

Exploring the relationship between immune heterogeneity characteristic genes of rheumatoid arthritis and acute myeloid leukemia

Chengzhi Jiang¹ · Wenjuan Jiang² · Pengtao Liu³ · Wenxue Sun¹ · Wenjie Teng³

Received: 15 September 2023 / Accepted: 11 December 2023

Published online: 02 January 2024

© The Author(s) 2023 **OPEN**

Abstract

Background People with autoimmune diseases are prone to cancer, and there is a close relationship between rheumatoid arthritis (RA) and acute myeloid leukemia (AML). The bone marrow (BM) is affected throughout the course of RA, with a variety of hematologic involvement. Hopes are pinned on rheumatoid arthritis research to obtain BM biomarkers for AML.

Methods Synovial transcriptome sequencing data for RA and osteoarthritis (OA), and single-cell sequencing data for RA and controls were obtained from the GEO database. Bone marrow sequencing data for AML patients and normal subjects were obtained from the UCSC Xena database. The final immune heterogeneity characteristics of RA were determined through ssGSEA analysis, gene differential expression analysis, fuzzy c-means clustering algorithm, and XGboost algorithm. Random Ferns classifiers (RFs) are used to identify new bone marrow markers for AML.

Results SELL, PTPRC, IL7R, CCR7, and KLRB1 were able to distinguish leukemia cells from normal cells well, with AUC values higher than 0.970.

Conclusion Genes characterizing the immune heterogeneity of RA are associated with AML, and KLRBA may be a potential target for AML treatment.

Keywords Biomarkers · Acute myeloid leukemia · Machine learning · Therapeutic target

1 Introduction

A growing number of studies have highlighted a significant correlation between autoimmune diseases and malignancies. Specifically, cohort studies have demonstrated an elevated risk of leukemia in individuals with rheumatoid arthritis (RA), and an alarming surge of approximately 52% in acute myeloid leukemia (AML) incidence over the last decade (2009–2018) [1]. Importantly, both RA and AML involve profound alterations in the bone marrow microenvironment. In the context of RA, this autoimmune disease is characterized by the aggregation and infiltration of immune cells within the synovium prior to the manifestation of inflammation. Conversely, osteoarthritis (OA) is characterized by sterile inflammation without an immune response. Therefore, the synovium emerges as a pivotal site harboring potential immune

✉ Wenjie Teng, wfteng@wfmcc.edu.cn; Chengzhi Jiang, 617893076@qq.com; Wenjuan Jiang, 1078595795@qq.com; Pengtao Liu, lpt1978@126.com; Wenxue Sun, 2639973659@qq.com | ¹School of Public Health, Shandong Second Medical University, Weifang, Shandong 261053, People's Republic of China. ²Qingdao Municipal Hospital (Group), Qingdao, Shandong 266000, People's Republic of China. ³Computer Teaching and Research Office, Shandong Second Medical University, Weifang, Shandong 261053, People's Republic of China.



heterogeneity differences between RA and OA. Although we have identified immunological heterogeneity features in RA, the expression and molecular mechanisms of these features in the context of AML remain unexplored.

In this study, we aimed to obtain relevant biomarkers by studying the immune heterogeneity of RA, to explore the relationship between these biomarkers and AML, to analyze whether they have the potential to be used as diagnostic biomarkers and therapeutic targets for AML, and to compare their diagnostic capabilities with those of the biomarkers that have been previously identified, such as CD13, CD33, CD14, and others [2–7]. To ensure the rigor and accuracy of the study, we adapted each machine learning model and mined a large amount of data in the study.

2 Materials and methods

2.1 Data acquisition

RNA sequencing datasets GSE36700 and GSE55584 for RA and OA were obtained from the GEO database (<http://www.ncbi.nlm.nih.gov>) with sample sizes of 25 and 16, respectively. The AML dataset GSE6891 (n = 536) with different cytological genetic risk profiles and CEBPA mutation types was also obtained from the GEO database. Mouse bone marrow single cell sequencing data GSM6893371 and GSM6893373 were also obtained from the GEO database. RNA sequencing dataset GDC TCGA Acute Myeloid Leukemia for AML was obtained from UCSC Xena (<https://xenabrowser.net/datapages/>) with a sample size of 151. RNA sequencing dataset for normal human was obtained GTEX-TOIL RSEM tpm and screened bone marrow sequencing samples with a sample size of 70. All data were converted to TPM data and $\log_2(\text{TPM} + 1)$ processed. This batch effect was removed in all data set merging processes.

2.2 Gene differential expression analysis

Firstly, gene differential expression analysis was realized using the 'limma' package in R Studio. The cut-off criteria were $|\log_2\text{fold-change}| > 1$ and $P < 0.05$. Secondly, after the gene differential expression analysis was completed, PPI networks of differential genes were constructed using the STRING protein interactions database (<https://cn.string-db.org/>), and key genes were screened using the 11 topology methods of the CytoHubba plugin in Cytoscape. Finally, Metascape software (<https://metascape.org/>) was used for gene function enrichment analysis.

2.3 Fuzzy C-Means clustering

The 'Mfuzzy' package [8] in R Studio was originally developed as a clustering method for processing gene expression or protein expression profile data, with the core algorithm based on Fuzzy C-Means Clustering (FCM). This method is used to cluster the key genes screened by CytoHubba to further determine the final features.

2.4 XGboost

eXtreme Gradient Boosting (XGBoost), is a Boosting algorithm that implements machine learning algorithms in the framework of Gradient Boosting. XGBoost provides Parallel Tree Boosting (also known as GBDT, GBM) to solve regression or classification problems quickly and accurately. In this study, after the fuzzy C-mean clustering analysis, XGboost regression models were constructed separately for each cluster and the optimal parameters were determined using fivefold cross-validation and grid search method.

2.5 Recursive feature elimination based on cross validation

Recursive feature elimination is essentially a greedy algorithm for finding the optimal subset. The main idea is to repeatedly construct the model and then select the best (or worst) features, select the selected features, and then repeat the process on the remaining features until all the features have been traversed. The order in which the features are eliminated in this process is the ranking of the features. Since purely recursive feature elimination is prone to overfitting problems, cross validation based recursive feature elimination has emerged. In this study, recursive feature elimination based on fivefold cross-validation is used to filter the variables.

2.6 Random ferns classifier

Random Ferns is a classifier based on the plain Bayesian algorithm and is often viewed as a non-hierarchical random forest [9]. The process of constructing the 'fern' differs from the decentralized structure of the 'tree' constructed by the random forest, but a vertical structure. In addition, the key parameter 'depth' must be in the range [1, 16]. In order to find the optimal parameters, tenfold cross-validation repeated three times and grid search method were applied to determine the optimal parameters. In this study, a random fern classification model was constructed to identify bone marrow diagnostic biomarkers for AML.

2.7 Dimensionality reduction algorithm

t-SNE, UMAP, and PCA are three different dimensionality reduction algorithms. Dimensionality reduction algorithms actually recalculate two or three new variables for a number of variables in the sample to achieve the purpose of extracting and synthesizing valid information and rejecting useless information. In other words, the generation of new variables after dimensionality reduction is based on all the previous variables, and together with the grouping information of the sample, the purpose of checking whether the grouping is correct or not can be achieved.

2.8 Online analysis platform

Drug sensitivity analysis was performed by CTD database (<http://ctdbase.org>) and CTRP database based on GSCA platform (<http://bioinfo.life.hust.edu.cn/GSCA/#/>). SNV analysis was performed based on GSCA platform. Validation of gene expression was performed based on the GEPIA2 platform (<http://gepia2.cancer-pku.cn/#index>) and BioGPS database (biogps.org/). ceRNA network construction was performed based on data from the ENCORI database (<http://starbase.sysu.edu.cn/>), miRWalk database (<http://mirwalk.umm.uni-heidelberg.de/>), and HMDD database (<http://www.cuilab.cn/hmdd/>). Survival and prognostic analysis based on Kaplan Meier Plotter database (<http://kmplot.com/analysis/index.php?p=background>).

3 Results

3.1 Gene differential expression analysis

We evaluated the immune infiltration status of each sample using ssGSEA (Fig. 1A). In order to reduce research error, we excluded three OA samples, so that the group with high immune infiltration scores and the group with low scores were divided into the RA group and the OA group, respectively. Gene differential expression analysis was conducted on the above groups, resulting in 566 differentially expressed genes. Differential gene expression analysis was performed on RA and OA groups without ssGSEA, and 642 differentially expressed genes (DEGs) were identified (Fig. 1B). Take the intersection of the two parts to obtain 544 DEGs (Fig. 2A), and then use the STRING database and the CytoHubba plugin in Cytoscape to obtain 50 hub genes (Fig. 2B). Functional and pathway enrichment analysis of these differential genes showed that most of these genes were enriched in adaptive immune response, lymphocyte activation, positive regulation of immune response, primary immune response, lymphocyte activation (Fig. 2C).

3.2 Screening of pivotal genes

Fuzzy C-means clustering analysis was performed on these 50 pivotal genes, and the results showed that these 50 pivotal genes were clustered into 4 classes (Fig. 3A). The XGboost regression model was constructed separately for each cluster, and the optimal model parameters were determined using fivefold cross-validation and grid search method. The results of the optimal parameters are shown in the table below (Table 1). Then recursive feature elimination based on fivefold cross validation was used to filter the variables, and the results of the number of variables selected according to recursive feature elimination for each model are shown. The first two important variables of each model were selected according to the ranking of importance, and the final identified key genes for RA immune heterogeneity were IFNG, IL7R, CCR7, KLRB1, CXCL9, CXCL13, SELL, and PTPRC (Fig. 3B–E).

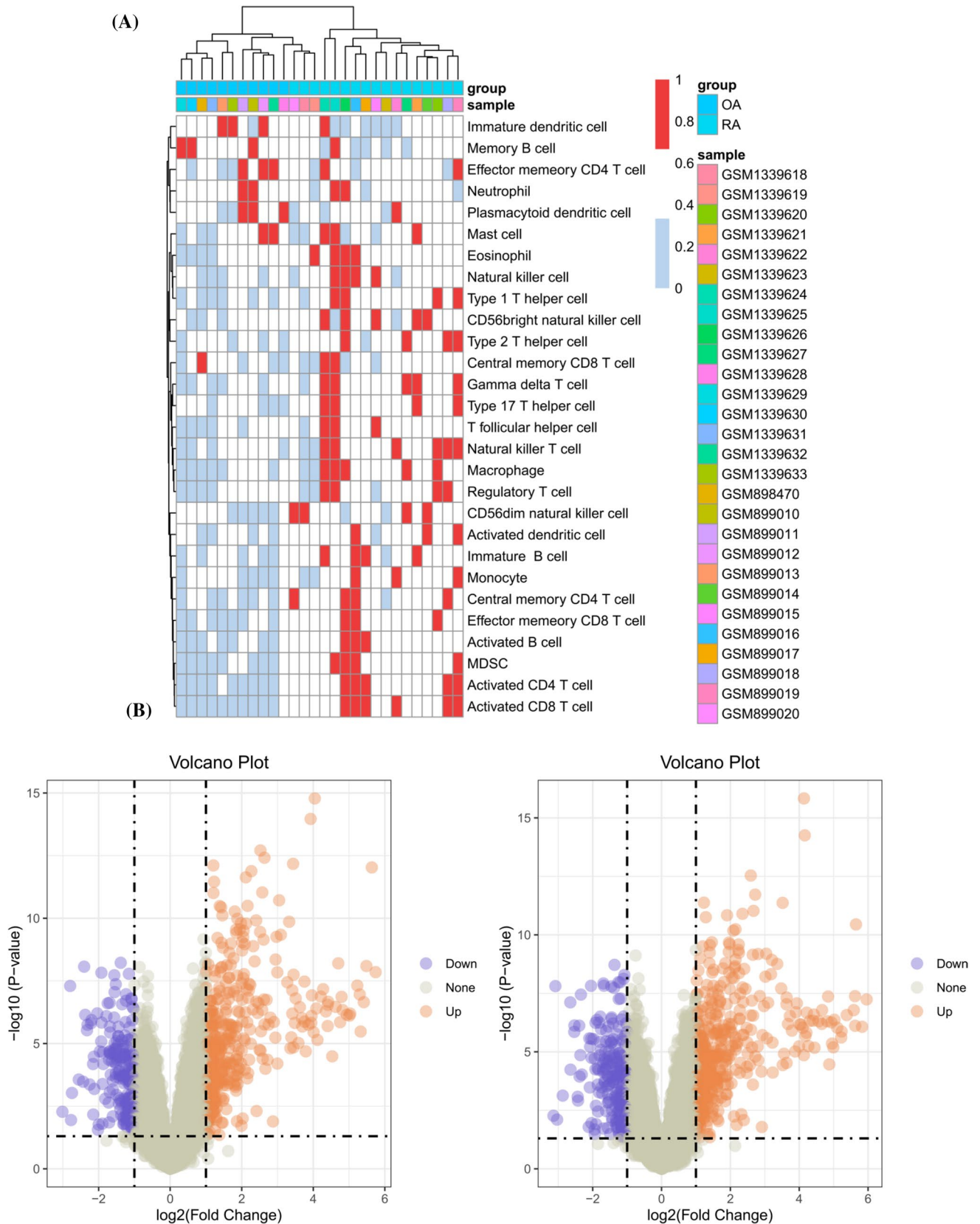


Fig. 1 Results of two gene differential expression analyses **(A)** Immune infiltration status of all samples **(B)** Number of differential genes up-regulated and down-regulated

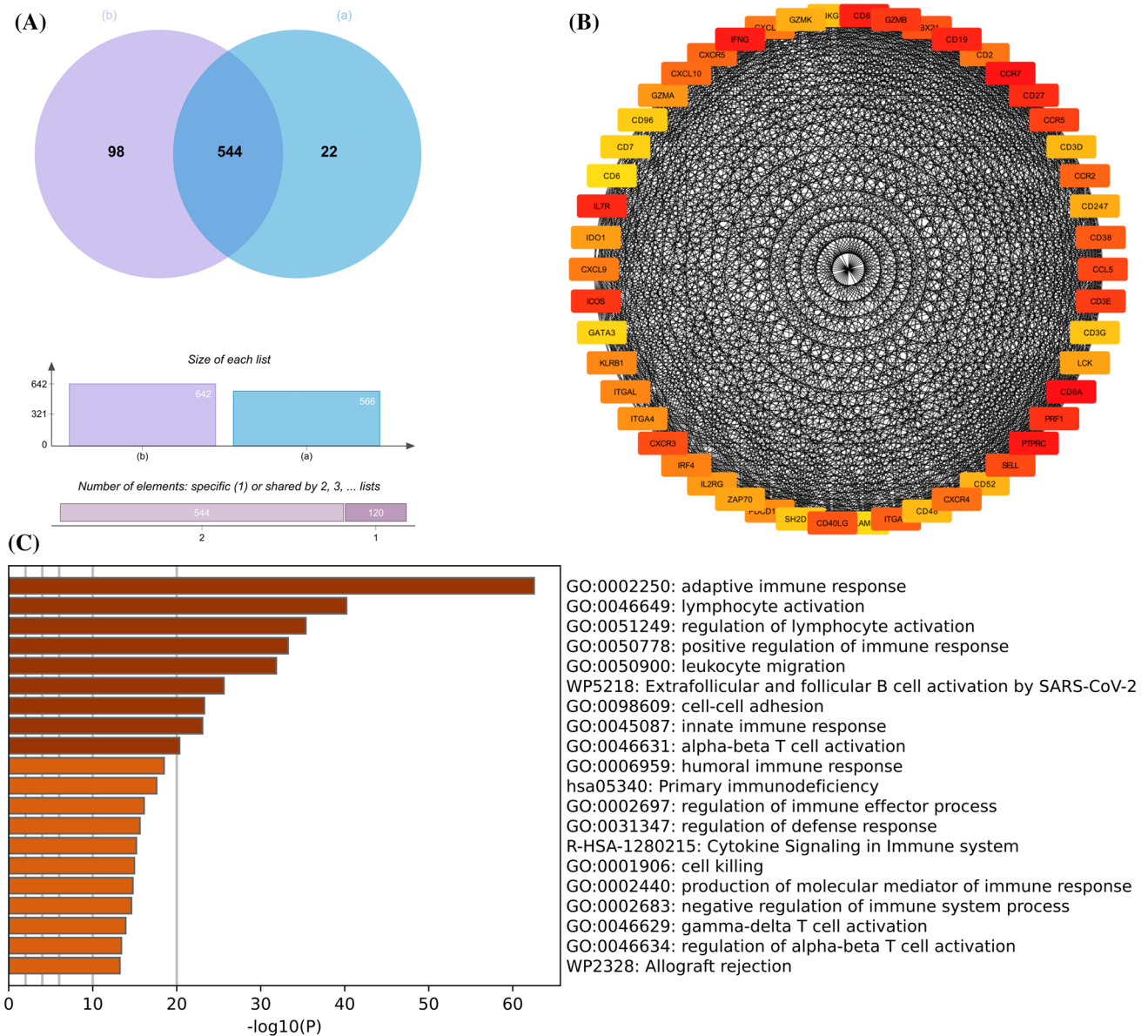


Fig. 2 Functional and pathway enrichment analysis of differentially expressed genes in RA and OA **(A)** Number of overlapping differentially expressed genes in RA and OA before and after immune infiltration assessment **(B)** Protein Interaction Network Screening for pivotal genes **(C)** Functional and pathway enrichment analysis of pivotal genes

3.3 Immune heterogeneity characterizing RA in AML

Use the 8 genes identified above as the immune heterogeneity characteristic gene set of RA to study their expression levels in AML. The analysis of data in GEPIA2 showed that there was no statistically significant difference in the expression of CXCL9, CXCL13, and IFNG in AML bone marrow (Fig. 4A). There is a statistically significant difference in the expression of IL7R, CCR7, KLRB1, SELL, and PTPRC in AML bone marrow and normal human bone marrow, and the lowest expression value in AML is almost higher than the highest value in the normal group, and this expression difference is more pronounced in AML than in other tumors (Fig. 4B–F). To validate this result, we chose another bone marrow RNA sequencing dataset of AML (n = 151) to merge with the normal samples and remove the batch effect. The results were confirmed by a wilcoxon test of the expression of the two groups, which showed that the differences in the expression of IL7R, CCR7, KLRB1, SELL, and PTPRC were statistically significant between the AML group and the normal group (Fig. 4G). In addition, we conducted another validation of the expression of the above five genes by the BioGPS database (GEO-based cohort

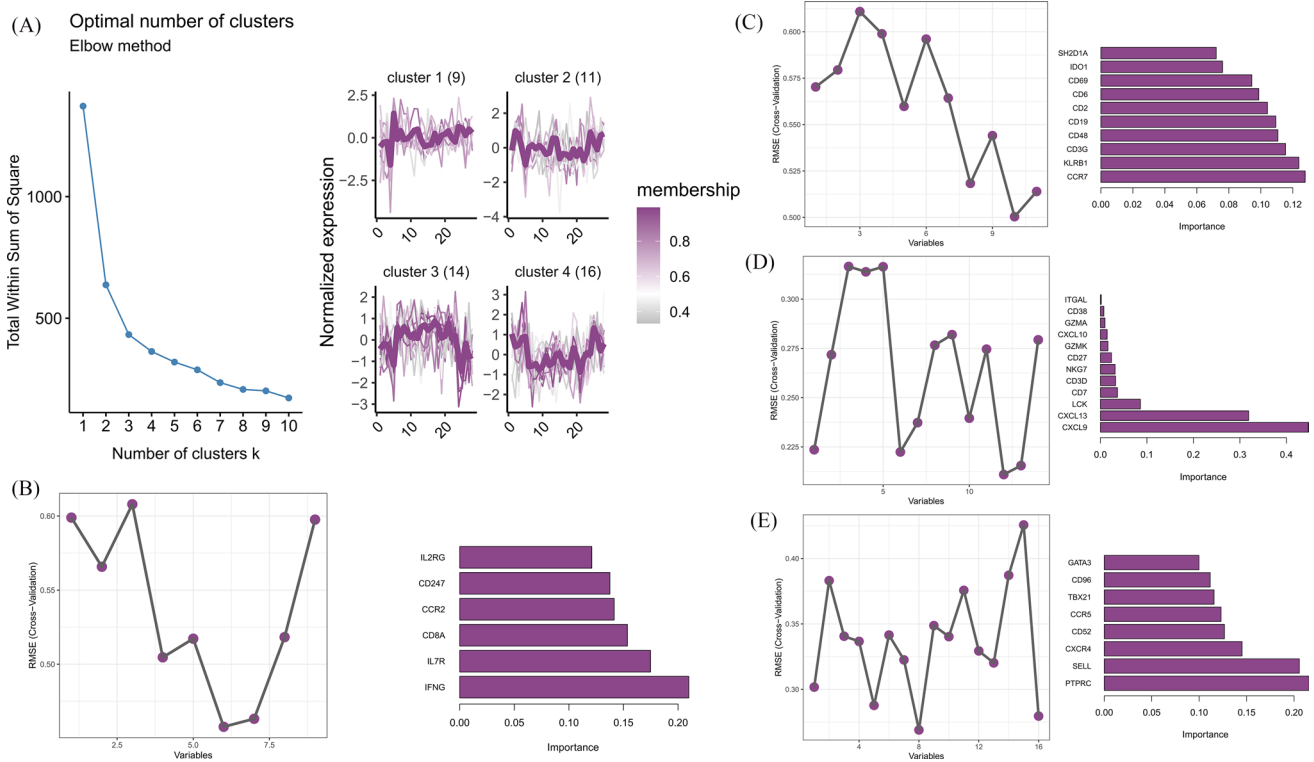


Fig. 3 Identify key genes (a) Fuzzy C-means clustering results (b–e) The importance of genes in each cluster

Table 1 Parameters of each XGboost regression model

	Cluster1	Cluster2	Cluster3	Cluster4
nrounds	50	50	50	100
Max_depth	3	1	1	2
Eta	0.3	0.4	0.3	0.4
Gamma	0	0	0	0
Colsample_bytree	0.8	0.6	0.6	0.6
Min_child_weight	1	1	1	1
Subsample	0.5	0.5	0.5	1

data). This validation data included 29 normal and 99 AML cohorts, and the validation results were consistent with our previously obtained results (Fig. 4H).

3.4 Machine learning to identify AML patients

To identify bone marrow diagnostic biomarkers for AML, a random fern classifier was constructed to distinguish AML patients from normal individuals. The process of constructing the random fern classification model uses tenfold cross-validation repeated three times and grid search method to determine the optimal parameters (Fig. 4I–M). The merged dataset from the previous step was split 70% as the training set and the whole dataset was used as the validation set. The results showed that the AUC value of each biomarker in the validation set was above 0.970 (Fig. 5 A). In addition, the dimensionality reduction algorithm was also used to check whether the grouping was correct. Three algorithms, t-SNE, UMPA, and PCA, were used to analyze the grouping of AML and normal individuals based on these five features and the grouping information of each sample. The results showed that regardless of which algorithm the AML group and the normal group were clearly distinguished (Fig. 5B–D), which further illustrates the value of these five features in diagnosing AML.

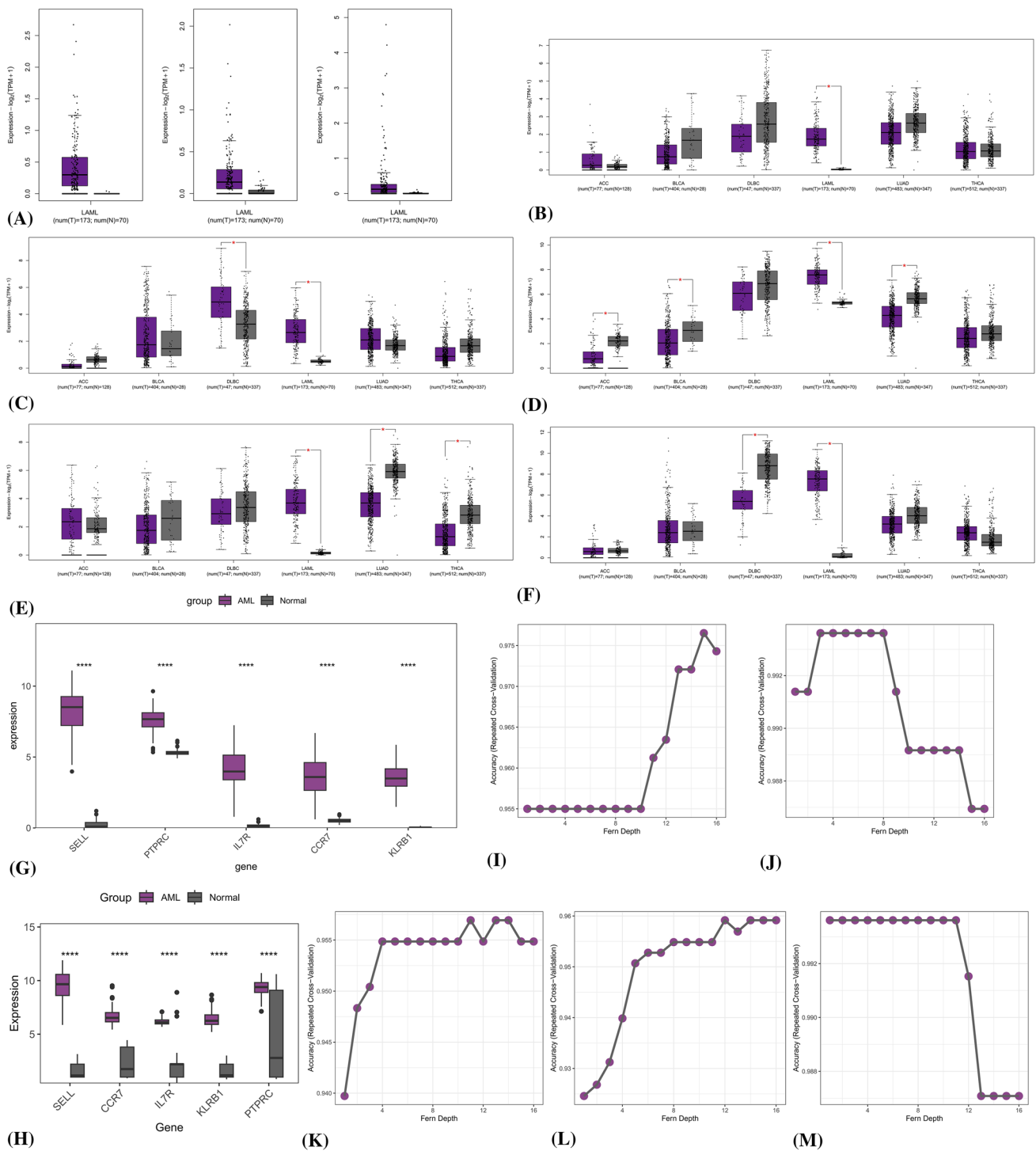


Fig. 4 Determination and validation of gene expression and search for the optimal parameters of RFs models (A) Expression of IFNG, CXCL9, and CXCL13 in AML bone marrow (B–F) Expression of IL7R, CCR7, KLRB1, SELL, PTPRC in AML bone marrow and other tumors (G) Validation of IL7R, CCR7, KLRB1, SELL, PTPRC expression (H) Validation of expression of SELL, PTPRC, IL7R, CCR7, and KLRB1 using BioGPS database (I–M) Determine the optimal parameters for RFs classification models

In addition, a comparison between old and new AML bone marrow biomarkers was performed. CD33, CD34, and CD14 are AML diagnostic biomarkers that have long been proven by research. Following the same research methodology described above, the optimal parameters of the randomized fern model are shown below (Fig. 5E–G), and the calculated AUC values for CD33, CD34, and CD14 were 0.893, 0.923, and 0.950, respectively (Fig. 5H).

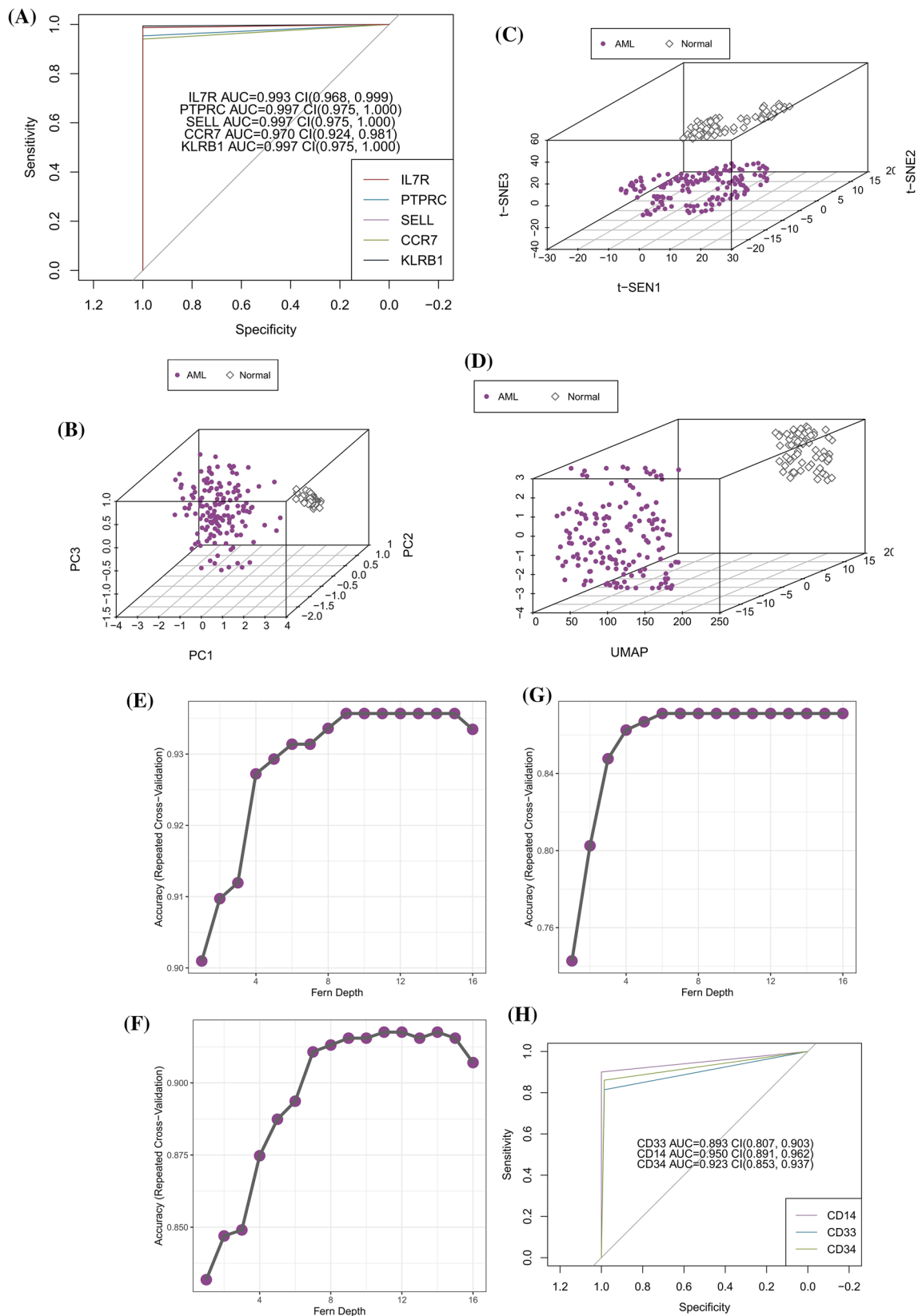


Fig. 5 Diagnostic value of bone marrow biomarkers in AML **(A)** ROC analysis results of bone marrow biomarkers in AML **(B–D)** UMAP, PCA, t-SNE analysis results **(E–G)** Selection of the optimal parameter ‘depth’ for random fern model **(H)** ROC analysis results of traditional AML biomarkers

3.5 Immune infiltration analysis

The results indicate that PTPRC is positively correlated with MDSC, and negatively correlated with Activated CD4 T cells and CD56bright natural killer cells. IL7R showed a positive correlation with Activated CD8 T cell, Activated B cell, and Activated CD4 T cell, while a negative correlation with Immature dendritic cell, Macrophage, Monocell, Eosinophel, and Memory B cell. The relationship between the expression of CCR7 and immune cells is similar to that of IL7R, while the expression of SELL shows a negative correlation with all immune cells. The expression of KLRB1 is positively correlated with Activated CD8 T cells and negatively correlated with all other immune cells (Fig. 6A). The expression of these five biomarkers was positively correlated with the expression of common immune checkpoints. In comparison, the expression of SELL is negatively correlated with the expression of more immune checkpoints. PTPRC did not show a particularly significant correlation with the expression of CD family genes, while KLRB1 did not show a particularly significant correlation with the expression of MHC family genes (Fig. 6B–E).

3.6 Drug sensitivity analysis and SNV analysis

The relationship between the expression of IL7R, CCR7, KLRB1, SELL, PTPRC and drug sensitivity was analyzed by CTRIP database. The results showed that the expression of IL7R showed a negative correlation with the sensitivity of apicidin, belinosta, ciclopirox, cytarabine hydrochloride, and panobinostat, and that IL7R may affect the therapeutic efficacy of AML patients (Fig. 7A). A search for drugs that could reduce IL7R expression in the CTD database showed that acetyl-cysteine was an effective drug for reducing IL7R expression. Missense mutations were the most common in pan-cancer SNV analysis of target genes, and PTPRC had the highest proportion of SNVs in pan-cancer samples, which was particularly prominent in small cell lung cancer (Fig. 7B).

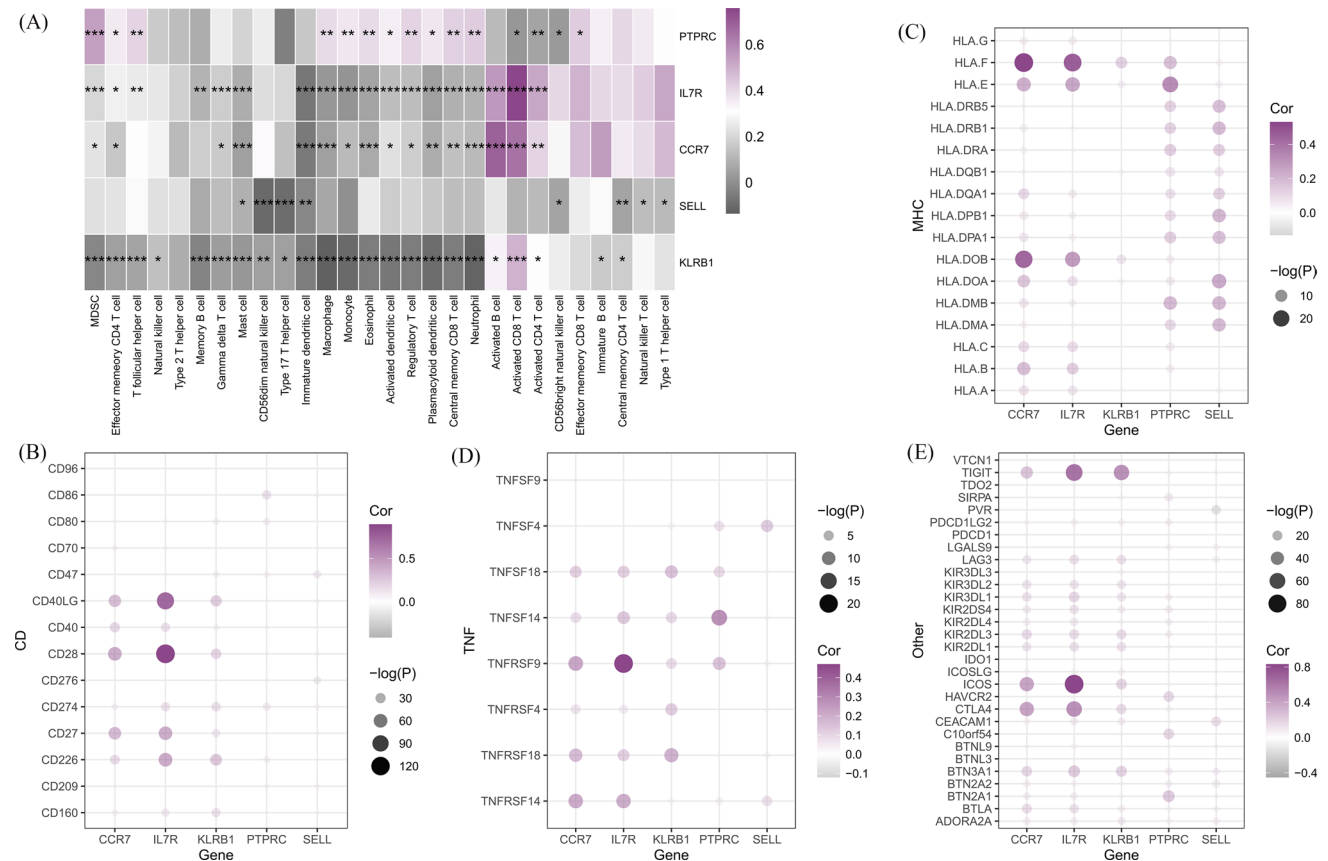


Fig. 6 Immune infiltration analysis results (A) The correlation between diagnostic genes and immune cells (B–E) Correlation between diagnostic genes and immune checkpoints

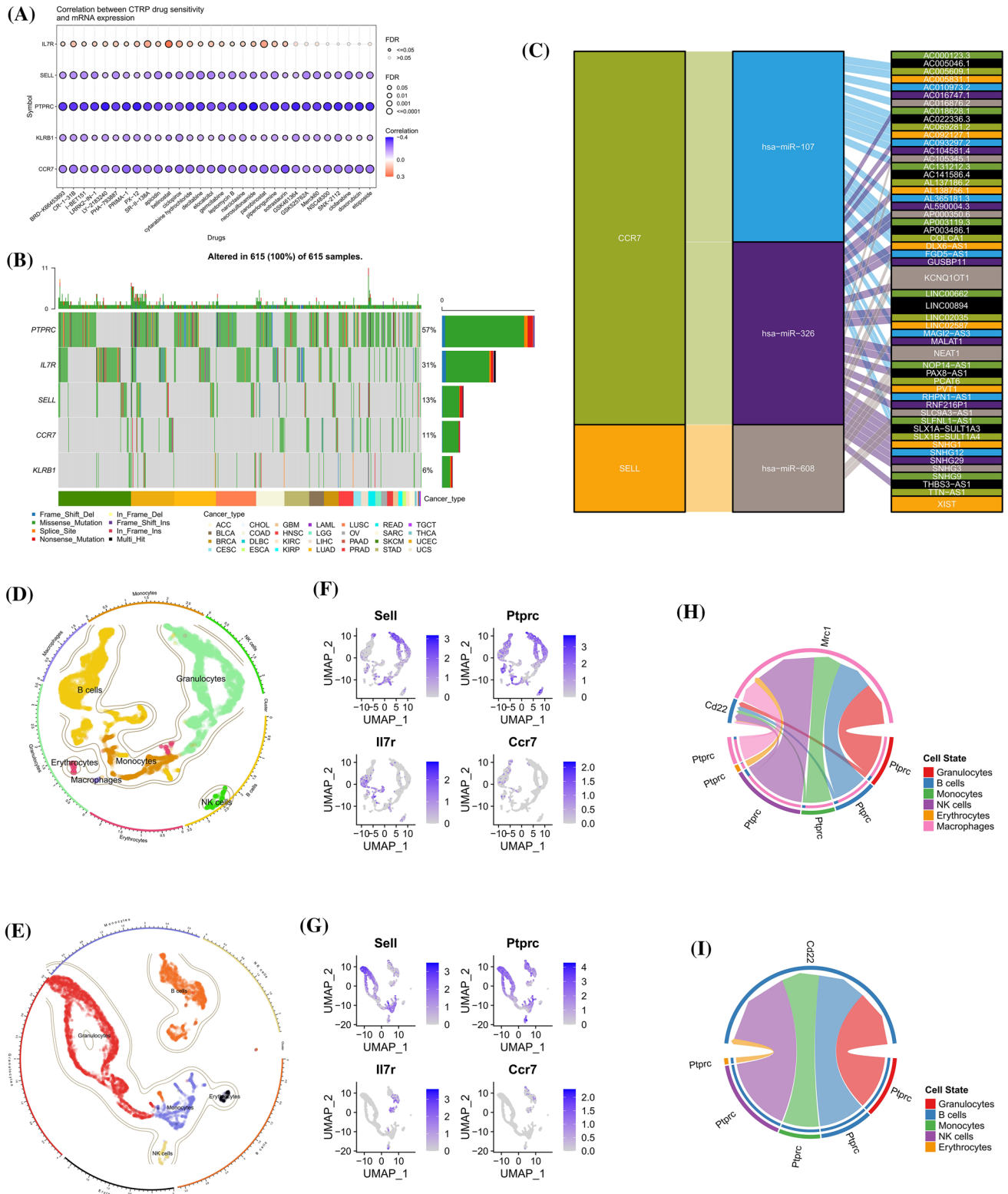


Fig. 7 GSEA online analysis and single cell sequencing analysis results **(A)** Drug sensitivity analysis results **(B)** SNV analysis results **(C)** Building a ceRNA network **(D)** Single cell sequencing analysis results in the control group **(E)** Single cell sequencing analysis results of RA group **(F)** Expression distribution of target genes in various cells of the control group **(G)** Expression distribution of target genes in various cells of the RA group **(H)** CD45 signaling pathway in the control group **(I)** CD45 signaling pathway in the RA group

3.7 Construction of ceRNA network

1386 miRNAs related to diagnostic genes were obtained from the miRWalk database, and 178 miRNAs related to AML were obtained from the HMDD database. 2 mRNAs were obtained by taking their intersection: *SELL* and *CCR7*, and 4 miRNAs: *hsa-miR-107*, *hsa-miR-1299*, *hsa-miR-608*, *hsa-miR-326*. 3 miRNAs and 54 lncRNAs were obtained by searching for miRNA-lncRNA relationship pairs from ENCORI database for these 4 miRNAs then the ceRNA network composed of mRNA-miRNA-lncRNA was constructed based on the above results. The results are shown below (Fig. 7C).

3.8 Analysis of single cell sequencing data for RA and control

Analysis of bone marrow single-cell sequencing data from RA and control mice showed that the expression of *SELL* and *PTPRC* was significantly stronger in the RA group than in the control group (Fig. 7D–G). Important interaction pairs associated with the CD45 signaling pathway in the RA group were exclusively contributed by *Ptprc*-*Cd22*, whereas important interaction pairs associated with the CD45 signaling pathway in the control group were mainly contributed by *Ptprc*-*Mrc1*, with *Ptprc*-*Cd22* contributing only a very small fraction (Fig. 7H, I).

3.9 The correlation between the expression of immune heterogeneity characteristic genes and different subtypes of AML

We first analyzed the expression of *SELL*, *CCR7*, *IL7R*, *KLRB1*, and *PTPRC* in relation to different cytogenetic risks in dataset GSE6891. We found that the expression of *SELL*, *IL7R*, and *PTPRC* was significantly higher in patients with cytogenetic risk of poor than in patients with cytogenetic risk of good (Fig. 8A–E). We divided AML patients into two groups of high and low expression according to the median of their respective expression of these five genes, and then analyzed the relationship between different expression status of different genes and different cytogenetic risks, we found that only the expression of *SELL* was associated with different cytogenetic risks ($r=0.104$, $P<0.05$) (Table 2). Then we further investigated the association of *SELL*, *CCR7*, *IL7R*, *KLRB1*, *PTPRC* with different *CEBPA* mutation types. The results showed that for different mutation types, the expression of the above five genes was significantly different (Fig. 8F). According to the same method described above, AML patients were categorized into high and low expression groups, and the expression of all genes except *KLRB1* was correlated with different types of *CEBPA* mutation, and the expression of *SELL* and *PTPRC* was more correlated with different types of *CEBPA* mutation compared with the other three genes (Table 2). Finally, we analyzed the relationship between high and low expression of the above five genes and the prognosis of AML patients according to different *CEBPA* mutation types. The results showed that the survival probability of patients with high expression of *IL7R* and *KLRB1* was significantly higher in patients with wild-type mutations, which may be associated with a favorable prognosis of the patients, and that the survival probability of patients with high expression of *PTPRC* was significantly lower, which may be associated with a poor prognosis of the patients (Fig. 8G). In patients with mutation type mutant, high expression of *PTPRC* also predicted a significantly lower probability of survival, which may also indicate an association with a poor prognosis for the patient (Fig. 8H).

4 Discussion

RA is an autoimmune disease that is associated with an increased risk of not only lymphoproliferative disorders, but also myeloid malignancies, increasing the risk of developing AML [10, 11]. As these studies continue to grow and intensify, this seems to open up the possibility of exploring diagnostic biomarkers from one disease to another. Due to the chronic immune stimulation or bone marrow infiltration of RA, some risk features at the molecular level are progressively manifested before it causes AML, and these risk features are not found in normal individuals, and these molecular features underpinning the development of AML remain largely speculative. Some of the differences between RA and normal individuals at the molecular level are reflected in the immune heterogeneity of RA. Therefore, we will first investigate the immunological heterogeneity features of RA and then use this as a starting point to explore whether these features are relevant to the diagnosis of AML.

Firstly, in this study, we performed two differential expression analyses on RA and OA samples before and after calculating the immune infiltration scores of each sample, and 544 DEGs were obtained by intersecting the results

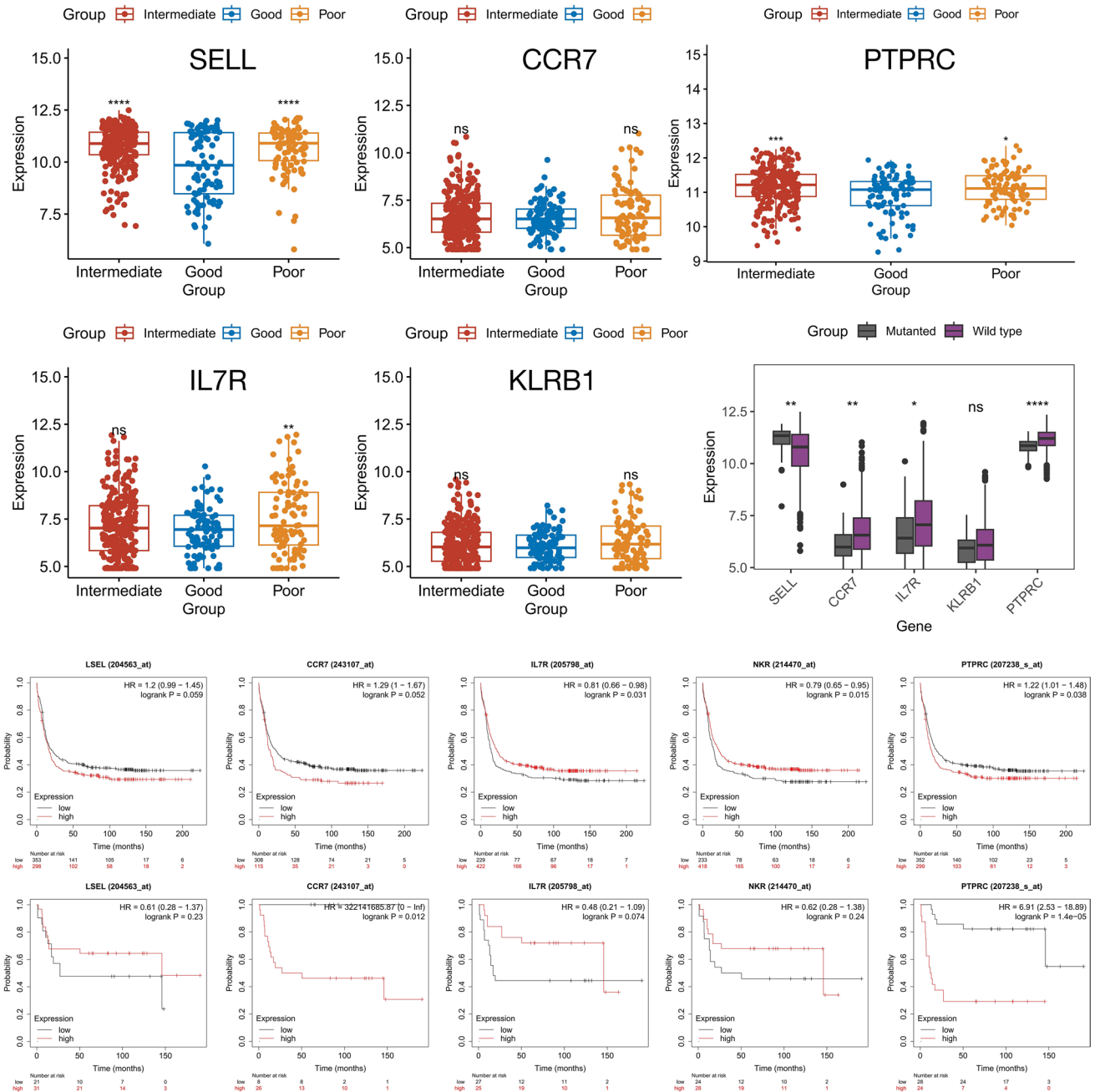


Fig. 8 The relationship between different gene expressions and AML subtypes **(A–E)** Differential expression of different genes under different cytogenetic risk groupings **(F)** Differential expression of different genes under different CEBPA mutation types **(G)** Survival and prognosis analysis of patients with wild-type mutations under different gene expression patterns **(H)** Survival and prognosis analysis of patients with mutated mutations in different gene expression patterns

Table 2 The relationship between the expression patterns of different genes and the cytological genetic risk and different CEBPA mutation states in AML patients

	Cytogenetic Risk		CEBPA mutation	
	Cor (Spearman)	<i>P</i>	Cor (Cramer's V)	<i>P</i>
SELL	0.104	<0.05	0.152	<0.05
CCR7	0.014	>0.05	0.100	<0.05
IL7R	0.028	>0.05	0.100	<0.05
KLRB1	0.035	>0.05	0.040	>0.05
PTPRC	0.056	>0.05	0.238	<0.05

of the two differential analyses. 544 DEGs were obtained through functional and pathway enrichment analyses of these DEGs, and the results showed that these DEGs were mainly enriched in the adaptive immune response, positive regulation of immune response, primary immunodeficiency, leukocyte migration, cytokine signaling in the immune system, and the expression of the immune infiltration scores of each sample response, positive regulation of immune response, primary immunodeficiency, leukocyte migration, and cytokine signaling in Immune system. These enrichment entries are from different databases but the results of functional and pathway enrichment are not very different. In addition, it has been shown that the biological processes of the immune system are crucial for the formation of a complex bone marrow micro-environment [12, 13], so that understanding the immunological heterogeneity of RA will help to develop effective bone marrow markers for AML diagnosis.

Secondly, eight genes associated with RA immune heterogeneity were obtained by fuzzy C-mean clustering and post-tuning XGboost regression model calculations: IFNG, IL7R, CCR7, KLRB1, CXCL9, CXCL13, SELL, and PTPRC. With these eight immune heterogeneity genes of RA as a set of characterized genes to be investigated in AML, we evaluated their expression in the expression in AML. Except for IFNG, CXCL9, CXCL13, which were not statistically significant in AML above normal, the difference between AML and normal was statistically significant for the remaining five genes, and this difference was highly significant. We also validated the expression of these 5 genes using a new dataset and analyzed their diagnostic value for AML using a random ferns classifier. Typically, shed forms of SELL are often studied in conjunction with AML, and plasma from AML patients shows high levels of shed forms of selections [14–16]. Little attention has been paid to the amount of SELL expressed in AML bone marrow. In this study, after analyzing a large amount of data, we found that the expression of SELL is much higher in AML bone marrow than in normal subjects, which provides a new biomarker for bone marrow diagnosis of AML. In addition, it has been found that interferon α up-regulates the expression of SELL [17], and interferon α is often used in the treatment of RA [18, 19], which may bring a potential risk of AML development in RA patients. In this study, KLRB1 and CCR7 were found to be barely expressed in normal human bone marrow, but highly expressed in AML bone marrow, and besides AML, they were also over-expressed in a few types of cancers, such as testicular germ cell tumors (TGCT), kidney renal clear cell carcinoma (KIRC), pancreatic cancer (PAAD), etc., and the tissues of the three cancers studied were not closely related to the bone marrow, so KLRB1 and CCR7 will be bone marrow biomarkers helpful for the diagnosis of AML. Previous studies have shown that over-expression of PTPRC can predict poor prognosis and serve as a therapeutic target for pediatric AML [20]. In the present study PTPRC can be used as a biomarker for AML diagnosis. Earlier studies have shown that IL7R is extremely low in bone marrow [21], which corroborates with this paper, so the presence of a significant increase in bone marrow of AML patients combined with the results of ROC analysis (AUC = 0.993) suggests that IL7R is a suitable bone marrow biomarker for AML diagnosis.

And then, this study analyzed the relationship between these five genes and the AML immune microenvironment and AML therapeutic drug sensitivity, and attempted to construct a ceRNA network based on these five genes. This study found that KLRB1 can be a target for AML therapy. In a previous study on glioma, it was shown that activation of CD161 would weaken T-cell responses to tumor cells, thereby inhibiting the utility of immunotherapy. Knockdown of KLRB1, the gene encoding CD161, strongly enhanced the ability of T cells to attack glioma cells and reduced T cell exhaustion [22]. In this study, KLRB1 showed almost negative correlation with different types of T cells and positive correlation with almost all common immune checkpoints, and its expression level was also much higher than that of normal individuals. Therefore, this study suggests that KLRB1 could be a new target for AML treatment or could be used in the development and design of novel immunotherapeutic drugs. This study also found that IL7R affects chemotherapy resistance. In fact, some studies have already confirmed that IL7R is related to chemotherapy resistance [23–26], so this study found a drug that can reduce the expression of IL7R through the Drug Interaction Database: acetylcysteine, which may be helpful in reducing chemotherapy resistance. To fully understand the role of these diagnostic genes in AML, this study also constructed a ceRNA network. miRNAs and lncRNAs play key roles in gene regulation and cancer biology. miRNAs are small non-coding RNAs that are known to bind and control mRNA expression [27–31]. In the ceRNA network, lncRNAs act as miRNAs to regulate gene expression and participate in cancer development. The construction of the ceRNA network will better investigate the role of diagnostic gene genes in AML and their potential mechanisms. In the results of the cell communication analysis of RA and control mice in this study, it is not difficult to see that completely different from the control group, the CD45 signaling pathway in RA is exclusively contributed by Ptprc-Cd22, which in turn is a biomarker for a variety of malignant tumors such as leukemia and non-Hodgkin's lymphoma, which may suggest that the CD45 signaling pathway is of great importance to be investigated in the transition from RA to leukemia.

Finally, we analyzed the relationship between the expression of these five biomarkers and the different cytogenetic risks of AML patients with different CEBPA mutation status. We found that the expression of SELL was higher the worse

the cytogenetic risk of AML, while the expression of PTPRC was highest in the group with Intermediate cytogenetic risk. The expression of SELL was significantly higher in patients with CEBPA mutation type as mutated than in wild type patients. The expression of all genes except SELL was significantly higher in patients with wild type than in patients with mutated type. In addition, different expression patterns of biomarkers other than KLRBA correlated with CEBPA mutated type, although this correlation was not strong. The above results demonstrate that the expression of genes characterizing the immune heterogeneity associated with RA is associated with different cytogenetic risks and CEBPA mutation types in AML.

This study explored the relationship between genes characterizing the immune heterogeneity of RA and AML, but further research is needed for an in-depth exploration of the transition from RA to AML. Indeed, the low frequency of RA developing into AML may in some cases be associated with certain somatic mutations, such as UBA1 in Vexas syndrome. One study identified Vexas syndrome as a myeloid-driven inflammatory disorder caused by somatic mutations in the UBA1 gene, further exposing the increasingly recognized overlap between hematologic disorders and autoimmune and/or autoinflammatory manifestations [32]. It has also been demonstrated that there is a commonality in the mechanistic pathways involved in RA, AML, and Vexas syndrome [33], and that all involve a profound alteration of the inflammatory response and the bone marrow microenvironment. Somatic mutations in the ASXL1 gene may also be a contributing factor in the progression of RA to AML. Studies have shown that somatic mutations in the ASXL1 gene correlate with RA [34], and ASXL1 mutations are highly specific in diagnosing AML [35]. In addition, mutations in DNMT3A and TET2, which are closely associated with RA, may also be important in the progression of RA to AML [36, 37].

Although this study analyzed a large amount of data using machine learning and bioinformatics methods to identify five biomarkers associated with AML, there are certain shortcomings. One is that the study may have been affected by potential confounding factors. The first is that although OA was chosen as a control in this study to reflect the immune heterogeneity of RA, after all, the mechanistic pathways of RA and OA are different, and to some extent this may interfere with the data analysis. Secondly, it is not clear whether the RA or OA samples had received treatment, as some drugs may have affected their immune function during the treatment process, which may have resulted in altered expression of certain genes. Despite some confounding, this study explored the relationship between genes that characterize immune heterogeneity in RA and AML, demonstrating that these genes are useful in distinguishing healthy cells from leukemic cells.

5 Conclusion

Current research evidence clearly demonstrates the association of SELL, PTPRC, IL7R, CCR7, and KLRB1 with AML, which are critical for distinguishing healthy cells from leukemic cells in AML. In addition, KLRB1 is expected to be a new target for AML treatment.

Acknowledgements Special thanks to the Computer Science Department of Shandong Second Medical University for their efforts in revising this study.

Author contributions CZJ, WJJ, WXS, and WJT conceived the entire research design. CZJ analyzed all data and contributed to writing the original draft manuscript. WXS and WJT critically reviewed and revised the draft manuscript, and WJJ provided guidance and suggestions for revision. PTL provided guidance for the final revision of this study. All authors contributed to writing the manuscript, read and approved the final version for submission.

Funding No funding.

Data availability In this study data are available from the GEO database (<http://www.ncbi.nlm.nih.gov>) and ucsc xena database (<https://xenabrowser.net/datapages/>) to download.

Declarations

Competing interests The authors declare that there is no conflict of interest.

Ethics approval All data in this study are from public databases and no additional ethical proof is required.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Luo X, He Y, Xu W, Liu M, Zhao Z, Peng L, He C, Chen J. The risk of leukemia in patients with rheumatoid arthritis: a systematic review and meta-analysis. *Clin Rheumatol*. 2021;40(4):1283–9.
2. Li J, Xu J, Abruzzo LV, Tang G, Li S, You MJ, Lu G, Jabbour EJ, Deng Q, Bueso-Ramos CE, Medeiros LJ, Yin CC. Acute myeloid leukemia with t(4;12)(q12;p13): an aggressive disease with frequent involvement of PDGFRA and ETV6. *Oncotarget*. 2017;9(13):10987–94.
3. Ehninger A, Kramer M, Röllig C, Thiede C, Bornhäuser M, von Bonin M, Wermke M, Feldmann A, Bachmann M, Ehninger G, Oelschlägel U. Distribution and levels of cell surface expression of CD33 and CD123 in acute myeloid leukemia. *Blood Cancer J*. 2014;4(6): e218.
4. Estey EH. Acute myeloid leukemia: 2013 update on risk-stratification and management. *Am J Hematol*. 2013;88(4):318–27.
5. Khara R, Ahmed F, Mundada MC, Devi SG, Murthy SS, Lavanya N, Rajappa SJ, Mallavarapu KM, Santa A. Multiplex approach in classification, diagnosis, and prognostication in acute myeloid leukemia: an experience from tertiary cancer center in south India. *Indian J Med Paediatric Oncol*. 2017;38(3):266–72.
6. Ghosh S, Shinde SC, Kumaran GS, Sapre RS, Dhond SR, Badrinath Y, Ansari R, Kumar A, Mahadik S, Chougule AB, Nair CN. Haematologic and immunophenotypic profile of acute myeloid leukemia: an experience of Tata Memorial Hospital. *Indian J Cancer*. 2003;40(2):71–6.
7. Yuanfeng L, Yuanhai Z, Yanzhen Z, et al. Application of immunotyping in the differential diagnosis of monocyte associated acute myeloid leukemia (M4/M5). *J Lab Med Clin Med*. 2013;10(19):2520–3.
8. Kumar L, Futschik ME. Mfuzz: a software package for soft clustering of microarray data. *Bioinformatics*. 2007;21(1):5–7.
9. Li Y, Wang Z, You ZH, Li LP, Hu X. Predicting protein-protein interactions via random ferns with evolutionary matrix representation. *Comput Math Methods Med*. 2022;2022:7191684.
10. Ramadan SM, Fouad TM, Summa V, Hasan S, Lo-Coco F. Acute myeloid leukemia developing in patients with autoimmune diseases. *Haematologica*. 2012;97(6):805–17.
11. Boddu PC, Zeidan AM. Myeloid disorders after autoimmune disease. *Best Pract Res Clin Haematol*. 2019;32(1):74–88.
12. Papadaki HA, Marsh JC, Eliopoulos GD. Bone marrow stem cells and stromal cells in autoimmune cytopenias. *Leuk Lymphoma*. 2002;43(4):753–60.
13. Takayanagi H. Osteoimmunology as an intrinsic part of immunology. *Int Immunol*. 2021;33(12):673–8.
14. Buccisano F, Maurillo L, Tamburini A, Del Poeta G, Del Principe MI, Ammatuna E, Consalvo MI, Campagna S, Ottaviani L, Sarlo C, Renzi D, Faccia S, Fraboni D, Lo Coco F, Amadori S, Venditti A. Evaluation of the prognostic relevance of L-selectin and ICAM1 expression in myelodysplastic syndromes. *Eur J Haematol*. 2008;80(2):107–14.
15. Kupsa T, Vasatova M, Karesova I, Zak P, Horacek JM. Baseline serum levels of multiple cytokines and adhesion molecules in patients with acute myeloid leukemia: results of a pivotal trial. *Exp Oncol*. 2014;36(4):252–7.
16. Deak D, Gorcea-Andronic N, Sas V, Teodorescu P, Constantinescu C, Iluta S, Pasca S, Hotea I, Turcas C, Moisoiu V, Zimta AA, Galdean S, Steinheber J, Rus I, Rauch S, Richlitzki C, Munteanu R, Jurj A, Petrushev B, Selicean C, Marian M, Soritau O, Andries A, Roman A, Dima D, Tanase A, Sigurjonsson O, Tomuleasa C. A narrative review of central nervous system involvement in acute leukemias. *Ann Transl Med*. 2021;9(1):68.
17. Martín-Henao GA, Quiroga R, Sureda A, González JR, Moreno V, García J. L-selectin expression is low on CD34+ cells from patients with chronic myeloid leukemia and interferon- α up-regulates this expression. *Haematologica*. 2000;85(2):139–46.
18. Xiao Q, Li X, Li Y, Wu Z, Xu C, Chen Z, He W. Biological drug and drug delivery-mediated immunotherapy. *Acta Pharm Sin B*. 2021;11(4):941–60.
19. Cooles FAH, Tarn J, Lendrem DW, Naamane N, Lin CM, Millar B, Maney NJ, Anderson AE, Thalayasingam N, Diboll J, Bondet V, Duffy D, Barnes MR, Smith GR, Ng S, Watson D, Henkin R, Cope AP, Reynard LN, Pratt AG, Isaacs JD. Interferon- α -mediated therapeutic resistance in early rheumatoid arthritis implicates epigenetic reprogramming. *Ann Rheum Dis*. 2022;81(9):1214–23.
20. Guo G, Li B, Li Q, Li C, Guo D. PTPRC overexpression predicts poor prognosis and correlates with immune cell infiltration in pediatric acute myeloid leukemia. *Clin Lab*. 2022;68(7).
21. Barata JT, Durum SK, Seddon B. Flip the coin: IL-7 and IL-7R in health and disease. *Nat Immunol*. 2019;20(12):1584–93.
22. Mathewson ND, Ashenberg O, Tirosh I, Gritsch S, Perez EM, Marx S, Jerby-Arnon L, Chanoch-Myers R, Hara T, Richman AR, Ito Y, Pyrdol J, Friedrich M, Schumann K, Poitras MJ, Gokhale PC, Gonzalez Castro LN, Shore ME, Hebert CM, Shaw B, Cahill HL, Drummond M, Zhang W, Olawoyin O, Wakimoto H, Rozenblatt-Rosen O, Brastianos PK, Liu XS, Jones PS, Cahill DP, Frosch MP, Louis DN, Freeman GJ, Ligon KL, Marson A, Chiocca EA, Reardon DA, Regev A, Suvà ML, Wucherpfennig KW. Inhibitory CD161 receptor identified in glioma-infiltrating T cells by single-cell analysis. *Cell*. 2021;184(5):1281–98.
23. Li Y, Buijs-Gladdines JG, Canté-Barrett K, Stubbs AP, Vroegindeweij EM, Smits WK, van Marion R, Dinjens WN, Horstmann M, Kuiper RP, Buijsman RC, Zaman GJ, van der Spek PJ, Pieters R, Meijerink JP. IL-7 receptor mutations and steroid resistance in pediatric T cell acute lymphoblastic leukemia: a genome sequencing study. *PLoS Med*. 2016;13(12): e1002200.
24. Follini E, Marchesini M, Roti G. Strategies to overcome resistance mechanisms in T-cell acute lymphoblastic leukemia. *Int J Mol Sci*. 2019;20(12):3021.
25. van der Zwet JCG, Buijs-Gladdines J, Cordo V, Debets DO, Smits WK, Chen Z, Dylus J, Zaman GJR, Altelaar M, Oshima K, Bornhauser B, Bourquin JP, Cools J, Ferrando AA, Vormoor J, Pieters R, Vormoor B, Meijerink JPP. MAPK-ERK is a central pathway in T-cell acute lymphoblastic leukemia that drives steroid resistance. *Leukemia*. 2021;35(12):3394–405.

26. Fu D, Yan J, Zhang Z, Liu Y, Ma X, Ding J, Yang S, Zhao R, Chang A, Gao C, Liu J, Zhao T, Wang X, Huang C, Gao S, Ma Y, Tang B, Feng Y, Wang H, Hao J. Nuclear PLD1 combined with NPM1 induces gemcitabine resistance through tumorigenic IL7R in pancreatic adenocarcinoma. *Cancer Biol Med*. 2023.
27. Tapeh BEG, Alivand MR, Solali S. The role of microRNAs in acute lymphoblastic leukaemia: from biology to applications. *Cell Biochem Funct*. 2020;38(4):334–46.
28. Asl ER, Amini M, Najafi S, Mansoori B, Mokhtarzadeh A, Mohammadi A, Lotfinejad P, Bagheri M, Shirjang S, Lotfi Z, Rasmi Y, Baradaran B. Interplay between MAPK/ERK signaling pathway and MicroRNAs: a crucial mechanism regulating cancer cell metabolism and tumor progression. *Life Sci*. 2021;278: 119499.
29. Yadav VK, Kumar A, Tripathi PP, Gupta J. Long noncoding RNAs in intestinal homeostasis, regeneration, and cancer. *J Cell Physiol*. 2021;236(11):7801–13.
30. Stasevich EM, Murashko MM, Zinevich LS, Demin DE, Schwartz AM. The role of non-coding RNAs in the regulation of the proto-oncogene MYC in different types of cancer. *Biomedicines*. 2021;9(8):921.
31. Fazaeli H, Sheikholeslami A, Ghasemian F, Amini E, Sheykhasan M. The emerging role of LncRNA FENDRR in multiple cancers: a review. *Curr Mol Med*. 2023;23(7):606–29.
32. Koster MJ, Warrington KJ. VEXAS within the spectrum of rheumatologic disease. *Semin Hematol*. 2021;58(4):218–25.
33. Stubbins RJ, Platzbecker U, Karsan A. Inflammation and myeloid malignancy: quenching the flame. *Blood*. 2022;140(10):1067–74.
34. Rossi M, Meggendorfer M, Zampini M, Tettamanti M, Riva E, Travaglino E, Bersanelli M, Mandelli S, Antonella Galbussera A, Mosca E, Saba E, Chierighin C, Manes N, Milanese C, Ubezio M, Morabito L, Peano C, Soldà G, Asselta R, Duga S, et al. Clinical relevance of clonal hematopoiesis in persons aged ≥ 80 years. *Blood*. 2021;138(21):2093–105.
35. Lindsley RC, Mar BG, Mazzola E, Grauman PV, Shareef S, Allen SL, Pigneux A, Wetzler M, Stuart RK, Erba HP, Damon LE, Powell BL, Lindeman N, Steensma DP, Wadleigh M, DeAngelo DJ, Neuberg D, Stone RM, Ebert BL. Acute myeloid leukemia ontogeny is defined by distinct somatic mutations. *Blood*. 2015;125(9):1367–76.
36. Brunetti L, Gundry MC, Goodell MA. DNMT3A in Leukemia. *Cold Spring Harb Perspect Med*. 2017;7(2): a030320.
37. Desai P, Mencia-Trinchant N, Savenkov O, Simon MS, Cheang G, Lee S, Samuel M, Ritchie EK, Guzman ML, Ballman KV, Roboz GJ, Hassane DC. Somatic mutations precede acute myeloid leukemia years before diagnosis. *Nat Med*. 2018;24(7):1015–23.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.