissue plays a major role in immune microenvironmental homeostasis, the bone microenvironment is an ideal fertile soil for both primary and secondary tumors to seed. Many malignant tumors occur in the bone, including osteosarcoma (OS), Ewing's sarcoma (EW), and multiple myeloma (MM), with an incident rate of 4.0–5.4 per million, 1.5 per million (Ottaviani and Jaffe, 2009; Grünewald et al., 2018), and 43 per million (Kyle et al., 2004). Meanwhile, cancer metastasis remains the major cause of cancer-related mortality, and multiple cancers predispose to bone metastasis; approximately 70% of breast and prostate cancers are associated with bone metastases (Macedo et al., 2017; Fornetti et al., 2018; Hernandez et al., 2018).

In current clinical practice, pathology, immunohistochemistry, and radiography are essential to establish diagnosis and differential diagnosis for bone-related malignant diseases, yet difficulties often accompany the diagnostic process. By far, studies have indicated that there is a high rate of misdiagnosis and missed diagnosis of OS based on imaging and medical history, especially in elderly patients, with an incident rate of 23–43% (Chen and Mellman, 2013; Wu and Dai, 2017; Clara et al., 2020). In EW, Wurtz et al. reported that the average delay in diagnosis of EW was about 10 months (Wurtz et al., 1999), and Widhe et al. found a misdiagnosis rate of 80.77% in EW (Wurtz et al., 1999; Widhe et al., 2007). It is also often difficult to diagnose bone metastatic carcinoma with an unknown primary focus, and confirmation of diagnosis usually requires a combination of PET-CT and other tests (Guzik and Barańska, 2013). The immune infiltration characteristics of different tumors may provide an important tool for the differential diagnosis of tumors. Existing studies suggest that MM is a malignant tumor with a predominance of B cells (Kastrinakis et al., 2000). In contrast, cancer bone metastasis is usually associated with extensive immunosuppression (Liu and Cao, 2016), EW presents extensive inflammatory features (Durbin et al., 1998; Jordanov et al., 2009; Huang et al., 2013), yet the main components of OS are osteoblasts and osteoclasts, which originate mainly from the myeloid cell system, and macrophages make up the highest percentage of its immune microenvironment (Zhang et al., 2020). Besides, the immune microenvironment of bone-associated malignancies also influences the biological behavior of the tumor and the patient's prognosis. B cells, MDSC, and other cells in the MM microenvironment can facilitate its migration and proliferation by secreting a variety of cytokines (Kastrinakis et al., 2000; Kumar et al., 2017); targeted intervention of tumor microenvironment B cells has significant efficacy in MM treatment (Shah et al., 2020). For OS, clinical studies have shown that patients with increased infiltration of CD163-positive macrophages have a better prognosis and that stimulation of

monocyte and macrophage infiltration in OS by mifamurtide can prolong the disease-free survival of patients (Buddingh et al., 2011; Gomez-Brouchet et al., 2017). In addition, the immune microenvironment may also contribute to the cancer bone metastasis process (Janiszewska et al., 2019). In breast cancer, bone metastases can be significantly suppressed by restoring the inherent IFN signaling pathway in tumor cells and activating both the innate and acquired immune responses (Bidwell et al., 2012). Therefore, the use of the immune microenvironment in bone-associated tumors for diagnosis and prognosis has significant clinical potential.

There are various methods to study the bone immune microenvironment, such as single-cell sequencing, CYTOF, and computer-assisted algorithms such as CIBERSORT and single-sample gene set enrichment analysis (ssGSEA). Among them, ssGSEA's unique non-parametric algorithm enables us to perform immuno-infiltration microenvironmental analysis across batches of independent samples and makes full use of the existing microarray and sequencing results. A large number of bone-related tumor microarrays and sequencing datasets have been published, including MM, OS, EW, and prostate cancer bone metastases (BM), making the development of immune infiltration-based differential diagnostic models very cost-effective and with great potential for application. In addition, the rapid development of machine learning algorithms has provided a unique opportunity for molecular-based cancer differential diagnosis. However, no studies to date have examined whether immune microenvironment profiles can be used for the differential diagnosis of bone-related tumors and used for prognostic analysis across datasets.

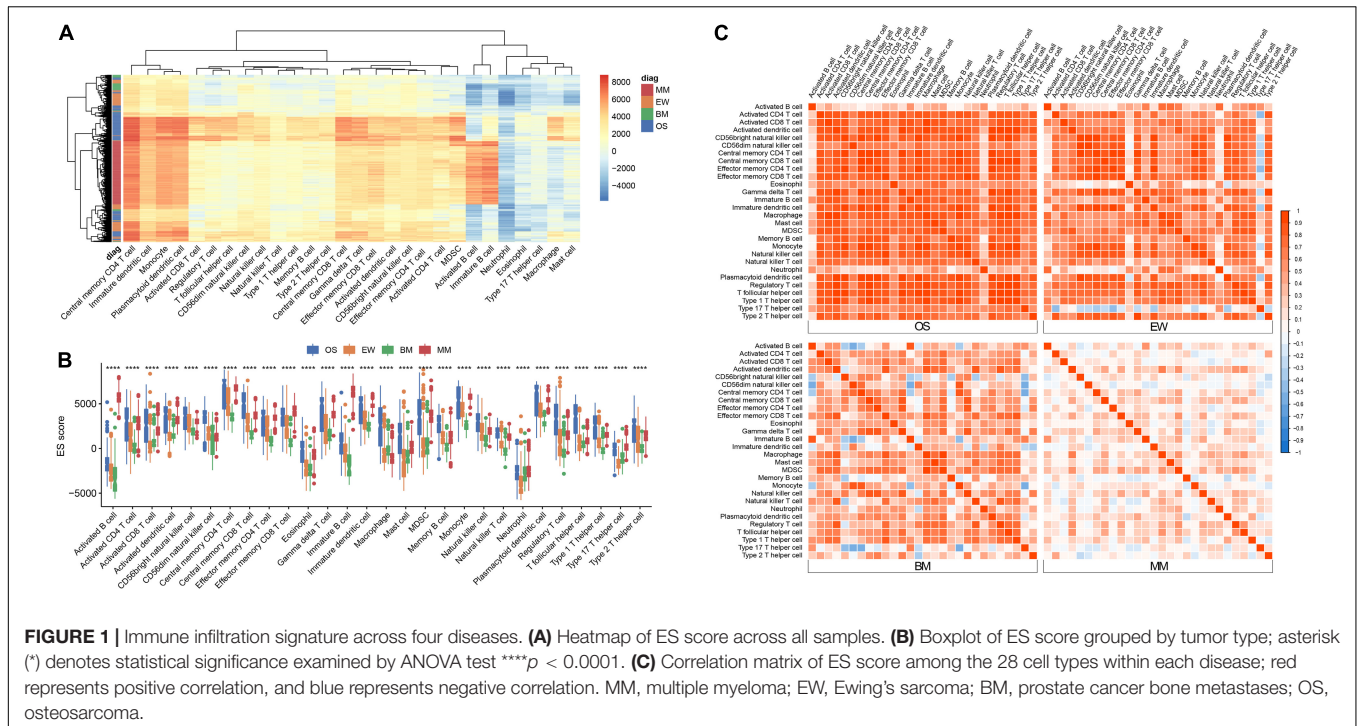
We hypothesize that immune microenvironment profiles can be used in the differential diagnosis of bone-associated tumors and have a significant prognosis for patients with bone-associated malignancies. Therefore, this study intends to use the ssGSEA algorithm for the immune microenvironment profiling of four tumors, combined with random forest machine learning methods to develop a bone-related tumor differential diagnosis model. On this basis, we will explore the influence of the immune microenvironment on the prognosis of tumor patients. It will provide new insight and theoretical basis for the diagnosis and treatment of related tumors.

RESULTS

Immune Infiltration Profiling

After downloading the corresponding datasets and standardizing the data, we removed the cross-platform batch effects *via* the ComBat method in R. Thereafter, the immune enrichment scores (ES score) of 28 immune cells in each of the 1,459 samples were obtained using the ssGSEA algorithm. Unsupervised clustering was performed using Euclidean distances (Figure 1A); the result showed the 28 types of immune cells that are prevalent in all four types of tumors, and similar enrichment profiles can be observed within and across tumor types.

B cells are known to predominate in MM and can be used as a reliable positive control for cross-tumor comparisons.



To clarify the differences in the infiltration of each immune cell in different tumors, we used ANOVA for comparison (Figure 1B). As predicted, the ES score of B cells was significantly elevated in MM, but not in OS, EW, or BM. CD56dim NK cells, macrophages, and Treg cells are mostly enriched in OS. BM has relatively low infiltration of various T cells, B cells, and plasmacytoid dendritic cells, which is consistent with the prevailing knowledge of the suppressive immune microenvironment in bone metastatic cancer.

Correlation analysis revealed the presence of distinct immune cell correlation signatures for four different diseases (Figure 1C). In OS and EW, there is a general positive correlation between various immune cells. In MM, the immune cells showed weak or no correlation. Meanwhile, the BM correlation matrix reveals a more complex immune cell interaction network. Moreover, the pair plot was drawn based on ES score in each sample, and the result showed that MM can be clearly distinguished from others based on activated B cell and/or immature B cell enrichment score. Additionally, clear BM, OS, and EW clusters can be observed (Figure 2 and Supplementary Figure 1). These results suggest that the four diseases have significant differences in immune microenvironmental infiltration and are potentially susceptible to differential diagnosis by their immune microenvironmental profiles.

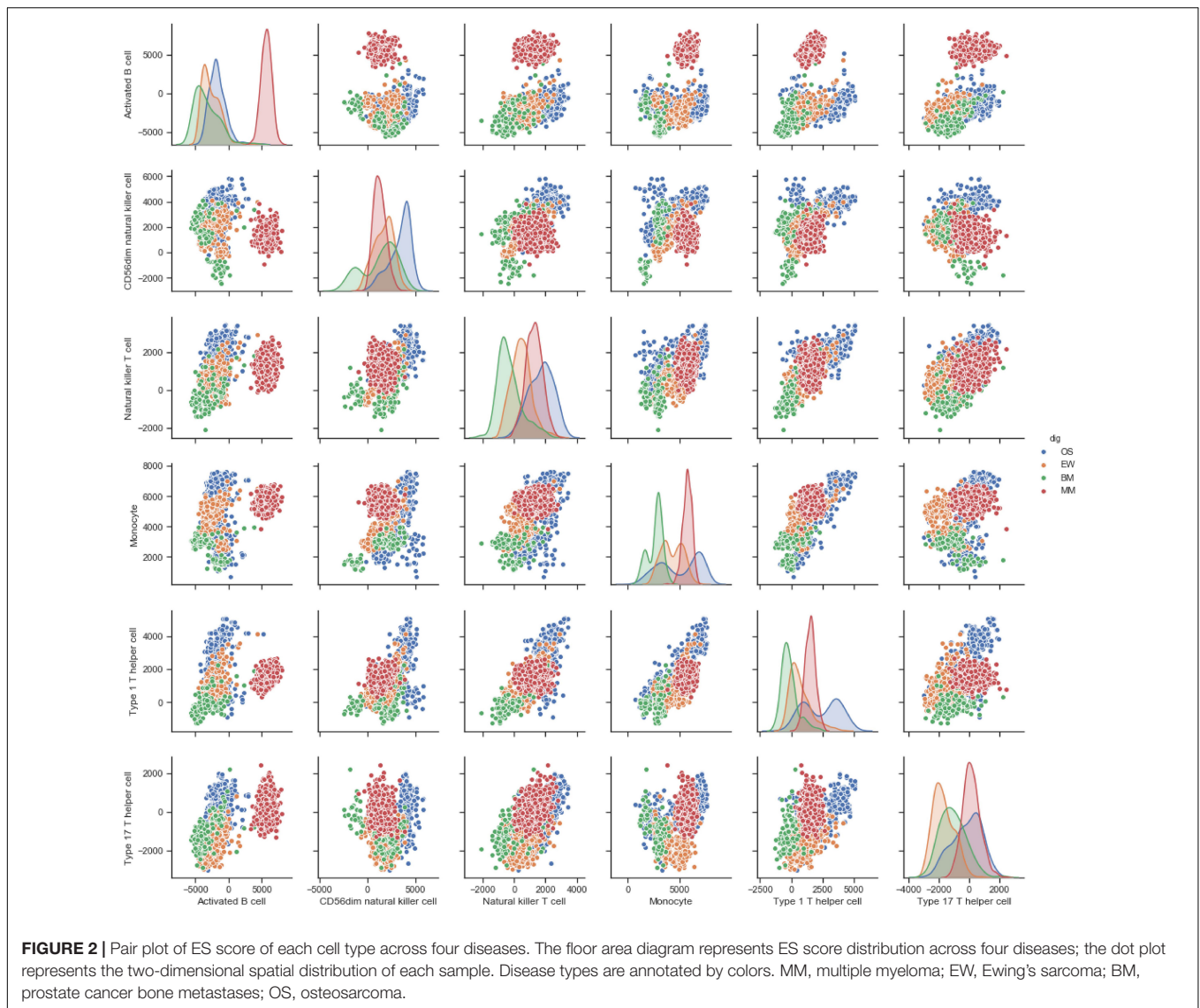
Development of the Diagnostic Model

To develop a multicategorical differential diagnostic model, we used a random forest algorithm to construct the model from a training set of 998 samples. To initially check the feasibility of constructing a classification model, MDS plots were drawn. The result showed that most of the samples can be distinguished

base on the immune infiltration profile (Figure 3A). During model tuning, at $n_{tree} = 150$ and $m_{try} = 5$, the diagnostic model had a stable performance with an out-of-bag (OOB) error rate of 2.3% (Figures 3B,C). Continuing to increase the n_{tree} parameter had no significant effect on the diagnostic performance of the model. Meanwhile, increasing the m_{try} parameter will significantly reduce its accuracy. To further clarify the factors that play a key role in the model, the variable importance histogram was drawn (Figure 3D). Among all 28 immune cell enrichment scores, monocyte, CD56dim NK cell, and activated B cell contributed the most to model accuracy, whereas activated B cell, immature B cell, and CD56dim NK cell contributed the most to sample variance.

Validation of the Diagnostic Model

Internal validation and external validation were carried out, respectively. Firstly, a total of 387 samples in the test dataset were used for internal validation. The overall accuracy of the model was 97.42% (95% CI, 95.3–98.75%, $p < 2.2 \times 10^{-9}$), the sensitivity was 0.828, 0.954, 1, and 0.992 on BM, EW, MM, and OS, respectively, and the specificity was 1, 0.997, 0.996, and 0.970 on BM, EW, MM, and OS, respectively. ROC curve showed that the RF model had a significantly better performance compared with single ES score diagnostic models developed with each of the five selected variables (Figures 4A–D). Furthermore, an addition of 74 samples was used for independent external validation (Figures 4E–H). The overall accuracy was 98.65% (95% CI, 92.7–99.97%, $p < 2.2 \times 10^{-16}$), with a sensitivity of 1, 1, 0.941, and 1 and a specificity of 1, 0.984, 1, and 1 on BM, EW, MM, and OS, respectively. Table 1 shows the F_1 score, precision, and



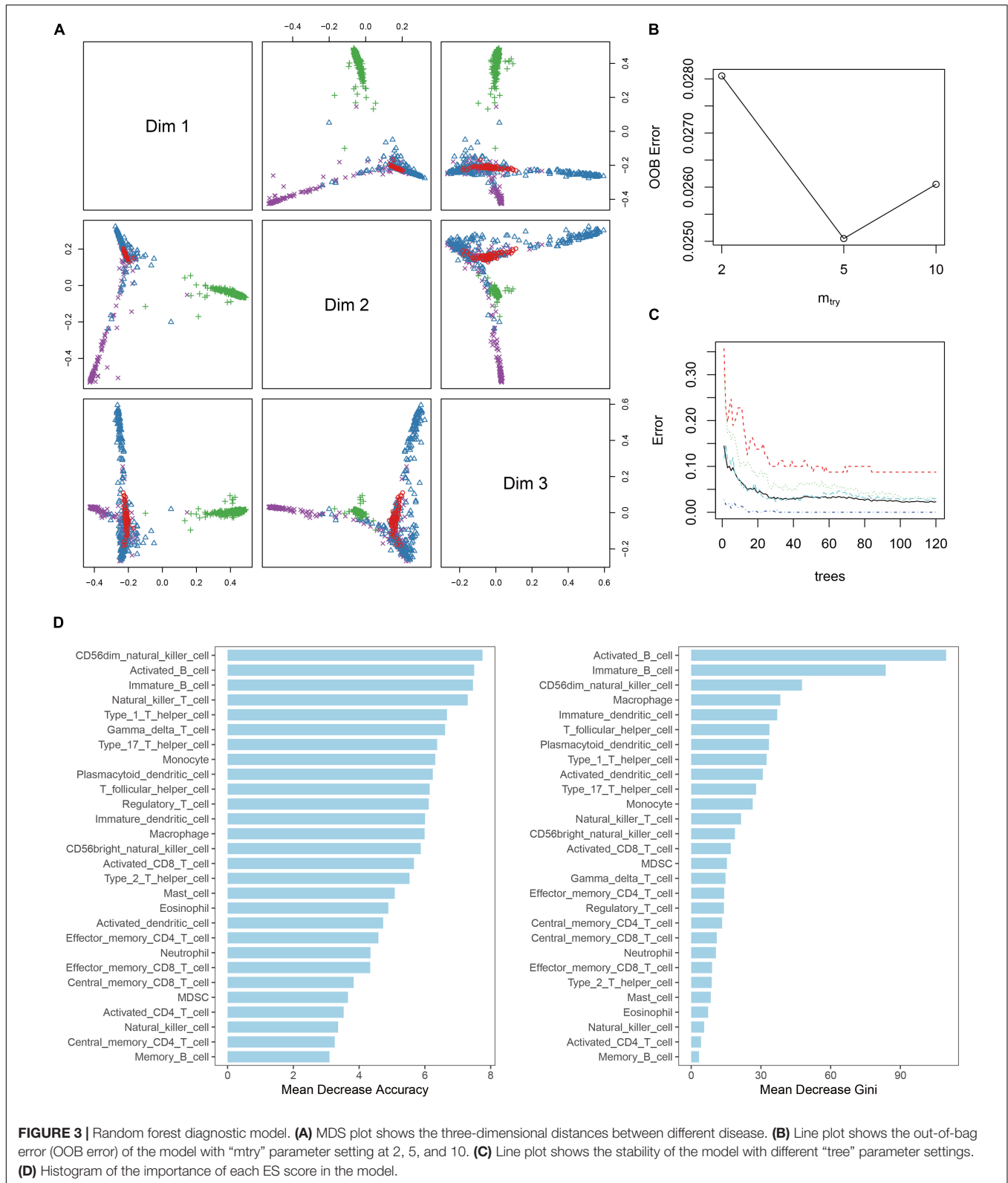
recall values across all diseases in the training, testing, and validation datasets.

Prognostic Value of Immune Infiltration Score in OS and EW

Among all 22 datasets, four OS datasets and two EW datasets contain clinical overall survival data. Cox regression analysis was carried out per each dataset. The prognosis value of immune infiltration scores differs across the four OS datasets, as well as between the two EW datasets (Figure 5A). KM analysis was conducted per each dataset using the median ES score as the cutoff value. The result showed that among four OS datasets, macrophage have a similar correlation with patients' survival (Figures 5B1–B4), as well as activated B cells, yet only activated B cells in dataset GSE39055 has statistical significance (Figures 5C1–C4). A similar result was observed among the two EW datasets (Figures 5D1,D2).

We further combined all datasets per disease and carried out KM analysis using the median ES score of the full sample of the four diseases as the cutoff value, which is more objective and without human intervention. The result indicated that the ES score of both macrophages and activated B cells is positively correlated with OS patient's survival, with a p -value of 0.009 and 0.032, respectively (Figures 5E,F). Also, among the 28 cell types, 17 of them were negatively correlated with EW patients' survival (Figure 5G and Supplementary Figure 2).

The OS dataset was accompanied by more detailed clinical data. To further clarify the relationship between the immune infiltration score and the clinical characteristics of the patients, we performed t -tests for different subgroup conditions. The results showed that there was a difference in CD4T cell infiltration between male and female patients, but no significant difference in the infiltration of other immune cells. In addition, using the age of 18 years as a cutoff, patients were divided into younger and older patients. t -test results showed that younger patients



generally had lower immune infiltration in 26 of the 28 cell types. Furthermore, when grouped according to tumor metastasis status, the comparison revealed significantly lower ES score of

activated B cell, immature B cell, activated DC, effector memory CD8 T cell, MDSC, natural killer T cell, neutrophil, and type 2 T helper cell. Whereas the ES score of the plasmacytoid dendritic

TABLE 1 | Precision, recall, and F_1 score value of the RF model in the training, testing, and validation datasets.

	OS			EW			MM			BM		
	Training	Testing	Validation	Training	Testing	Validation	Training	Testing	Validation	Training	Testing	Validation
Precision	0.951	0.930	1.000	0.969	1.000	0.909	0.997	1.000	1.000	0.987	1.000	1.000
Recall	0.970	1.000	1.000	0.961	0.954	1.000	1.000	1.000	0.941	0.925	1.000	1.000
F_1 score	0.961	0.964	1.000	0.965	0.976	0.952	0.999	1.000	0.970	0.955	0.906	1.000

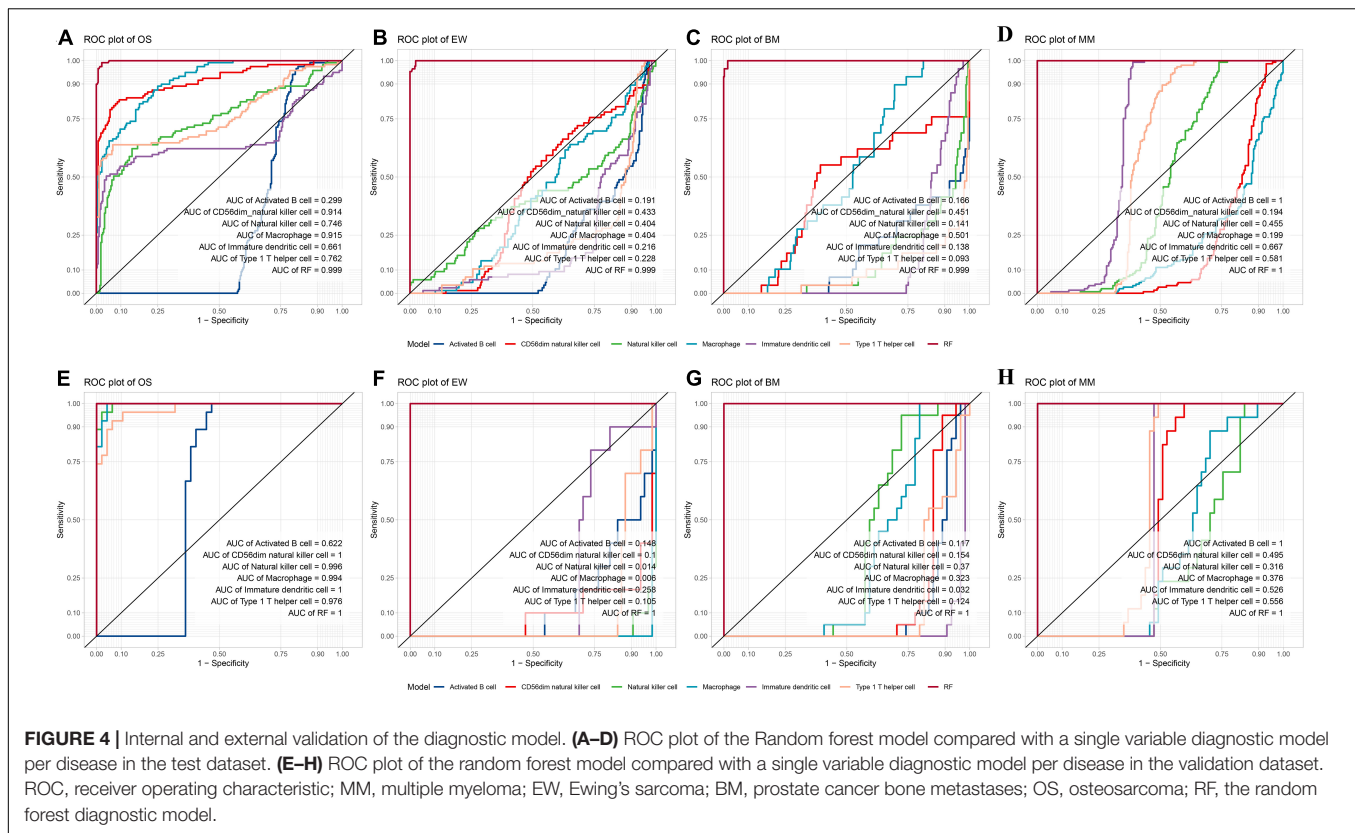
cell was significantly elevated. Finally, patients were divided into Huvos grades I/II and III/IV by using tumor necrosis rate of 90% as the cutoff; the results showed more infiltration of CD8 T cell, mast cell, and type 2 T helper cell in patients with high Huvos grade (Figure 6).

DISCUSSION

Molecular profiling and computer-assisted algorithms are of great significance for fundamental and clinical cancer research (Zhang et al., 2018) and have been widely used for cancer driver gene analysis (Cava et al., 2018), cancer subtyping, and prognosis (Cascione et al., 2013; Junker et al., 2013; Fennell et al., 2019). The immune infiltration microenvironment differs significantly between different malignant tumors, and even in the identical tumor, different immune subtypes exist (Kim et al., 2019). Therefore, the immune infiltration of tumors can potentially be an effective marker for differential diagnosis and prognosis of patients. This study focuses on bone-related malignancies, including OS, EW, MM, and BM; the four diseases have similar tissue conditions, but the origin and progression of these tumors are drastically diverse. Precise differential diagnosis models may be developed based on their immune infiltration profiles. In this study, we obtained 28 immune infiltration scores for each sample across 22 datasets, with batch effect correction and ssGSEA immune infiltration profiling algorithm. The disease diagnosis model was then constructed using Random forest, and evaluated in both internal test datasets and external validation datasets. The association between immune infiltration and patients' clinical characteristics was evaluated by *t*-test, ANOVA test, and survival analysis. Our results suggested a significantly different immune microenvironment among the four malignancies. In addition, this project developed a diagnostic model with an overall accuracy higher than 97%. On this basis, this study correlated the immune microenvironment with patients' clinical data and prognosis and found that the highly inflammatory environment in EW was strongly associated with poor prognosis. The effector memory CD8 T cells and type 2 T helper cells in OS were associated with tumor chemotherapy responsiveness and tumor metastasis, while macrophages and activated B cells were closely associated with patient prognosis.

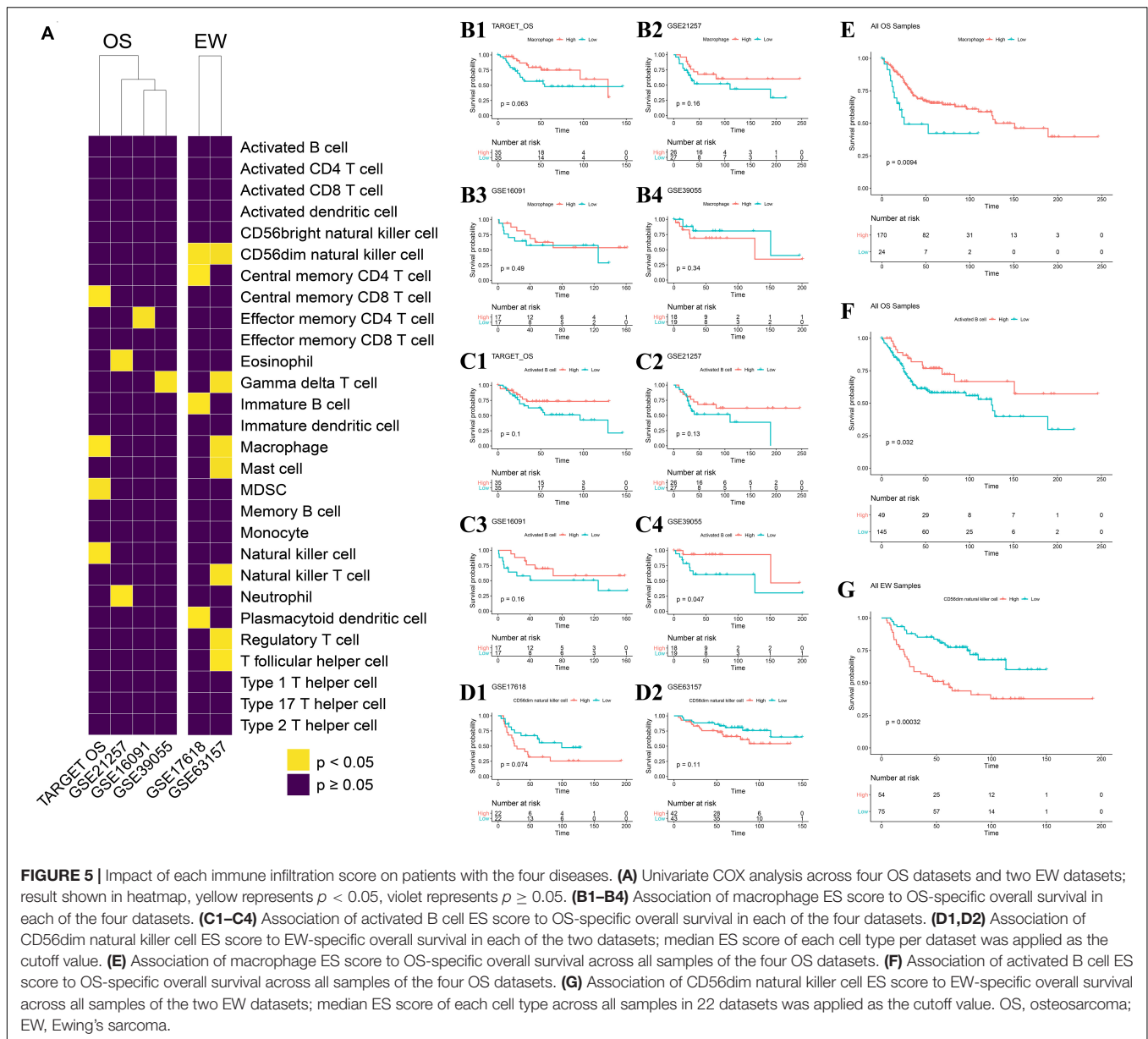
In this study, the overall accuracy of the model based on sample immune infiltration scores was higher than 97%. Studies have shown that ssGSEA is one of the most reliable methods for unsupervised single-sample gene enrichment analysis (Barbie et al., 2009), and the distribution of gene expression ranks inside and outside the gene set. The analysis does not require the entire matrix as input data, and the results are robust (Foroutan et al., 2018). In future applications, the microarray and sequencing data from a single patient can be used for diagnosis. It is an excellent complimentary test and computer-aided diagnostic method.

The immune microenvironment varies significantly among tumors, such as among the four bone-related malignancies included in this study, MM is a malignancy characterized by B cell bone marrow infiltration (Angell et al., 2020), and the



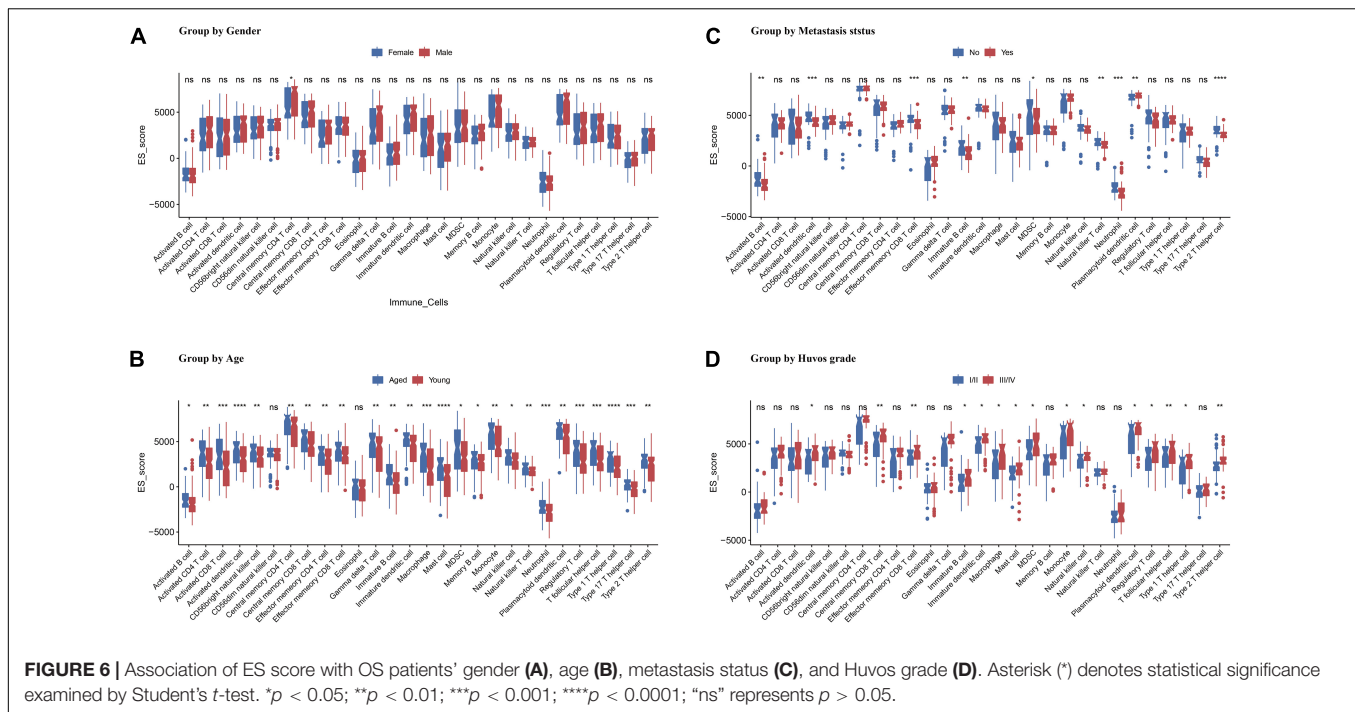
proportion and status of B cells in bone marrow aspirate samples can be used to make an accurate diagnosis of the disease. Bone metastatic carcinoma has an extensive immunosuppressive microenvironment, and EW often presents with osteomyelitis-like changes (Kyle et al., 2004; Ottaviani and Jaffe, 2009; Grünwald et al., 2018), accompanied by high expression of inflammatory factors such as IL-6 (Macedo et al., 2017). Finally, OS originates from the bone marrow mesenchyme, and the tumor potentially originates from the monocyte-macrophage lineage (Han et al., 2019), whose immune microenvironment is dominated by monocyte-macrophages and T cells. The results of the present study also revealed that there was a significant enrichment of B cells in MM, and the accurate diagnosis of MM could be made by B cell ratio, with a specificity and sensitivity of 1, which is consistent with the existing literature. Also, in the bone metastatic lesions of prostate cancer, a significant reduction in the infiltration of multiple immune cells, especially T cells, B cells, and DC cells, was observed, consistent with the reported suppressive immune microenvironment of bone metastatic cancer (Fornetti et al., 2018). For OS, its macrophages were most significantly enriched among the four tumors, accompanied by a higher infiltration of T helper cells, mast cells, etc. Notably, unlike the predominant B cells in MM and the significant enrichment of macrophages and CD56dim natural killer cells in OS, the present study did not identify a dominant cell subpopulation in the EW microenvironment, which indicated that a single diagnostic indicator of the immune microenvironment could not be used in the diagnosis of EW.

Tumor immune infiltration is also closely related to patient treatment and prognosis, and the evidence in the literature indicates that increased immune infiltration of M2-like macrophages, CD8 T, NK, mast cells, B cells, and type 2 T helper cells in the EW microenvironment are significantly associated with poor prognosis in EW patients (Kastrinakis et al., 2000; Guzik and Barańska, 2013; Hernandez et al., 2018). Survival analysis of the individual and combined datasets in this study also suggested this result, with increased scores in 17 of 28 immune infiltration scores all suggesting a poor prognosis. Inflammation may play a key role in the progression of EW and is potentially a factor that promotes cancer progression. In addition, the literature indicates a close relationship between OS and myeloid cells, and *in vitro* studies have shown that macrophages present a pro-cancerous role in the progression and evolution of OS, while some clinical studies have shown that activation of monocyte and macrophage infiltration in OS can significantly inhibit OS progression (Liu and Cao, 2016). Adjuvant treatment with macrophage activators such as mifamurtide can improve the survival of patients with highly malignant OS (Jordanov et al., 2009; Huang et al., 2013), while OS patients with high expression of macrophage-related markers such as CD163 had longer overall survival (Durbin et al., 1998). The results of the present study suggest that survival analysis using the median macrophage immune infiltration of each dataset as the cutoff was statistically insignificant and the prognostic trends were divergent. However, by using the median macrophage immune infiltration from four tumors as the cutoff for survival analysis,



it was found that patients with low macrophage infiltration had a significantly shorter survival. This result also supports the thesis that macrophages play a protective role in OS. Notably, B cells also suggested significant prognostic value in this study and showed statistically significant differences in both the GSE30699 and combined datasets, suggesting that B cells may have a key role in the prognosis and treatment of OS. B cells are the main component of humoral immunity, and previous research on B cells in OS has been extremely rare. However, in recent years, there has been a spate of breakthroughs in B cell research in oncology. Studies have shown that B cell markers and pathological tertiary lymphoid structure have significant prognostic value (Cabrita et al., 2020) and are important predictors of immunotherapy efficacy (Helminck et al., 2020). Petitprez et al. (2020) identified multiple unique tumor

immune subtypes in soft tissue sarcoma and confirmed that B cell infiltration and tertiary lymphoid structures have the strongest correlation with the prognosis of soft tissue sarcoma patients. The present study reveals the prognostic role of B cells in OS at the molecular immune microenvironment level, also suggesting that further studies targeting B cells in OS have a good scientific value. Furthermore, existing studies have shown that age correlates with the level of immune infiltration in patients with OS, and the results of the correlation analysis performed on the integrated dataset in this study are also consistent with this assertion, with younger patients being associated with less immune infiltration (Wu et al., 2020). Moreover, in the present study, effector memory CD8 T cells and type 2 T helper cells were found to be associated with the efficacy of tumor chemotherapy and the development of tumor metastasis. Type 2 T helper cells



in OS are still rarely studied; they have long been considered to play a key role in B cell growth, differentiation, and isotype switching and may exert OS suppressive effects in concert with B cells. As for CD8 cells, a study by our colleagues Li et al. (2018) has shown that OS cells can interfere with CD8 T cell infiltration by inhibiting CXCL12 expression and that increasing CXCL12 expression significantly enhances CD8 T cell infiltration and further enhances the eradication of early lung metastases. Although the role of chemotherapy on CD8 T cells is still controversial in academia (Wang et al., 2019; Deng et al., 2020), current studies generally support the anti-tumor effect presented by CD8 T cells. The mechanism by which CD8 T cells, as well as type 2 T helper cells, affect the efficacy and prognosis of chemotherapy in OS still requires further experimental studies.

In conclusion, this paper employed a large dataset of the four tumors for the description of the immune microenvironment of bone-associated tumors and the construction of differential diagnostic models. For the first time, the immune microenvironment differences between the four tumors are analyzed in terms of molecular and molecular-based immune infiltration. The potential of machine learning for disease differential diagnosis was explored. The model showed high accuracy, specificity, and sensitivity. It is a reliable differential diagnosis model with good performance in both the internal test dataset and the external validation dataset and can be used as an auxiliary differential diagnostic tool based on pathological biopsy to further improve the accuracy of diagnosis. The results of the analysis of the integrated dataset are also consistent with many existing published articles and show good scientific validity, meanwhile suggesting that suppressing the highly inflammatory environment of EW and promoting macrophage and B cell infiltration in OS have good potential

to be exploited as a therapeutic strategy. The limitation of the present study is that the current diagnostic model cannot be used as a direct substitute for examinations such as clinical pathology; further studies need to be conducted prospectively by collecting samples of the four tumors to further validate the accuracy of the model and advance the application in clinical practice. Also, since the study used publicly available datasets, their clinical data registry was incomplete and therefore unable to compare the changes of immune infiltration in the four tumors of different stages, the association between the immune infiltrative microenvironment of tumors, and the pathological and imaging manifestations of tumor samples also still needs further investigation. Meanwhile, further analysis of tumor-infiltrating cells and immune-related genes and proteins in patients' biopsy samples by gene expression analysis, flow cytometry, and *in situ* immunohistochemistry is of great importance for the diagnosis and prognosis of the four diseases and the development of novel targeted therapies.

MATERIALS AND METHODS

Gene Expression Data Preprocessing

A total of 22 datasets were included in the study, including OS, EW, BM, and MM samples, as shown in **Supplementary Table 1**. Microarray gene expression data and clinical information of these datasets were downloaded from the Gene Expression Omnibus (GEO) database¹ and the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) database². All samples

¹<http://www.ncbi.nlm.nih.gov/geo/>

²<https://ocg.cancer.gov/programs/target/>

were carefully recurated to avoid errata when using public datasets (Nicolle et al., 2019). The cell line samples, purified tumor cell samples, duplicated samples, samples of normal tissue, and metastasis sample of sites other than bone were excluded from the datasets, only primary bone tumor and cancer bone metastasis samples were kept for further analysis.

Entrez IDs were used to represent genes across different platforms. If multiple probe sets correspond to the same Entrez ID, the one with the highest mean signal was selected as the expression level of the corresponding gene (Miller et al., 2011). The batch effect between different platforms as well as between different datasets in the same platform were adjusted by the ComBat method (Johnson et al., 2007). All expression data were normalized using $\log_2(\text{expression} + 1)$ per each dataset before batch effect removal.

Immune Infiltration Profiling

The immune signature gene set for 28 immune cell types was obtained from Bindea et al. (2013). The immune infiltration profiling was conducted by the ssGSEA method in R package GSEA (Hänzelmann et al., 2013). Briefly, the gene expression values for a given sample were rank normalized, and an enrichment score was produced using the empirical cumulative distribution functions of the genes in the signature and the remaining genes. This procedure is similar to GSEA, but the list is ranked by absolute expression in each sample. In its final step, within dataset normalization will be carried out routinely to make the results more interpretable. However, since the ssGSEA score normalization within the dataset will introduce unnecessary bias, and study have shown ssGSEA score without normalization is more robust to estimate pathway enrichment (Foroutan et al., 2018), in the present study, ssGSEA scores of 28 immune cell types without normalization (ES score) were used for further analysis.

Construction of the Diagnostic Model

The random forest method is applied to develop the diagnostic model. Random forest is an ensemble learning method for classification and regression tasks. It operates by selecting random samples from a given dataset, then constructing numerous decision trees during training and voting on each prediction, and finally outputting the class of each tree. It is a widely used classifier for multiple classification tasks and can better cope with sample imbalance and model overfitting.

The datasets were divided into a discovery cohort for model training and testing and an independent external validation cohort for model validation. A total of 1,385 and 74 samples were included in the discovery cohort and validation cohort, respectively. The discovery cohort which contains 421 OS samples, 538 MM samples, 317 EW samples, and 109 BM samples were further divided into a training dataset ($n = 998$) and a testing dataset ($n = 387$) in a 7:3 ratio, by stratified sampling. The model was constructed on the training data using randomForest package in R, with parameters of ntree varying from 10 to 500, and mtry from 0 to 15, meanwhile using parameter tuning function to find out the best combination. The model was then tested for performance on the test dataset and

the external validation dataset. ROC curves were visualized using the pROC package.

Statistical and Survival Analysis

Continuous variables were analyzed using Student's *t*-tests or ANOVA test using the "stats" package. Four OS datasets (TARGET OS, GSE21257, GSE16091, GSE29055) and two EW dataset (GSE17618, GSE63157) contain the clinical overall survival data, and the prognostic analyses were performed using Kaplan-Meier survival analysis and Cox univariate analyses using the "survival" package. All analyses were carried out using R v.3.6.4. $p < 0.05$ considered being statistically significant. The result was visualized using the "pheatmap" and "survminer" package.

CONCLUSION

The random forest diagnostic model based on immune infiltration can accurately perform the differential diagnosis of bone-related malignancies. The immune microenvironment of osteosarcoma and Ewing's sarcoma has an important impact on patient prognosis. Suppressing the highly inflammatory environment of Ewing's sarcoma and promoting macrophage and B cell infiltration in osteosarcoma has good potential to be a novel adjuvant treatment option.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

Z-MY and B-HL design and supervised the study. G-QL collected all datasets and wrote the article with the help of Y-KW and HZ. G-QL, Y-KW, HZ, MA, and W-KZ performed experiments and data visualization. H-YL, YW, L-GJ, and C-YW validated the data. G-QL, L-GJ, and C-YW revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Key Research and Development Program of China (2018YFC1105400 to Z-MY) and the National Natural Science Foundation of China (81802684 to B-HL and 81872173 to Z-MY).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcell.2021.630355/full#supplementary-material>

REFERENCES

- Angell, H. K., Bruni, D., Barrett, J. C., Herbst, R., and Galon, J. (2020). The immunoscore: colon cancer and beyond. *Clin. Cancer Res.* 26, 332–339. doi: 10.1158/1078-0432.ccr-18-1851
- Barbie, D. A., Tamayo, P., Boehm, J. S., Kim, S. Y., Moody, S. E., Dunn, I. F., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* 462, 108–112. doi: 10.1038/nature08460
- Bidwell, B. N., Slaney, C. Y., Withana, N. P., Forster, S., Cao, Y., Loi, S., et al. (2012). Silencing of Irf7 pathways in breast cancer cells promotes bone metastasis through immune escape. *Nat. Med.* 18, 1224–1231. doi: 10.1038/nm.2830
- Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenaus, A. C., et al. (2013). Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* 39, 782–795. doi: 10.1016/j.immuni.2013.10.003
- Buddingh, E., Kuijjer, M., Duim, R., Bürger, H., Agelopoulos, K., Myklebost, O., et al. (2011). Tumor-infiltrating macrophages are associated with metastasis suppression in high-grade osteosarcoma: a rationale for treatment with macrophage activating agents. *Clin. Cancer Res.* 17, 2110–2119. doi: 10.1158/1078-0432.ccr-10-2047
- Cabrita, R., Lauss, M., Sanna, A., Donia, M., Skaarup Larsen, M., Mitra, S., et al. (2020). Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature* 577, 561–565. doi: 10.1038/s41586-019-1914-8
- Cascione, L., Gasparini, P., Lovat, F., Carasi, S., Pulvirenti, A., Ferro, A., et al. (2013). Integrated microRNA and mRNA signatures associated with survival in triple negative breast cancer. *PLoS One* 8:e55910. doi: 10.1371/journal.pone.0055910
- Cava, C., Bertoli, G., Colaprico, A., Olsen, C., Bontempi, G., and Castiglioni, I. (2018). Integration of multiple networks and pathways identifies cancer driver genes in pan-cancer analysis. *BMC Genomics* 19:25. doi: 10.1186/s12864-017-4423-x
- Chen, D. S., and Mellman, I. (2013). Oncology meets immunology: the cancer-immunity cycle. *Immunity* 39, 1–10. doi: 10.1016/j.immuni.2013.07.012
- Clara, J. A., Monge, C., Yang, Y., and Takebe, N. (2020). Targeting signalling pathways and the immune microenvironment of cancer stem cells - a clinical update. *Nat. Rev. Clin. Oncol.* 17, 204–232. doi: 10.1038/s41571-019-0293-2
- Deng, C., Xu, Y., Fu, J., Zhu, X., Chen, H., Xu, H., et al. (2020). Reprogramming the tumor immunologic microenvironment using neoadjuvant chemotherapy in osteosarcoma. *Cancer Sci.* 111, 1899–1909. doi: 10.1111/cas.14398
- Durbin, M., Randall, R. L., James, M., Sudilovsky, D., and Zoger, S. (1998). Ewing's sarcoma masquerading as osteomyelitis. *Clin. Orthop. Relat. Res.* 357, 176–185.
- Fennell, L., Dumenil, T., Wockner, L., Hartel, G., Nones, K., Bond, C., et al. (2019). Integrative genome-scale DNA methylation analysis of a large and unselected cohort reveals 5 distinct subtypes of colorectal adenocarcinomas. *Cell. Mol. Gastroenterol. Hepatol.* 8, 269–290. doi: 10.1016/j.jcmgh.2019.04.002
- Fornetti, J., Welm, A. L., and Stewart, S. A. (2018). Understanding the bone in cancer metastasis. *J. Bone Miner. Res.* 33, 2099–2113. doi: 10.1002/jbmr.3618
- Foroutan, M., Bhuva, D. D., Lyu, R., Horan, K., Cursons, J., and Davis, M. J. (2018). Single sample scoring of molecular phenotypes. *BMC Bioinformatics* 19:404. doi: 10.1186/s12859-018-2435-4
- Gomez-Brouchet, A., Illac, C., Gilhodes, J., Bouvier, C., Aubert, S., Guinebretiere, J. M., et al. (2017). CD163-positive tumor-associated macrophages and CD8-positive cytotoxic lymphocytes are powerful diagnostic markers for the therapeutic stratification of osteosarcoma patients: an immunohistochemical analysis of the biopsies from the French OS2006 phase 3 trial. *Oncoimmunology* 6:e1331193. doi: 10.1080/2162402x.2017.1331193
- Grünewald, T. G. P., Cidre-Aranaz, F., Surdez, D., Tomazou, E. M., de Álava, E., Kovar, H., et al. (2018). Ewing sarcoma. *Nat. Rev. Dis. Primers* 4:5.
- Guzik, G., and Barańska, B. (2013). Metastasis of adenocarcinoma of an unknown primary site to the right crus simulating sarcoma - a case report. *Ortop. Traumatol. Rehabil.* 15, 273–279.
- Han, Y., Feng, H., Sun, J., Liang, X., Wang, Z., Xing, W., et al. (2019). Lkb1 deletion in periosteal mesenchymal progenitors induces osteogenic tumors through mTORC1 activation. *J. Clin. Invest.* 129, 1895–1909. doi: 10.1172/jci124590
- Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 14:7. doi: 10.1186/1471-2105-14-7
- Helms, B., Reddy, S., Gao, J., Zhang, S., Basar, R., Thakur, R., et al. (2020). B cells and tertiary lymphoid structures promote immunotherapy response. *Nature* 577, 549–555.
- Hernandez, R. K., Wade, S. W., Reich, A., Pirolli, M., Liede, A., and Lyman, G. H. (2018). Incidence of bone metastases in patients with solid tumors: analysis of oncology electronic medical records in the United States. *BMC Cancer* 18:44. doi: 10.1186/s12885-017-3922-0
- Huang, P., Wu, P., Chen, C., Lee, F., Wu, H., Liu, C., et al. (2013). Osteomyelitis of the femur mimicking bone tumors: a review of 10 cases. *World J. Surg. Oncol.* 11:283. doi: 10.1186/1477-7819-11-283
- Janiszewska, M., Tabassum, D. P., Castaño, Z., Cristea, S., Yamamoto, K. N., Kingston, N. L., et al. (2019). Subclonal cooperation drives metastasis by modulating local and systemic immune microenvironments. *Nat. Cell Biol.* 21, 879–888. doi: 10.1038/s41556-019-0346-x
- Johnson, W. E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8, 118–127. doi: 10.1093/biostatistics/kxj037
- Jordanov, M., Block, J., Gonzalez, A., and Green, N. (2009). Transarticular spread of Ewing sarcoma mimicking septic arthritis. *Pediatr. Radiol.* 39, 381–384. doi: 10.1007/s00247-008-1124-2
- Junker, K., Ficarra, V., Kwon, E. D., Leibovich, B. C., Thompson, R. H., and Oosterwijk, E. (2013). Potential role of genetic markers in the management of kidney cancer. *Eur. Urol.* 63, 333–340. doi: 10.1016/j.eururo.2012.09.040
- Kastrinakis, N. G., Gorgoulis, V. G., Foukas, P. G., Dimopoulos, M. A., and Kittas, C. (2000). Molecular aspects of multiple myeloma. *Ann. Oncol.* 11, 1217–1228.
- Kim, I. S., Gao, Y., Welte, T., Wang, H., Liu, J., Janghorban, M., et al. (2019). Immuno-subtyping of breast cancer reveals distinct myeloid cell profiles and immunotherapy resistance mechanisms. *Nat. Cell Biol.* 21, 1113–1126. doi: 10.1038/s41556-019-0373-7
- Kumar, S. K., Rajkumar, V., Kyle, R. A., van Duin, M., Sonneveld, P., Mateos, M. V., et al. (2017). Multiple myeloma. *Nat. Rev. Dis. Primers* 3:17046.
- Kyle, R., Therneau, T., Rajkumar, S., Larson, D., Plevak, M., and Melton, L. (2004). Incidence of multiple myeloma in Olmsted County, Minnesota: trend over 6 decades. *Cancer* 101, 2667–2674. doi: 10.1002/cncr.20652
- Li, B., Wang, Z., Wu, H., Xue, M., Lin, P., Wang, S., et al. (2018). Epigenetic regulation of CXCL12 plays a critical role in mediating tumor progression and the immune response in osteosarcoma. *Cancer Res.* 78, 3938–3953. doi: 10.1158/0008-5472.can-17-3801
- Liu, Y., and Cao, X. (2016). Characteristics and significance of the pre-metastatic niche. *Cancer Cell* 30, 668–681. doi: 10.1016/j.ccell.2016.09.011
- Macedo, F., Ladeira, K., Pinho, F., Saraiva, N., Bonito, N., Pinto, L., et al. (2017). Bone metastases: an overview. *Oncol. Rev.* 11, 321–321.
- Miller, J. A., Cai, C., Langfelder, P., Geschwind, D. H., Kurian, S. M., Salomon, D. R., et al. (2011). Strategies for aggregating gene expression data: the collapseRows R function. *BMC Bioinformatics* 12:322. doi: 10.1186/1471-2105-12-322
- Nicolle, R., Raffenne, J., Paradis, V., Couvelard, A., de Reynies, A., Blum, Y., et al. (2019). Prognostic biomarkers in pancreatic cancer: avoiding errata when using the TCGA dataset. *Cancers (Basel)* 11:126. doi: 10.3390/cancers11010126
- Ottaviani, G., and Jaffe, N. (2009). The epidemiology of osteosarcoma. *Cancer Treat. Res.* 152, 3–13. doi: 10.1007/978-1-4419-0284-9_1
- Petitprez, F., de Reynies, A., Keung, E., Chen, T., Sun, C., Calderaro, J., et al. (2020). B cells are associated with survival and immunotherapy response in sarcoma. *Nature* 577, 556–560.
- Shah, N., Chari, A., Scott, E., Mezzi, K., and Usmani, S. Z. (2020). B-cell maturation antigen (BCMA) in multiple myeloma: rationale for targeting and current therapeutic approaches. *Leukemia* 34, 985–1005. doi: 10.1038/s41375-020-0734-z
- Wang, J., Hu, C., Wang, J., Shen, Y., Bao, Q., He, F., et al. (2019). Checkpoint blockade in combination with doxorubicin augments tumor cell apoptosis in osteosarcoma. *J. Immunother.* 42, 321–330. doi: 10.1097/cji.0000000000000459

- Widhe, B., Widhe, T., and Bauer, H. C. (2007). Ewing sarcoma of the rib-initial symptoms and clinical features: tumor missed at the first visit in 21 of 26 patients. *Acta Orthop.* 78, 840–844. doi: 10.1080/17453670710014644
- Wu, C. C., Beird, H. C., Andrew Livingston, J., Advani, S., Mitra, A., Cao, S., et al. (2020). Immuno-genomic landscape of osteosarcoma. *Nat. Commun.* 11: 1008.
- Wurtz, L. D., Peabody, T. D., and Simon, M. A. (1999). Delay in the diagnosis and treatment of primary bone sarcoma of the pelvis. *J. Bone Joint Surg. Am* 81, 317–325. doi: 10.2106/00004623-199903000-00003
- Wu, T., and Dai, Y. (2017). Tumor microenvironment and therapeutic response. *Cancer Lett.* 387, 61–68. doi: 10.1016/j.canlet.2016.01.043
- Zhang, C., Zheng, J. H., Lin, Z. H., Lv, H. Y., Ye, Z. M., Chen, Y. P., et al. (2020). Profiles of immune cell infiltration and immune-related genes in the tumor microenvironment of osteosarcoma. *Aging (Albany NY)* 12, 3486–3501. doi: 10.18632/aging.102824
- Zhang, J., Zhang, Z., Wang, Z., Liu, Y., and Deng, L. (2018). Ontological function annotation of long non-coding RNAs through hierarchical multi-label classification. *Bioinformatics* 34, 1750–1757. doi: 10.1093/bioinformatics/btx833

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Li, Wang, Zhou, Jin, Wang, Albahde, Wu, Li, Zhang, Li and Ye. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.