



*Research article*

## **Immune classification of osteosarcoma**

**Trang Le, Sumeyye Su and Leili Shahriyari\***

Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, MA MA 01003-9305, USA

\* **Correspondence:** Email: [lshahriyari@umass.edu](mailto:lshahriyari@umass.edu).

**Abstract:** Tumor immune microenvironment has been shown to be important in predicting the tumor progression and the outcome of treatments. This work aims to identify different immune patterns in osteosarcoma and their clinical characteristics. We use the latest and best performing deconvolution method, CIBERSORTx, to obtain the relative abundance of 22 immune cells. Then we cluster patients based on their estimated immune abundance and study the characteristics of these clusters, along with the relationship between immune infiltration and outcome of patients. We find that abundance of CD8 T cells, NK cells and M1 Macrophages have a positive association with prognosis, while abundance of  $\gamma\delta$  T cells, Mast cells, M0 Macrophages and Dendritic cells have a negative association with prognosis. Accordingly, the cluster with the lowest proportion of CD8 T cells, M1 Macrophages and highest proportion of M0 Macrophages has the worst outcome among clusters. By grouping patients with similar immune patterns, we are also able to suggest treatments that are specific to the tumor microenvironment.

**Keywords:** osteosarcoma; immune pattern; gene expression analysis

---

### **1. Introduction**

Osteosarcoma is the most common type of malignant bone tumors that appears most frequently in children and young adults. It usually starts in the femur, the tibia and the humerus, and less commonly the skull, the jaw or the pelvis [1]. There are some risk factor such as age, height, gender, and heritable syndromes that predispose to for osteosarcoma, however, the cause of osteosarcomas is still not clear [2].

Although neoadjuvant chemotherapy in treatment of osteosarcoma has shown better results, overall survival of patients with metastasis still remains in low rate in the last three decades [3, 4]. Immunotherapy and targeted therapy have recently shown significant improvements for some malignancies [5, 6]. Therefore, they are increasingly attractive treatment options for osteosarcoma

patients. However, they are not effective for many patients and there is no evidence that proves the effectiveness of a specific molecular target [7].

It has been shown in many studies that cancer cells and tumor infiltrating immune cells (TIICs) have an important role in cancer progression and determination of malignant tumor types [8–10]. CD8 T cells, which is an important component of TIICs, has been found to associate with better clinical outcomes and response to immunotherapy in many cancer types [11, 12]. Treatments that includes the antitumor immunocompetence of innate immune cells, which includes NK cells and  $\gamma\delta$  T cells, are found favorable for osteosarcoma tumors [13, 14]. These studies show the importance of finding further biomarkers of tumor cells such as PD-1 and CTLA-4 and of tumor infiltrating immune cells such as T cells and Macrophages to improve new treatment strategies and to choose the best treatment strategy for osteosarcoma patients.

Several computational methods, alternative ways to immunohistochemistry and flow cytometry [15], have been recently developed to derive tumor immune infiltrates using gene expression profiles of bulk tumors. In this study, we use the latest popular deconvolution method [16–20], CIBERSORTx [21], to investigate the immune patterns of tumors and analyze the relationship between immune composition and clinical features of osteosarcoma patients.

There have been many other studies that utilize a deconvolution method to study the tumor microenvironment of osteosarcoma. A number of them use the estimated immune infiltrations calculated using CIBERSORT and/or immune scores calculated using ESTIMATE to find immune-related genes that can predict the prognosis of osteosarcoma [22–25]. Another set of studies find genes with prognostic values by applying Cox model on survival data or performing differentially expressed genes analysis between two groups of interest, and then investigate the relationship between these genes and estimated immune infiltrates [26–30]. Others study the association of immune abundance with clinical information directly [24, 31–34]. Our work falls somewhat into the third category.

Among the studies that directly investigate the relationship between immune infiltrations and clinical information, three of them use TARGET data set [24, 31, 33], one uses data from GSE21257 [32], and one uses data from GSE39058 [34]. Most of these studies use CIBERSORT [24, 32–34], while the other uses ssGSEA and ImmuCellAI along with expression of immune marker genes to get estimated abundance of immune cells [31]. Our study, on the other hand, utilizes the latest and best performing version of deconvolution methods called CIBERSORTx B-mode, and conducts our analyses on both TARGET data set and GSE21257 data set. We also perform K-means clustering using the estimated immune abundance to study the clinical characteristics of different immune patterns in osteosarcoma. Another study has used hierarchical clustering on immune abundance, but they estimate immune abundance using ssGSEA and do not focus on the clinical difference between clusters [35].

## 2. Materials and methods

### 2.1. Data collection and processing

The gene expression data sets in this study are obtained from 2 cohorts: TARGET (cohort 1) and GSE21257 (cohort 2). Cohort 1 includes FPKM normalized RNA-seq data of 88 osteosarcoma patients downloaded from the UCSC Xena web portal and their corresponding clinical data

downloaded from the GDC data portal. Cohort 2 includes microarray data and corresponding clinical features of 53 osteosarcoma samples, downloaded from GEO website. Cohort 2's gene expression data were previously normalized with robust spline normalization before being downloaded.

## 2.2. Relative abundance of immune cells

We calculate the relative frequencies of 22 immune cell types by applying CIBERSORTx [21] B-mode algorithm with immune signature matrix LM22 on gene expression data from both cohort 1 and 2. CIBERSORTx is the latest version of CIBERSORT (Cell type Identification By Estimating Relative Subsets Of RNA Transcripts) [36], where batch correction method Combat has been introduced in the algorithm to reduce the effect of cross-platform variation between gene expression data and signature matrix. There are 2 modes of this algorithm: B-mode and S-mode, whose difference lies in the way batch correction is applied in the algorithm. Our recent study has shown that CIBERSORTx B-mode gives good estimates of immune abundance in both RNA-Seq and microarray data with the use of LM22, and in fact outperforms CIBERSORT and other tumor deconvolution methods [16]. We obtain estimated immune fractions by running the algorithm on their website with 100 permutations. Similar to CIBERSORT, CIBERSORTx outputs a p-value for each deconvolved sample as an indicator of confidence of the results. We use samples with p-value less than 0.05 for analyses in this study.

## 2.3. Identification of immune patterns in osteosarcoma

To calculate the abundance of each cell type, which has several subtypes with a small abundance, we sum the proportions of their subtypes obtained from CIBERSORTx. The abundance of B cells is the summation of naive and memory B cells; NK cell is the summation of resting and activated NK cells; Mast cells is the summation of resting and activated Mast cells; Dendritic cells is the summation of resting and activated Dendritic cells; and CD4 T cells is the summation of follicular helper T cells, regulatory T cells, naive CD4 T cells, resting and activated memory CD4 T cells. We do not combine subtypes of Macrophages because M1 and M2 Macrophages have very different functions and Macrophages make up the majority of immune cells in osteosarcoma.

We then use K-means clustering to identify various immune patterns of osteosarcoma based on the estimated immune cells' abundance. The number of clusters in K-means is determined using elbow method. A t-SNE visualization of the clusters is also included to see how well K-means algorithm distinguishes between samples with different immune patterns.

## 2.4. Immune scores of osteosarcoma tumors

Immune scores of all samples in cohort 1 and 2 are computed from ESTIMATE algorithm [37]. In order to do so, we run *estimate* package locally from R. We also divide all patients from both cohorts into low immune score and high immune score group using the median immune score as cut-off to study the relationship between immune score and survival outcome.

## 2.5. Statistical analysis

Chi-square test is used to analyze the relationship between categorical variables in this study. We employ Mann-Whitney-Wilcoxon test to detect any significant difference between groups of continuous variables, such as immune fractions, gene expression, age and immune score. Pearson

correlation and corresponding p-value are used to study the correlation between different immune infiltrates.

To investigate the impact of immune infiltrates on survival, for each immune cell, we split all patients into high and low abundance group using the median value as cut-off and perform log-rank test to find significant difference in survival between groups. Kaplan-Meier curves are also plotted to visualize the differences between these groups.

All analyses in this study are conducted on all samples in both cohorts 1 and 2, except for metastasis-free survival analysis which is applied only on cohort 2 since cohort 1 does not include data on the time of metastasis development. All statistical analyses and visualizations are carried out in Python.

### 3. Results

#### 3.1. *The most abundant immune cells in osteosarcoma are Macrophages and CD4 T cells*

Results of CIBERSORTx B-mode on gene expression profiles of 141 osteosarcoma patients (cohorts 1 and 2) demonstrates that M0 Macrophages is the most frequent immune cell in osteosarcoma tumors with an average of 40% of total immune cells, followed by M2 Macrophages and CD4 T cells (Figure 1B, C). Unsupervised hierarchical clustering of immune cell fractions shows that most abundant cells tend to be clustered together, as is shown in Figure 1A, where M0 and M2 Macrophages are clustered together and then grouped with CD4 T cells and other immune cells. In addition, the most frequent cells also have the highest variation in abundance across osteosarcoma tumors (Figure 1B).

#### 3.2. *Correlation between immune infiltrates in osteosarcoma*

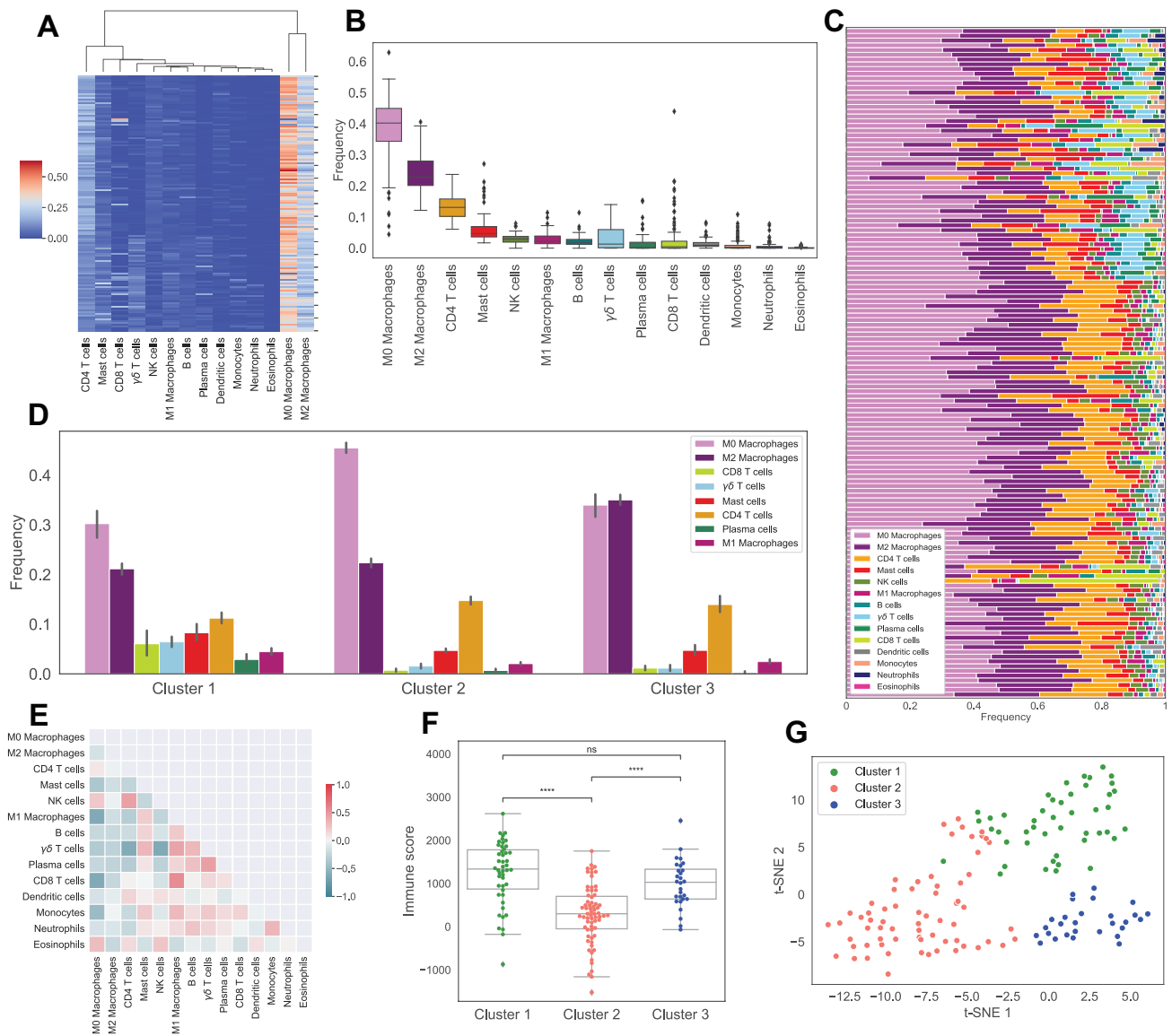
According to the CIBERSORTx B-mode results, abundance of CD8 T cells is negatively correlated with M0 Macrophages and positively correlated with M1 Macrophages with Pearson correlation coefficients of  $-0.62$  and  $0.55$ , with p-values of  $4.8e-16$  and  $1.1e-12$ , respectively (Figure 1E). The proportion of  $\gamma\delta$  T cells is also significantly negatively correlated with CD4 T cells and NK cells (Pearson coefficients of  $-0.6$  and  $-0.62$ , with p-values of  $4.3e-15$  and  $1.8e-16$ ). Interestingly, frequencies of M0 and M1 Macrophages exhibit a negative correlation of  $-0.61$  with p-value  $1e-15$ .

#### 3.3. *There are 3 immune patterns of osteosarcoma*

K-means clustering of immune cell proportions in osteosarcoma tumors indicates the existence of three distinct immune classes (Figure 1D), namely: Cluster 1, which has the highest proportions of CD8 T cells,  $\gamma\delta$  T cells, M1 Macrophages, Mast cells and Plasma cells and the lowest proportion of M0 Macrophages; Cluster 2, in which the percentage of M0 Macrophages is the highest; and Cluster 3, which has the highest percentage of M2 Macrophages. A t-SNE plot of immune cell proportions suggests that K-means clustering algorithm successfully separates osteosarcoma patients with different immune patterns (Figure 1G).

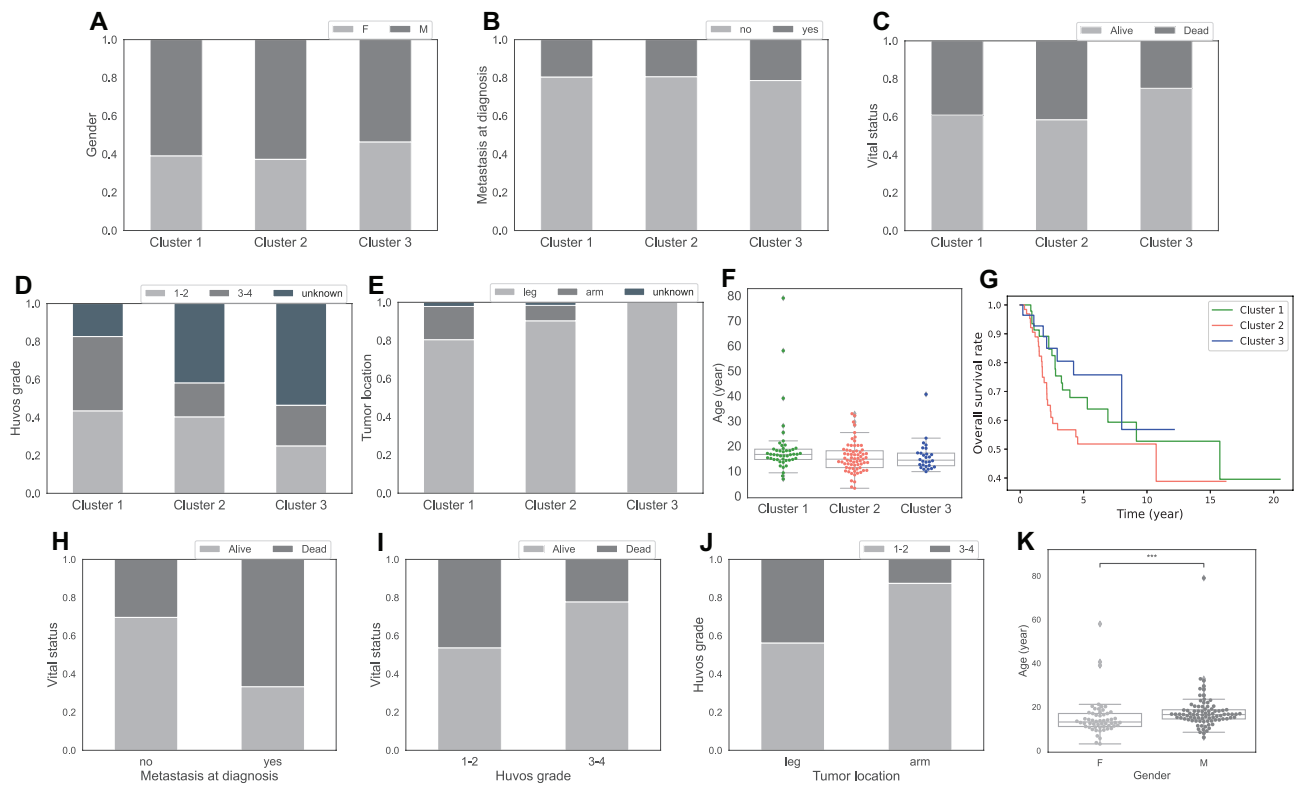
#### 3.4. *Cluster 2 has the worst survival outcome among all clusters*

While there is no significant difference in gender, age and proportion of metastasis at diagnosis between clusters (Figure 2A, B, F), we observe some differences in survival outcomes among clusters.



**Figure 1.** Immune pattern of osteosarcoma. Sub-figure A shows the hierarchical clustering of estimated immune cells' infiltration. Sub-figure B and C display the boxplot and stacked barchart of these immune cells' fractions. Sub-figures D shows the average frequencies of immune cells in 3 clusters obtained from K-means clustering. Sub-figure E indicates the correlation map of immune cell frequencies. Sub-figure F displays the boxplot of ESTIMATE immune scores in 3 clusters, with asterisks indicating significant difference from Mann-Whitney-Wilcoxon test (ns: no significance, \*:  $0.01 < p \leq 0.05$ , \*\*:  $0.001 < p \leq 0.01$ , \*\*\*:  $0.0001 < p \leq 0.001$ , \*\*\*\*:  $p \leq 0.0001$ ). Sub-figure G shows t-SNE plot of estimated immune abundance, color coded by cluster.

Kaplan-Meier curves indicate that cluster 2 has the worst survival probability throughout time out of all clusters (Figure 2G). In addition, cluster 2, along with cluster 1, has higher percentage of dead patients at the last time of follow up than cluster 3 (Figure 2C). Interestingly, cluster 2 also has the lowest immune scores compared to other clusters (Figure 1F).



**Figure 2.** Clinical characteristics of clusters and relationship between clinical features of osteosarcoma. Sub-figure A-E show percentage of patients by gender (A), metastasis at diagnosis (B), vital status at the last time of follow-up (C), Huvos grade (D), primary tumor location (E), in the 3 clusters. Sub-figure F shows a boxplot of patients' age at diagnosis in each cluster. Sub-figure G displays Kaplan-Meier curves of overall survival across 3 clusters. Sub-figures H-K shows the association between clinical features, H: percentage of alive and dead patients by metastasis at diagnosis, I: percentage of alive and dead patients by Huvos grade, J: percentage of low and high Huvos grade by tumor location, K: boxplot of age at diagnosis by gender, with asterisks indicating significant difference from Mann-Whitney-Wilcoxon test (ns: no significance, \*:  $0.01 < p \leq 0.05$ , \*\*:  $0.001 < p \leq 0.01$ , \*\*\*:  $0.0001 < p \leq 0.001$ , \*\*\*\*:  $p \leq 0.0001$ ).

Cluster 3 appears to have the best outcome among clusters. It has the lowest percentage of dead patients at the last time of follow up among all clusters (Figure 2C), and better survival rate than cluster 2 over time (Figure 2G) with p-value 0.07 from the log-rank test.

Cluster 1, which has the highest amount of CD8 T cells,  $\gamma\delta$  T cells, M1 Macrophages and Mast cells, has slightly better overall survival time than cluster 2 ( $p = 0.16$ , Figure 2G). However, cluster

1 seems to have worse outcome than cluster 3 due to its higher percentage of dead patients at the last time of follow up (Figure 2C). It is worth noting that there is no significant difference in the survival rate between cluster 1 and 3 according to the log-rank test ( $p = 0.5$ , Figure 2G).

### 3.5. *There is a relationship between certain clinical features of osteosarcoma*

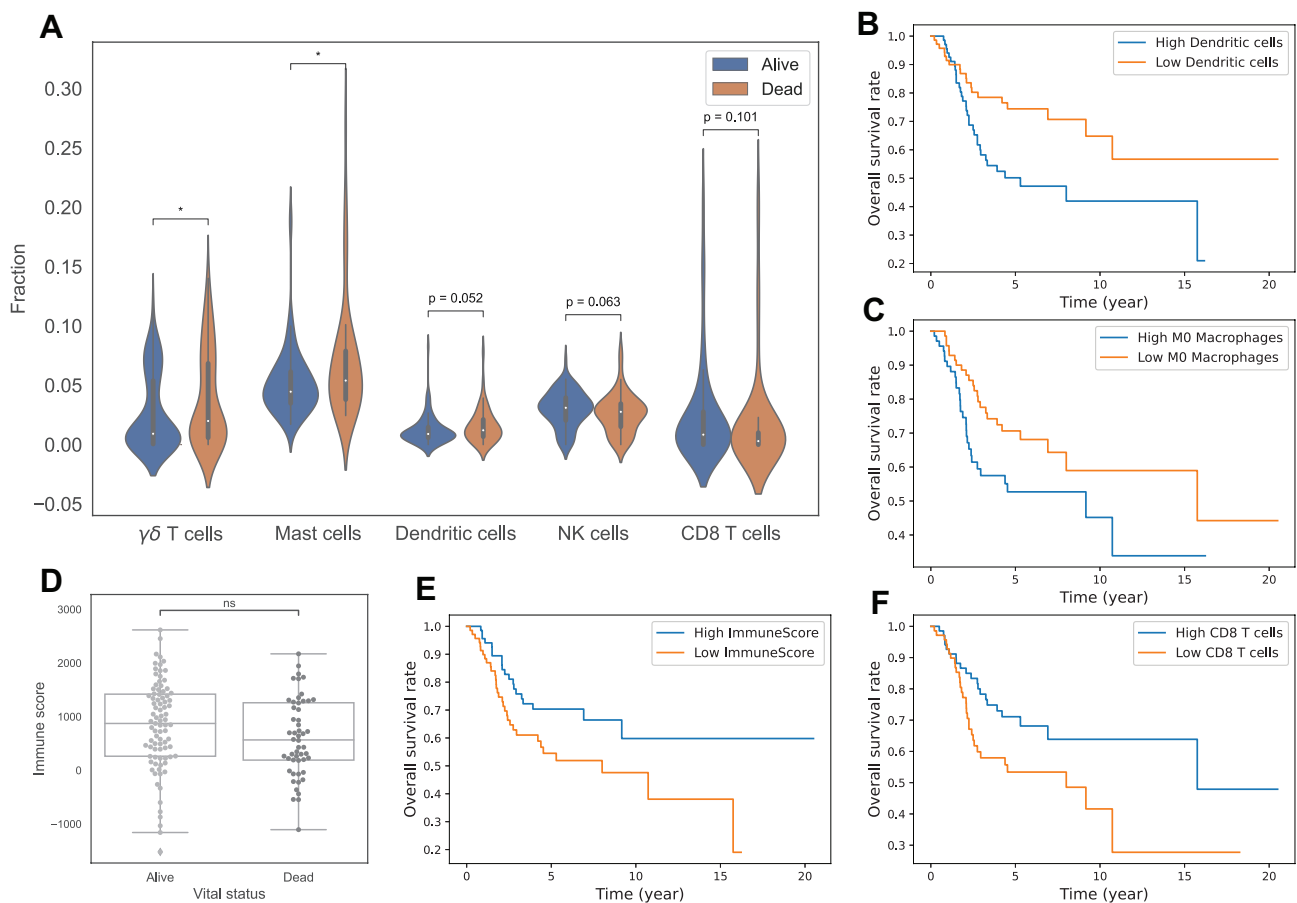
The Chi-square test and Mann-Whitney-Wilcoxon test show a relationship between metastasis at diagnosis and vital status ( $p = 0.001$ , Figure 2H), Huvos grade, which is a grading system to evaluate a patient's response to chemotherapy based on the percentage of necrosis in the tumor after treatment, and vital status ( $p = 0.036$ , Figure 2I), and between gender and age ( $p = 2.3e-4$ , Figure 2K) where male patients are older on average. We observe that patients with metastasis at diagnosis have much higher percentage of being dead at the last time of follow up than patients without metastasis (Figure 2H). This makes perfect sense since metastases have been known to associate with late stages of tumor and poor prognosis in many cancers. The other clinical feature with a relation to vital status is Huvos grade. Higher percentage of patients with high Huvos grade (3-4) are alive at the last time of follow up than patients with low Huvos grade (1-2) (Figure 2I), which is reasonable since a high Huvos grade means good response to chemotherapy. Figure 2J suggests that primary osteosarcoma tumors in the arm respond more poorly to chemotherapy than leg tumors, as illustrated by the high proportion of Huvos grade 1-2 in arm tumors. However, it is important to note that primary osteosarcoma tumors happen more often in the leg than in the arm (Figure 2E), and the observation in Figure 2J is based on a small number of arm tumor samples ( $n = 8$ ), thus the relationship between tumor location and Huvos grade is not considered significant by the Chi-square test ( $p = 0.18$ ).

### 3.6. *Immune score does not relate to vital status directly, but does relate to survival probability over time*

Figure 3D indicates no clear difference in immune score between alive and dead patients at the last time of follow up. However, Kaplan-Meier curves of high and low immune score with a median cut-off reveal that the high immune score group has a better outcome (Figure 3E). The log rank test supports this observation with a p-value of 0.03. Thus, higher immune score is associated with better survival probability throughout time. This is consistent with the observation of outcome in the clusters. Cluster 2 has significantly lower immune score than cluster 1 and 3 (Figure 1F), with p-values of  $5e-9$  and  $2.1e-5$  from Mann-Whitney-Wilcoxon test, and accordingly worse overall survival probability over time than cluster 1 and 3 (Figure 2G). Cluster 1, with the highest average immune score among clusters (Figure 1F), even though has about the same proportion of dead patients at the last time of follow up as cluster 2 (Figure 2C), shows better survival time than cluster 2 (Figure 2G).

### 3.7. *Relationship between immune infiltrates and survival outcome in osteosarcoma*

The Mann-Whitney-Wilcoxon test shows that there is a significant difference in the level of  $\gamma\delta$  T cells and Mast cells between alive and dead patients at the last time of follow up, with p-values of 0.045 and 0.022, where dead patients are associated with higher percentages of  $\gamma\delta$  T cells and Mast cells than alive patients (Figure 3A). Dendritic cells, NK cells and CD8 T cells also seem to associate with survival status in osteosarcoma. We observe higher level of NK cells ( $p = 0.063$ ), CD8 T cells ( $p = 0.1$ ) and lower level of Dendritic cells ( $p = 0.052$ ) in alive patients than in dead patients (Figure 3A).



**Figure 3.** Relationship of immune infiltrations with survival in osteosarcoma. Sub-figure A displays violin plots of fractions of  $\gamma\delta$  T cells, Mast cells, Dendritic cells, NK cells, CD8 T cells between alive and dead patients. Sub-figures B, C, E, F show Kaplan-Meier curves of overall survival between 2 groups, B: high vs low Dendritic cells, C: high vs low M0 Macrophages, E: high vs low ESTIMATE immune score, F: high vs low CD8 T cells. Sub-figure D shows a boxplot of ESTIMATE immune scores between alive and dead patients. Note: asterisks indicate significant difference from Mann-Whitney-Wilcoxon test (ns: no significance, \*:  $0.01 < p \leq 0.05$ , \*\*:  $0.001 < p \leq 0.01$ , \*\*\*:  $0.0001 < p \leq 0.001$ , \*\*\*\*:  $p \leq 0.0001$ ).

Kaplan-Meier curves (Figure 3B, C, F) and the log rank test indicate an association between survival outcomes and levels of Dendritic cells, M0 Macrophages and CD8 T cells, with a p-value of 0.01, 0.04 and 0.04, respectively. Low Dendritic cells, low M0 Macrophages and high CD8 T cells are associated with better survival probability over time in osteosarcoma patients. This is again consistent with the outcome of the clusters where cluster 2, with the highest level of M0 Macrophages and lowest CD8 T cells, has the worst overall survival.

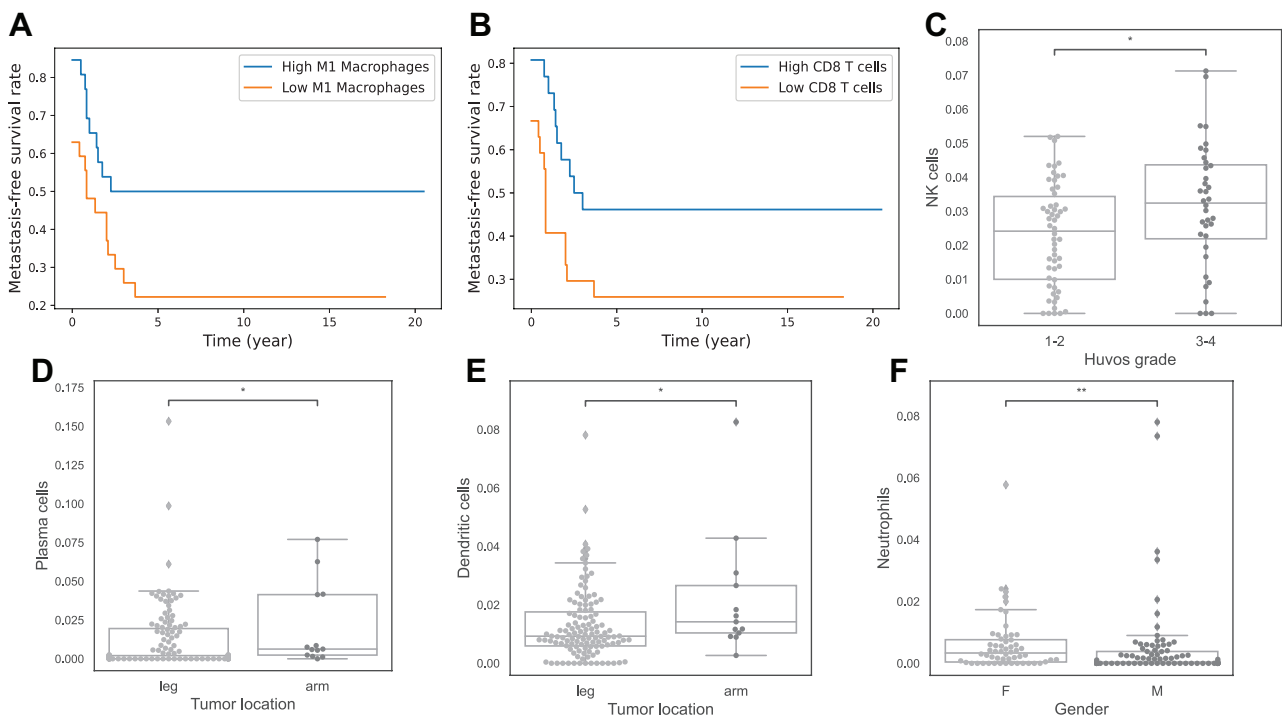
Overall, we found that  $\gamma\delta$  T cells, Mast cells, Dendritic cells, M0 Macrophages, NK cells and CD8 T cells have a relationship with the survival of osteosarcoma patients.



### 3.8. Association of immune infiltrates with other clinical features

We see no significant relationship between age or metastasis at diagnosis and the frequencies of immune cells. However, we notice an association of M1 Macrophages and CD8 T cells' frequencies to metastasis-free survival. High levels of M1 Macrophages and CD8 T cells are associated with better metastasis-free survival probability across time in osteosarcoma (Figure 4A, B), with p-values of 0.05 and 0.08 from the log-rank test, respectively. This means that patients with more M1 Macrophages and CD8 T cells are less likely to develop metastasis or die at any given time than patients with low percentage of these cells.

A relationship between some immune infiltrates and other clinical features of osteosarcoma has also been observed. Higher level of NK cells is associated with good response to chemotherapy ( $p = 0.035$ , Figure 4C). Patients with arm tumors have higher percentage of Plasma cells and Dendritic cells than patients with leg tumors (Figure 4D, E), with p-values of 0.046 and 0.033. Lastly, female patients are shown to have higher frequency of Neutrophils (Figure 4F), with p-value 0.0016.

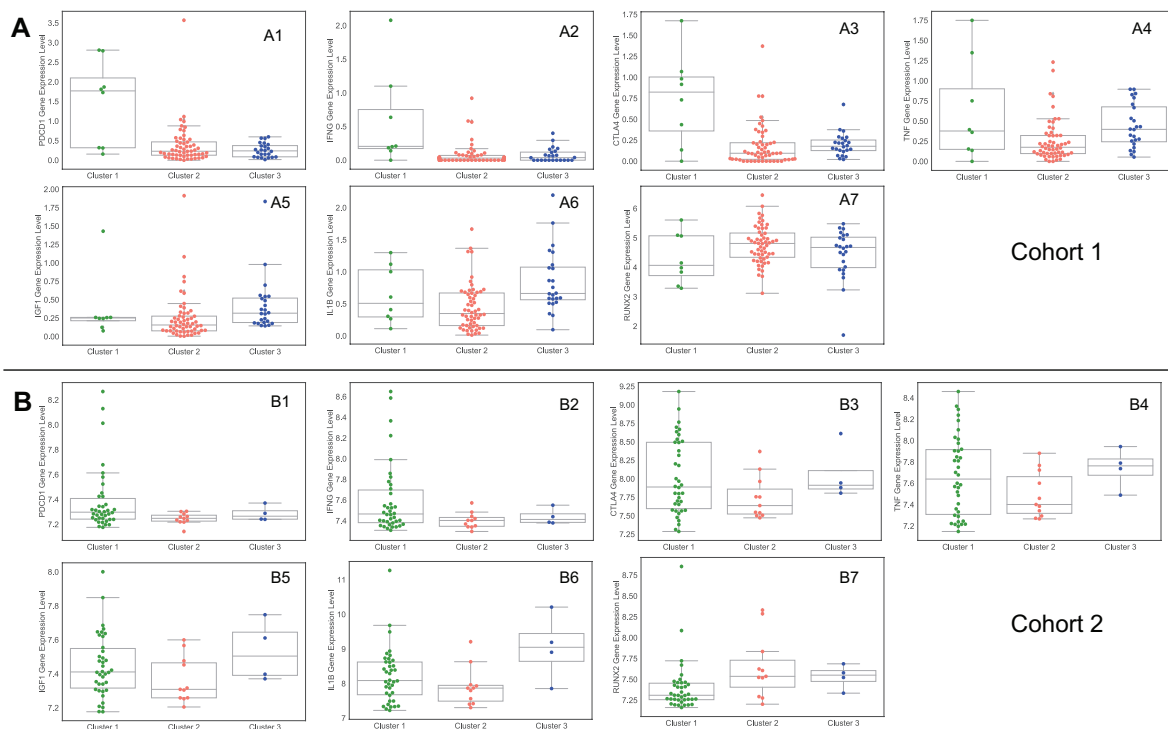


**Figure 4.** Association of immune infiltrations and other clinical features. Sub-figures A and B show Kaplan-Meier curves of metastasis-free-survival in cohort 1 between 2 groups, A: high vs low M1 Macrophages, B: high vs low CD8 T cells. Sub-figures C-F are boxplots to indicate relationship of immune infiltrates with Huvos grade (C), primary tumor location (D and E), and gender (F). Note: asterisks indicate significant difference from Mann-Whitney-Wilcoxon test (ns: no significance, \*:  $0.01 < p \leq 0.05$ , \*\*:  $0.001 < p \leq 0.01$ , \*\*\*:  $0.0001 < p \leq 0.001$ , \*\*\*\*:  $p \leq 0.0001$ ).

### 3.9. Expression level of genes encoding PD-1, $INF-\gamma$ , CTLA4, TNF, $IL1-\beta$ , IGF1, $IL-6$ and RUNX2 are significantly different for some clusters

We use the gene expression values for some important proteins, and we analyze gene expression value of the proteins separately for cohort 1 and 2 because they have different data types: RNA-Seq and microarray, respectively.

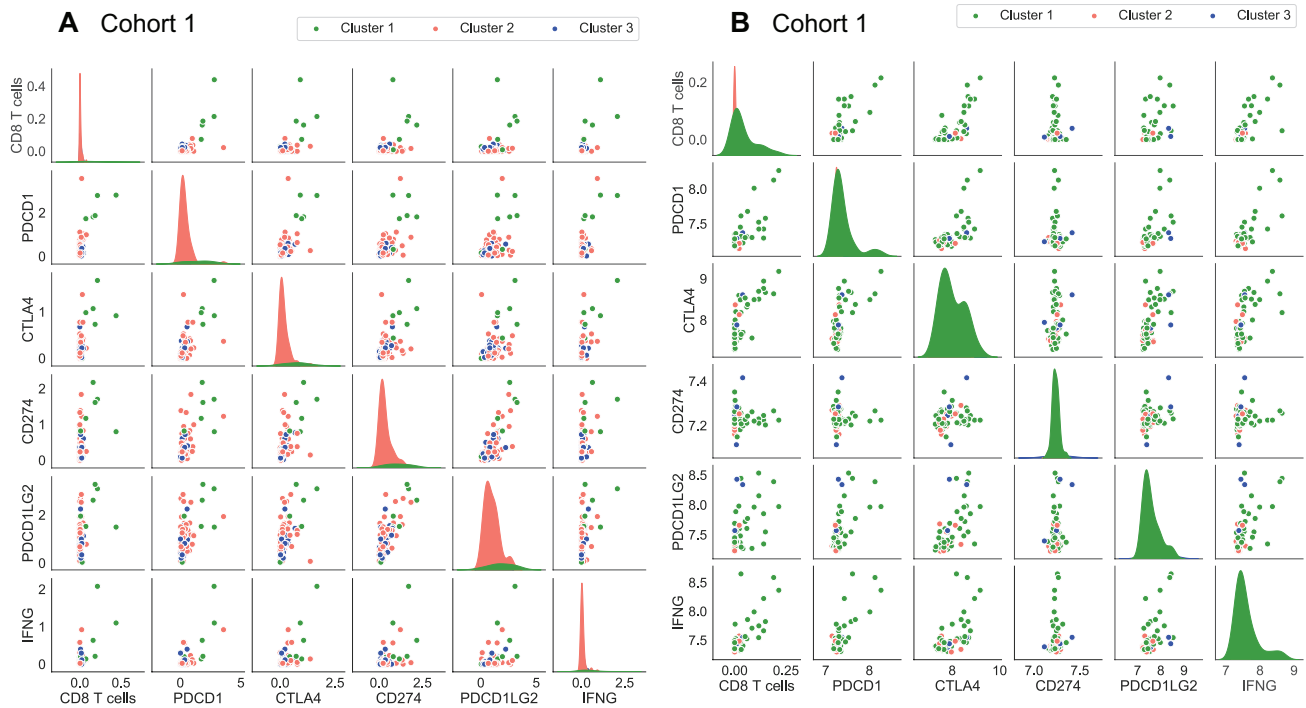
Programmed cell death protein 1 (PD-1) is a type of protein on T cells and cancer cells use it to bind with PD-1 ligand (PD-L1) and PD-2 ligand (PD-L2) to escape cell death by immune cells. There is a high correlation between PDCD1 gene, which encodes PD-1 protein, and CD8 T cells in both data sets with correlation coefficient of 0.70 and 0.77, respectively, and p-values less than 0.05 (Figure 6). As a result of this correlation, PDCD1 is the highest in cluster 1 (Figure 5A1, B1). Cluster 1 also has the highest expression of CTLA4 gene (Figure 5A3, B3) that encodes Cytotoxic T-Lymphocyte Associated Protein 4 (CTLA4), which is a member of immunoglobulin superfamily and has been found to significantly associate with the risk of osteosarcoma [38,39]. Moreover, we see that gene expression value of CTLA4 is significantly correlated with CD8 T cells in osteosarcoma tumors (Figure 6).



**Figure 5.** Gene expression values of important proteins in the clusters. Sub-figures (A1–A7) and (B1–B7) show the gene expression values that come from cohort 1 and cohort 2 data sets, respectively.

Interferon  $\gamma$  ( $INF-\gamma$ ), encoded by IFNG gene, has antiviral, immunoregulatory, and anti-tumor properties in the immune system and is secreted by mostly T cells and NK cells [40]. Importantly, CD8 T cells frequency and PDCD1 gene expression are significantly correlated with IFNG gene

expression so that cluster 1 has the highest level of IFNG compared the other clusters (Figure 5A2, B2). Beside these, we do not see any significant correlation between expression levels of CD274 and PDCD1LG2 genes, that encodes PD-L1 and PD-L2 respectively, with the expression levels of PDCD1 and IFNG, and the percentage of CD8 T cells in osteosarcoma tumors (Figure 6).



**Figure 6.** Correlation and distribution of important proteins in the clusters. Sub-figures A and B represent the cohort 1 and cohort 2 data sets respectively.

Tumor necrosis factor (TNF) is a cytokine that is mainly produced by Macrophages and has crucial roles in tumor development and tumor progression inducing apoptosis, necrosis, angiogenesis, immune cell activation, differentiation, and cell migration [41]. We notice that cluster 2 has the lowest TNF gene expression among clusters, while cluster 1 and 3 have roughly similar average expression of this gene (Figure 5A4, B4). In addition, we analyze gene *IL1B* that encodes cytokin protein Interleukin-1 beta ( $IL-1\beta$ ), which is produced by activated Macrophages [42], and see that *IL1B* gene expression is the lowest in cluster 2 and the highest in cluster 3 (Figure 5A6, B6).

Beside these, Insulin-like growth factor 1 (IGF-1) is a hormone that has important function in the development and function of many tissues and it has been used as a diagnostic marker for osteosarcoma [43, 44]. Similar to TNF and *IL1B* genes expression, cluster 2 has the lowest amount of IGF1 among other clusters (Figure 5A5, B5). Furthermore, we examine *RUNX2* oncogene that is associated with amplifications and it has been found to correlate to poor response to chemotherapy in osteosarcoma [45, 46]. In our analysis, cluster 1 has the lowest amount of *RUNX2* gene and cluster 2 and 3 show almost similar expression of *RUNX2* gene (Figure 5A7, B7).

#### 4. Discussion

The findings from analyses using estimated immune infiltrations in osteosarcoma have varied among studies, perhaps due to the small number of osteosarcoma tumors with available gene expression data in the literature. In this study, we find that infiltration of CD8 T cells, NK cells and M1 Macrophages have a positive association with prognosis, while infiltration of  $\gamma\delta$  T cells, Mast cells, M0 Macrophages and Dendritic cells have a negative association with prognosis. Yu et al. [24] also illustrates that high level of CD8 T cells is a good prognosis in their survival analysis, and results from [26] and [28] indirectly suggest the positive prognostic value of CD8 T cells. Tang et al. [26] reports that CD8 T cells infiltration has a positive correlation with CXCR3 expression which is related to good prognosis. Khader et al. [28] shows that low-risk patients have high level of CD8 T cells and NK cells, which supports with our conclusion on these cells. In agreement with our findings on prognostic value of M1 Macrophages, Song et al [31] demonstrates that high level of M1 Macrophages is associated with good prognosis, while Zhang et al. [22] and Tang et al. [26] imply the same from their results. Our conclusion about M0 Macrophages aligns with the results from [22] and [26], but contradicts with the finding from [34] that abundance level of M0 Macrophages is positively correlated with survival.

The observable difference in outcomes between clusters are likely due to the relationship between immune infiltrates and prognosis in osteosarcoma, because we cluster the patients based on their immune composition. Our results indicate that cluster 2 has the worst outcome, while cluster 3 seems to have the best outcome among clusters. The main difference in immune composition between these two clusters is that cluster 2 has much higher percentage of M0 Macrophages, and lower percentage of CD8 T cells and M1 Macrophages than cluster 3. In general, we found that high levels of CD8 T cells and M1 Macrophages are associated with good prognosis, while a high level of M0 Macrophages correlates with poor prognosis in osteosarcoma. These results make sense because CD8 T cells are known to kill cancer cells directly [47,48] and M1 Macrophages exhibit anti-tumor effects by producing cytokines that inhibit osteosarcoma growth [49]. These facts could also explain the observed differences between the outcomes of patients in clusters 2 and 3. Meanwhile, cluster 1 has worse outcome than cluster 3, but better outcome than cluster 2. This could be due to the fact that cluster 1 has both high level of immune cells associated with good prognosis such as CD8 T cells and M1 Macrophages, and high level of immune cells associated with poor prognosis according to our results such as  $\gamma\delta$  T cells and Mast cells. A high infiltration of mast cells has been associated with poor prognosis, low survival and increased metastasis in many cancers [50], while  $\gamma\delta$  T cells show dual effects on cancer growth [51]. Both mast cells and  $\gamma\delta$  T cells promote tumor development by supporting angiogenesis through angiogenic factors production [50, 51]. Mast cells also produce proteases, which lead to extracellular matrix degradation and tissue remodeling, and thus promote tumor growth [50].  $\gamma\delta$  T cells have been reported to secrete TGF- $\beta$  [51], which is a pro-tumor cytokine in osteosarcoma [52–54]. Overall, the clinical outcomes of the clusters are consistent with our findings and biological knowledge on prognostic values of immune cells in osteosarcoma.

On the other hand, we did not observe any difference in age or metastasis at diagnosis between clusters. This can be explained by the lack of correlation between immune infiltrates and these clinical variables, which suggests that the immune composition of the primary tumor has no effect on age or metastasis status at diagnosis.

Immune checkpoints have an important role in the immune system to prevent autoimmune diseases, but unfortunately they can allow immune tolerance against tumors. PD-1 and CTLA-4 are the main checkpoints that tumor cells use to block immune system [55–57]. Blocking PD-1 pathway has improved oncological survival of several patients with metastatic cancers, including melanoma, renal cell carcinoma, and colon cancer [58, 59]. Also, targeting CTLA-4 in patients with metastatic melanomas demonstrates significant development about overall survival [60].

It has been reported that osteosarcoma patients treated with an anti PD-1 drug, Pembrolizumab, show some improvement in disease progression [61]. Combination of PD-1 and CTLA-4 blockade therapy in bone sarcoma have shown better response compared to single checkpoint inhibitor therapy [62]. Note, patients in cluster 1 have the highest expression levels of IFNG, PDCD1 and CTLA-4 that are significantly correlated with CD8 T cells (Figures 5 and 6), and it has been shown that INF- $\gamma$  increases the CD8 T cells expansion [63]. Thus, patients in cluster 1 might respond well to combination of PD-1 and CTLA-4 blockade therapies.

It has been suggested in several studies that bacteria are able to activate anti-tumor immune responses [64, 65]. In a study with combination of Bacillus Calmette-Guerin (BCG) injection and tumor vaccine, 18% of the patients remained alive and disease-free and it has been reported that bacterial vaccine caused increased level of immunoregulatory cytokines such as TNF- $\alpha$ , IFN- $\gamma$ , and IL1- $\beta$  that might be involved in inducing tumor regression [66]. As a result, bacterial vaccine and inactivated tumor cells injection can be thought of as a treatment that activates anti-tumor immune responses [47]. In our results, cluster 2 has the lowest amount of gene expression level of immunoregulatory cytokines TNF- $\alpha$ , IFN- $\gamma$ , and IL1- $\beta$  (Figure 5) so that tumors in this cluster might be treated with bacterial vaccine. Otherwise, targeting RUNX2 oncogene with chemotherapy is suggested as a new therapeutic approach to osteosarcoma patients in recent studies [46, 67] and cluster 2 has the highest amount of RUNX2 gene expression values compared to other clusters (Figure 5A7–B7) so with the help of further studies, tumors similar to those in cluster 2 also might be good candidates to treat with targeting RUNX2 in conjunction to standard chemotherapy.

Targeting tumor associated macrophages (TAM) is another alternative treatment method for osteosarcoma tumors and treatments that suppress M2 Macrophages phenotype or block the polarization of M1 Macrophages to M2 Macrophages have shown positive results in several studies [68–71]. Thus, tumors in cluster 3, which has the highest amount of M2 macrophages (Figure 1D) can be treated with targeting TAMs.

## Acknowledgements

We would like to thank Dr. Marieke Kuijjer for the helpful clarifications on GSE21257 data set.

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under Award Number R21CA242933. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Conflict of interest

No potential conflict of interest was reported by the authors.

## References

1. M. Kansara, M. W. Teng, M. J. Smyth, D. M. Thomas, Translational biology of osteosarcoma, *Nat. Rev. Cancer*, **14** (2014), 722–735.
2. American Cancer Society, What Cause Osteosarcoma?, 2021. Available from: <https://www.cancer.org/cancer/osteosarcoma/causes-risks-prevention/what-causes.html>.
3. X. He, Z. Gao, H. Xu, Z. Zhang, P. Fu, A meta-analysis of randomized control trials of surgical methods with osteosarcoma outcomes, *J. Orthop. Surg. Res.*, **12** (2017), 5.
4. P. A. Meyers, C. L. Schwartz, M. D. Krailo, J. H. Healey, M. L. Bernstein, D. Betcher, et al., Osteosarcoma: the addition of muramyl tripeptide to chemotherapy improves overall survival—a report from the Children’s Oncology Group, *J. Clin. Oncol.*, **26** (2008), 633–638.
5. F. Conforti, L. Pala, V. Bagnardi, T. De Pas, M. Martinetti, G. Viale, et al., Cancer immunotherapy efficacy and patients’ sex: a systematic review and meta-analysis, *Lancet Oncol.*, **19** (2018), 737–746.
6. Y. T. Lee, Y. J. Tan, C. E. Oon, Molecular targeted therapy: treating cancer with specificity, *Eur. J. Pharmacol.*, **834** (2018), 188–196.
7. K. L. Davis, E. Fox, M. S. Merchant, J. M. Reid, R. A. Kudgus, X. Liu, et al., Nivolumab in children and young adults with relapsed or refractory solid tumours or lymphoma (ADVL1412): A multicentre, open-label, single-arm, phase 1–2 trial, *Lancet Oncol.*, **21** (2020), 541–550.
8. S. I. Grivnenikov, F. R. Greten, M. Karin, Immunity, inflammation, and cancer, *Cell*, **140** (2010), 883–899.
9. T. Kitamura, B. Z. Qian, J. W. Pollard, Immune cell promotion of metastasis, *Nat. Rev. Immunol.*, **15** (2015), 73–86.
10. J. B. Swann, M. J. Smyth, Immune surveillance of tumors, *J. Clin. Invest.*, **117** (2007), 1137–1146.
11. F. Pagès, A. Kirilovsky, B. Mlecnik, M. Asslaber, M. Tosolini, G. Bindea, et al., In Situ Cytotoxic and Memory T Cells Predict Outcome in Patients With Early-Stage Colorectal Cancer, *J. Clin. Oncol.*, **27** (2009), 5944–5951.
12. J. Yao, W. Xi, Y. Zhu, H. Wang, X. Hu, J. Guo, Checkpoint molecule PD-1-assisted CD8+ T lymphocyte count in tumor microenvironment predicts overall survival of patients with metastatic renal cell carcinoma treated with tyrosine kinase inhibitors, *Cancer Manag. Res.*, **10** (2018), 3419–3431.
13. N. Tarek, D. A. Lee, Natural killer cells for osteosarcoma, *Adv. Exp. Med. Biol.*, **804** (2014), 341–353.
14. Z. Li, Potential of human  $\gamma\delta$  T cells for immunotherapy of osteosarcoma, *Mol. Biol. Rep.*, **40** (2013), 427–437.
15. J. R. Heath, A. Ribas, P. S. Mischel, Single-cell analysis tools for drug discovery and development, *Nat. Rev. Drug. Discov.*, **15** (2016), 204–216.
16. T. Le, R. A. Aronow, A. Kirshtein, L. Shahriyari, A review of digital cytometry methods: estimating the relative abundance of cell types in a bulk of cells, *Brief. Bioinform.*, **2020** (2020), bbaa219.

17. S. Su, S. Akbarinejad, L. Shahriyari, Immune classification of clear cell renal cell carcinoma, 2020. Available from: <https://www.biorxiv.org/content/10.1101/2020.07.03.187047v1.abstract>.
18. A. Kirshtein, S. Akbarinejad, W. Hao, T. Le, S. Su, R. A. Aronow, et al., Data driven mathematical model of colon cancer progression, *J. Clin. Med.*, **9** (2020), 3947.
19. L. Li, L. Shen, J. Ma, Q. Zhou, M. Li, H. Wu, et al., Evaluating distribution and prognostic value of new tumor-infiltrating lymphocytes in HCC based on a scRNA-seq study with CIBERSORTx, *Front. Med.*, **7** (2020), 451.
20. L. Huang, H. Chen, Y. Xu, J. Chen, Z. Liu, Q. Xu, Correlation of tumor-infiltrating immune cells of melanoma with overall survival by immunogenomic analysis, *Cancer Med.*, **9** (2020), 8444–8456.
21. A. M. Newman, C. B. Steen, C. L. Liu, A. J. Gentles, A. A. Chaudhuri, F. Scherer, et al., Determining cell type abundance and expression from bulk tissues with digital cytometry, *Nat. Biotechnol.*, **37** (2019), 773–782.
22. C. Zhang, J. H. Zheng, Z. H. Lin, H. Y. Lv, Z. M. Ye, Y. P. Chen, et al., Profiles of immune cell infiltration and immune-related genes in the tumor microenvironment of osteosarcoma, *Aging*, **12** (2020), 3486–3501.
23. W. Hong, H. Yuan, Y. Gu, M. Liu, Y. Ji, Z. Huang, et al., Immune-related prognosis biomarkers associated with osteosarcoma microenvironment, *Cancer Cell Int.*, **20** (2020), 1–12.
24. Y. Yu, H. Zhang, T. Ren, Y. Huang, X. Liang, W. Wang, et al., Development of a prognostic gene signature based on an immunogenomic infiltration analysis of osteosarcoma, *J. Cell. Mol. Med.*, **24** (2020), 11230–11242.
25. C. Hu, C. Liu, S. Tian, Y. Wang, R. Shen, H. Rao, et al., Comprehensive analysis of prognostic tumor microenvironment-related genes in osteosarcoma patients, *BMC Cancer*, **20** (2020), 1–11.
26. Y. Tang, Z. Gu, Y. Fu, J. Wang, CXCR3 from chemokine receptor family correlates with immune infiltration and predicts poor survival in osteosarcoma, *Biosci. Rep.*, **39** (2019), 1–12.
27. J. Niu, T. Yan, W. Guo, W. Wang, Z. Zhao, T. Ren, et al., Identification of Potential Therapeutic Targets and Immune Cell Infiltration Characteristics in Osteosarcoma Using Bioinformatics Strategy, *Front. Oncol.*, **10** (2020), 1628.
28. L. Q. Li, L. H. Zhang, Y. Zhang, X. C. Lu, Y. Zhang, Y. K. Liu, et al., Construction of immune-related gene pairs signature to predict the overall survival of osteosarcoma patients, *Aging*, **12** (2020), 22906–22926.
29. T. Zhang, Y. Nie, H. Xia, Y. Zhang, K. Cai, X. Chen, et al., Identification of Immune-Related Prognostic Genes and LncRNAs Biomarkers Associated With Osteosarcoma Microenvironment, *Front. Oncol.*, **10** (2020), 1109.
30. W. Yuan, Y. Deng, E. Ren, G. Zhang, Z. Wu, Q. Xie, Analysis of Immune Infiltration Pattern in Osteosarcoma and Its Clinical Significance, *Res. Sq.*, **2020** (2020), 1–26.
31. Y. J. Song, Y. Xu, X. Zhu, J. Fu, C. Deng, H. Chen, et al., Immune Landscape of the Tumor Microenvironment Identifies Prognostic Gene Signature CD4/CD68/CSF1R in Osteosarcoma, *Front. Oncol.*, **10** (2020), 1198.
32. T. Chen, L. Zhao, Patrolling monocytes inhibit osteosarcoma metastasis to the lung, *Aging*, **12** (2020), 23004–23016.

33. C. Deng, Y. Xu, J. Fu, X. Zhu, H. Chen, H. Xu, et al., Reprograming the tumor immunologic microenvironment using neoadjuvant chemotherapy in osteosarcoma, *Cancer Sci.*, **111** (2020), 1899–1909.
34. X. Yang, W. Zhang, P. Xu, NK cell and macrophages confer prognosis and reflect immune status in osteosarcoma, *J. Cell. Biochem.*, **120** (2019), 8792–8797.
35. C. C. Wu, H. C. Beird, J. A. Livingston, S. Advani, A. Mitra, S. Cao, et al., Immuno-genomic landscape of osteosarcoma, *Nat. Commun.*, **11** (2020), 1–11.
36. A. M. Newman, C. L. Liu, M. R. Green, A. J. Gentles, W. Feng, Y. Xu, et al., Robust enumeration of cell subsets from tissue expression profiles, *Nat. Methods*, **12** (2015), 453–457.
37. K. Yoshihara, M. Shahmoradgoli, E. Martínez, R. Vegesna, H. Kim, W. Torres-Garcia, et al., Inferring tumour purity and stromal and immune cell admixture from expression data, *Nat. Commun.*, **4** (2013), 1–11.
38. G. Qiao, H. Miao, Y. Yi, D. Wang, B. Liu, Y. Zhang, et al., Genetic association between CTLA-4 variations and osteosarcoma risk: Case-control study, *Int. J. Clin. Exp. Med.*, **9** (2016), 9598–9602.
39. C. Zhang, W. H. Hou, X. X. Ding, X. Wang, H. Zhao, X. W. Han, et al., Association of cytotoxic T-lymphocyte antigen-4 polymorphisms with malignant bone tumors risk: A meta-analysis, *Asian Pac. J. Cancer Prev.*, **17** (2016), 3783–3789.
40. K. Schroder, P. J. Hertzog, T. Ravasi, D. A. Hume, Interferon- $\gamma$ : an overview of signals, mechanisms and functions, *J. Leukocyte Biol.*, **75** (2004), 163–189.
41. H. Wajant, The role of TNF in cancer, *Results Probl. Cell Differ.*, **49** (2009), 1–15.
42. National Center for Biotechnology Information, IL1B interleukin 1 beta, 2021. Available from: <https://www.ncbi.nlm.nih.gov/gene/3553>.
43. Y. S. Li, Q. Liu, H. B. He, W. Luo, The possible role of insulin-like growth factor-1 in osteosarcoma, *Curr. Prob. Cancer*, **43** (2019), 228–235.
44. T. Jentsch, B. Robl, M. Husmann, B. Bode-Lesniewska, B. Fuchs, Worse prognosis of osteosarcoma patients expressing IGF-1 on a tissue microarray, *Anticancer Res.*, **34** (2014), 3881–3890.
45. J. W. Martin, M. Zielenska, G. S. Stein, A. J. van Wijnen, J. A. Squire, The role of RUNX2 in osteosarcoma oncogenesis, *Sarcoma*, **2011** (2011), 1–13.
46. A. Roos, L. Satterfield, S. Zhao, D. Fuja, R. Shuck, M. J. Hicks, et al., Loss of Runx2 sensitises osteosarcoma to chemotherapy-induced apoptosis, *Br. J. Cancer*, **113** (2015), 1289–1297.
47. S. Miwa, T. Shirai, N. Yamamoto, K. Hayashi, A. Takeuchi, K. Igarashi, et al., Current and emerging targets in immunotherapy for osteosarcoma, *J. Oncol.*, **2019** (2019), 1–8.
48. K. Wang, A. T. Vella, Regulatory T cells and cancer: a two-sided story. *Immunol. Invest.*, **45** (2016), 797–812.
49. M. F. Heymann, D. Heymann, Immune environment and osteosarcoma, in *Osteosarcoma-Biology, Behavior and Mechanisms*, InTech: London, UK, (2017), 105–120.
50. T. T. Maciel, I. C. Moura, O. Hermine, The role of mast cells in cancers, *F1000Prime Rep.*, **7** (2015), 5–10.



51. Y. Zhao, C. Niu, J. Cui, Gamma-delta ( $\gamma \delta$ ) T Cells: friend or foe in cancer development?, *J. Transl. Med.*, **16** (2018), 1–13.
52. M. F. Heymann, F. Lézot, D. Heymann, The contribution of immune infiltrates and the local microenvironment in the pathogenesis of osteosarcoma, *Cell. Immunol.*, **343** (2019), 103711.
53. A. Lamora, J. Talbot, M. Mullard, B. L. Royer, F. Redini, F. Verrecchia, TGF- $\beta$  signaling in bone remodeling and osteosarcoma progression, *J. Clin. Med.*, **5** (2016), 96.
54. I. Corre, F. Verrecchia, V. Crenn, F. Redini, V. Trichet, The osteosarcoma microenvironment: a complex but targetable ecosystem, *Cells*, **9** (2020), 1–25.
55. D. S. Chen, I. Mellman, Oncology meets immunology: the cancer-immunity cycle. *Immunity*, **39** (2013), 1–10.
56. R. J. Motzer, B. Escudier, D. F. McDermott, S. George, H. J. Hammers, S. Srinivas, et al., Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma. *New Engl. J. Med.*, **373** (2015), 1803–1813.
57. J. Dine, R. Gordon, Y. Shames, M. Kasler, M. Barton-Burke, Immune checkpoint inhibitors: An innovation in immunotherapy for the treatment and management of patients with cancer, *Asia Pac. J. Oncol. Nurs.*, **4** (2017), 127–135.
58. S. L. Topalian, F. S. Hodi, J. R. Brahmer, S. N. Gettinger, D. C. Smith, D. F. McDermott, et al., Safety, activity, and immune correlates of anti-PD-1 antibody in cancer, *New Engl. J. Med.*, **366** (2012), 2443–2454.
59. S. Koyama, E. A. Akbay, Y. Y. Li, G. S. Herter-Sprue, K. A. Buczkowski, W. G. Richards, et al., Adaptive resistance to therapeutic PD-1 blockade is associated with upregulation of alternative immune checkpoints, *Nat. Commun.*, **7** (2016), 10501.
60. F. S. Hodi, S. J. O'Day, D. F. McDermott, R. W. Weber, J. A. Sosman, J. B. Haanen, et al., Improved survival with ipilimumab in patients with metastatic melanoma, *New Engl. J. Med.*, **363** (2010), 711–723.
61. H. A. Tawbi, M. Burgess, V. Bolejack, B. A. Van Tine, S. M. Schuetze, J. Hu, 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.*, **18** (2017), 1493–1501.
62. P. Thanindratarn, D. C. Dean, S. D. Nelson, F. J. Hornicek, Z. Duan, Advances in immune checkpoint inhibitors for bone sarcoma therapy, *J. Bone Oncol.*, **15** (2019), 100221.
63. Ö. Sercan, G. J. Hämmerling, B. Arnold, T. Schüler, Cutting Edge: Innate Immune Cells Contribute to the IFN- $\gamma$ -Dependent Regulation of Antigen-Specific CD8 + T Cell Homeostasis, *J. Immunol.*, **176** (2006), 735–739.
64. B. D. X. Lascelles, W. S. Dernell, M. T. Correa, M. Lafferty, C. M. Devitt, C. A. Kuntz CA, et al., Improved survival associated with postoperative wound infection in dogs treated with limb-salvage surgery for osteosarcoma, *Ann. Surg. Oncol.*, **12** (2005), 1073–1083.
65. Y. Chen, S. F. Xu, M. Xu, X. C. Yu, Postoperative infection and survival in osteosarcoma patients: Reconsideration of immunotherapy for osteosarcoma, *Mol. Clin. Oncol.*, **3** (2015), 495–500.
66. J. Karbach, A. Neumann, K. Brand, C. Wahle, E. Siegel, M. Maeurer, et al., Phase I clinical trial of mixed bacterial vaccine (Coley's toxins) in patients with NY-ESO-1 expressing cancers: Immunological effects and clinical activity, *Clin. Cancer Res.*, **18** (2012), 5449–5459.

67. Z. Ling, G. Fan, D. Yao, J. Zhao, Y. Zhou, J. Feng, et al., MicroRNA-150 functions as a tumor suppressor and sensitizes osteosarcoma to doxorubicin-induced apoptosis by targeting RUNX2, *Exp. Ther. Med.*, **19** (2019), 481–488.
68. Z. Wang, Z. Wang, B. Li, S. Wang, T. Chen, Z. Ye, Innate immune cells: A potential and promising cell population for treating osteosarcoma, *Front. Immunol.*, **10** (2019), 1–19.
69. Q. Zhou, M. Xian, S. Xiang, D. Xiang, X. Shao, J. Wang, et al., All-trans retinoic acid prevents osteosarcoma metastasis by inhibiting M2 polarization of tumor-associated macrophages, *Cancer Immunol. Res.*, **5** (2017), 547–559.
70. Y. Kimura, M. Sumiyoshi, Resveratrol prevents tumor growth and metastasis by inhibiting lymphangiogenesis and M2 macrophage activation and differentiation in tumor-associated macrophages, *Nutr. Cancer*, **68** (2016), 667–678.
71. Y. Kimura, M. Sumiyoshi, Antitumor and antimetastatic actions of dihydroxycoumarins (esculetin or fraxetin) through the inhibition of M2 macrophage differentiation in tumor-associated macrophages and/or G1 arrest in tumor cells, *Eur. J. Pharmacol.*, **746** (2015), 115–125.

## Appendix

### A.1. CIBERSORTx B-mode algorithm

CIBERSORTx is the latest version of CIBERSORT, which has been the most popular deconvolution method up until now. Both algorithms are based on the following model:

$$y = X\beta$$

where  $y$  is the vector of gene expression values of the bulk tumor, called mixture data,  $X$  is the signature matrix where each column is a vector of gene expression values of a cell type, and  $\beta$  is the vector of estimated cell fractions.

CIBERSORT solves for  $\beta$  using a machine learning technique called nu-Support Vector Machine ( $\nu$ -SVR). Specifically, CIBERSORT finds  $\beta$  by minimizing the objective function:

$$\frac{1}{2} \|\beta\|^2 + C \left( \nu\epsilon + \frac{1}{N} \sum_{i=1}^N (\xi_i + \xi_i^*) \right)$$

s.t.  $y_i - \hat{y}_i \leq \epsilon + \xi_i$  and  $\hat{y}_i - y_i \leq \epsilon + \xi_i^*$  and  $\xi_i, \xi_i^* \geq 0$ .

Here  $\epsilon$  is the margin of error,  $\nu$  is a model hyperparameter between 0 and 1,  $\hat{y}_i$  is the predicted value of  $y_i$ , and  $(\xi_i + \xi_i^*)$  is the absolute error of  $\hat{y}_i$  that lie outside margin of error  $\epsilon$  from  $y_i$ . By minimizing this objective function, CIBERSORT finds estimated fractions  $\hat{\beta}$  that result in the smallest absolute error between observed and predicted mixture data ( $y$  and  $\hat{y}$ ) where a margin of error  $\epsilon$  is tolerated. After this optimization step, all negative coefficients in the estimated  $\hat{\beta}$  are set to 0 and the coefficients are normalized to sum to 1 because  $\hat{\beta}$  represents cell fractions.

CIBERSORTx extends CIBERSORT by adding batch correction to adjust the potential cross-platform variations between mixture data  $y$  and signature matrix  $X$ . The algorithm of CIBERSORTx B-mode works as follows:

1. Given signature matrix  $X$  and mixture data  $y$ , obtain estimated cell fractions  $\hat{\beta}$  from CIBERSORT

2. Compute estimated mixture data  $\hat{y}$  as  $\hat{y} = X\hat{\beta}$
3. Use Combat, a well-known batch correction method, to adjust the technical variations between  $y$  and  $\hat{y}$ , outputting adjusted mixture data  $y_{adj}$
4. Obtain final estimated cell fractions from  $X$  and  $y_{adj}$  using CIBERSORT

### A.2. ESTIMATE algorithm for Immune Score

Given an immune gene set, which is a set of genes associated with the quantity of immune infiltrations, ESTIMATE uses single sample gene set enrichment analysis (ssGSEA) to compute the Immune Score from gene expression data of bulk tumor. A high Immune Score represents a high level of infiltrating immune cells in the tumor tissue. ssGSEA algorithm in ESTIMATE works as follows:

1. Order gene expression data of bulk tumor from highest to lowest and replace gene expression values with their ranks
2. For each gene  $i$ , compute the following:

$$P_G^w(G, y, i) = \frac{\sum_{r_j \in G, j \leq i} |r_j|^\alpha}{\sum_{r_j \in G} |r_j|^\alpha},$$

$$P_{NG}(G, y, i) = \sum_{r_j \notin G, j \leq i} \frac{1}{(N - N_G)},$$

where  $G$  is the given immune gene set which has  $N_G$  genes,  $y$  is the gene expression data of bulk tumor which has  $N$  genes,  $r_j$  is the rank of gene  $j$  and  $\alpha$  is a hyperparameter between 0 and 1.

3. The Immune Score of sample  $y$  given immune gene set  $G$  is:

$$IS(G, y) = \sum_{i=1}^N [P_G^w(G, y, i) - P_{NG}(G, y, i)].$$

### A.3. K-means clustering algorithm

Given a number of clusters  $k$ , K-means clustering works as follows:

1. Randomly choose  $k$  data points as centroids
2. Assign each data point to its closest centroid, resulting in  $k$  clusters based on  $k$  centroids
3. Compute a new centroid for each cluster by taking average of data points that belong to that cluster
4. Repeat step 2 and 3 until the centroids no longer change



AIMS Press

© 2021 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)