Clinical and prognostic relevance of *CXCL12* expression in acute myeloid leukemia

Shi-sen Wang [Equal first author, 1] , Zi-jun Xu [Equal first author, 2, 3] , Ye Jin [1, 3] , Ji-chun Ma [2, 3] , Pei-hui Xia [2, 3] , Xiangmei Wen [2, 3] , Zhen-wei Mao [Corresp., 2] , Jiang Lin [2, 3] , Jun Qian [Corresp. 1, 3]

[1] Department of Hematology, Affiliated People's Hospital of Jiangsu University, Zhenjiang, Jiangsu, China

[2] Laboratory Center, Affiliated People's Hospital of Jiangsu University, Zhenjiang, Jiangsu, China

[3] Zhenjiang Clinical Research Center of Hematology, Zhenjiang, Jiangsu, China

Corresponding Authors: Zhen-wei Mao, Jun Qian
Email address: maopen365@163.com, qianjun@jskfhn.org.cn

Background. Accumulating studies have been made to understand the association between *CXC* chemokine ligand-12 (*CXCL12*)/*CXC* chemokine receptor 4 (*CXCR4*) and acute myeloid leukemia (AML). However, large-scale data analysis of potential relationship between *CXCL12* and AML remains insufficient. **Methods.** We collected abundant *CXCL12* expression data and AML samples from several publicly available datasets. CIBERSORT algorithm was used to quantify immune cell fractions and the online website of STRING was utilized for gene ontology (GO) enrichment and Kyoto encyclopedia of genes and genomes (KEGG) analysis. The statistical analysis and graphical work were mainly performed via the R software. **Results.** *CXCL12* expression was extremely down-regulated in AML. Clinically, low *CXCL12* expression was correlated with higher white blood cells (WBCs) ($P < 0.0001$), more blasts in bone marrow (BM) ($P < 0.001$) and peripheral blood (PB) ($P < 0.0001$), *FLT3*-internal tandem duplications (*FLT3*-ITD) ($P = 0.010$) and *NPM1* mutations ($P = 0.015$). More importantly, reduced *CXCL12* expression predicted worse overall survival (OS) and event-free survival (EFS) in all AML, non-M3-AML, and cytogenetically normal (CN)-AML patients in three independent cohorts. As for immune cell infiltration, high *CXCL12* expressed groups tended to harbor more memory B cells and plasma cells infiltration while low *CXCL12* expressed groups exhibited more eosinophils infiltration. GO enrichment and KEGG pathways analysis revealed the potential biological progress the gene participating in. **Conclusions.** *CXCL12* is significantly down-regulated in AML and low *CXCL12* expression is an independent and poor predictor of AML prognosis. *CXCL12* expression level correlates with clinical and immune characteristics of AML, which could provide potential assistance for treatment. Prospective studies are needed to further validate the impact of *CXCL12* expression before routine clinical application in AML.

Clinical and prognostic relevance of *CXCL12*

expression in acute myeloid leukemia

Shi-sen Wang[1][†], Zi-jun Xu[2,3][†], Ye Jin[1,3], Ji-chun Ma[2,3], Pei-hui Xia[2,3], Xiang-mei Wen[2,3], Zhen-wei Mao[2], Jiang Lin[2,3], Jun Qian[1,3]

[1]Department of Hematology, Affiliated People's Hospital of Jiangsu University, Zhenjiang, Jiangsu, P.R. China

[2]Laboratory Center, Affiliated People's Hospital of Jiangsu University, Zhenjiang, Jiangsu, P.R. China

[3]Zhenjiang Clinical Research Center of Hematology, Zhenjiang, Jiangsu, P.R. China

[†]Shi-sen Wang and Zi-jun Xu should be considered joint first author

Corresponding Author:

Jun Qian[1,3]

8 Dianli Rd., Zhenjiang, Jiangsu, 212002, P.R. China

Email address: qianjun@jskfhn.org.cn

Zhen-wei Mao[2]

8 Dianli Rd., Zhenjiang, Jiangsu, 212002, P.R. China

Email address: maopen365@163.com

35

36

37

Abstract

39 **Background.** Accumulating studies have been made to understand the association between *CXC*

40 chemokine ligand-12 (*CXCL12*)/*CXC* chemokine receptor 4 (*CXCR4*) and acute myeloid

41 leukemia (AML). However, large-scale data analysis of potential relationship between *CXCL12*

42 and AML remains insufficient.

43 **Methods.** We collected abundant *CXCL12* expression data and AML samples from several

44 publicly available datasets. CIBERSORT algorithm was used to quantify immune cell fractions

45 and the online website of STRING was utilized for gene ontology (GO) enrichment and Kyoto

46 encyclopedia of genes and genomes (KEGG) analysis. The statistical analysis and graphical

47 work were mainly performed via the R software.

48 **Results.** *CXCL12* expression was extremely down-regulated in AML. Clinically, low *CXCL12*

49 expression was correlated with higher white blood cells (WBCs) ($P < 0.0001$), more blasts in

50 bone marrow (BM) ($P < 0.001$) and peripheral blood (PB) ($P < 0.0001$), *FLT3*-internal

51 tandem duplications (*FLT3*-ITD) ($P = 0.010$) and *NPM1* mutations ($P = 0.015$). More

52 importantly, reduced *CXCL12* expression predicted worse overall survival (OS) and event-free

53 survival (EFS) in all AML, non-M3-AML, and cytogenetically normal (CN)-AML patients in

54 three independent cohorts. As for immune cell infiltration, high *CXCL12* expressed groups

55 tended to harbor more memory B cells and plasma cells infiltration while low *CXCL12* expressed

56 groups exhibited more eosinophils infiltration. GO enrichment and KEGG pathways analysis

57 revealed the potential biological progress the gene participating in.

58 **Conclusions.** *CXCL12* is significantly down-regulated in AML and low *CXCL12* expression is

59 an independent and poor predictor of AML prognosis. *CXCL12* expression level correlates with

60 clinical and immune characteristics of AML, which could provide potential assistance for

61 treatment. Prospective studies are needed to further validate the impact of *CXCL12* expression

62 before routine clinical application in AML.

63

Introduction

65 Acute myeloid leukemia (AML) represents a heterogeneous disease characterizing by complex

66 molecular signatures, complicated pathogeny and aggressive progression, as well as a poor

67 clinical outcome (Döhner et al. 2015). Bone marrow (BM) microenvironment contributes

68 significantly to leukemogenesis and leukemia progression (Ayala et al. 2009; Meads et al. 2008).

69   Most adult patients are still terribly suffering from AML, even though tremendous advancements
70   such as novel chemotherapy drugs and hematopoietic stem cell transplantation (HSCT) have
71   been achieved recently (Döhner et al. 2017; Döhner et al. 2015). Until now, chromosome
72   aberration analysis remains the main method for diagnosis, classification, prognosis, and guiding
73   clinical therapy while $40-50\%$ of patients are lack representative chromosome aberrations
74   (Mrózek et al. 2004). So more specific and accurate biomarkers are necessary to be discovered
75   for AML. In recent years, immunophenotype and cytochemistry analysis are found valuable in
76   the diagnosis of AML, and immunophenotyping surface type molecules like CD123, CD45,
77   CD34, CD38 have been reported to help confirm the diagnosis of AML (Prada-Arismendy et al.
78   2017). Additionally, it is now evident that oncogene mutations affect the progression of AML
79   and some of these mutated genes are *FLT3*, *NPM1*, and *CEBPA* (Grove & Vassiliou 2014;
80   Meyer & Levine 2014). Unfortunately, despite extensive researches that have been carried out to
81   find prognostic biomarkers, AML is still a tough disease with a variable prognosis and poor
82   survival rate, with 5-year overall survival less than 50% and even lower in the elderly (Kell
83   2004). Considering all the above, further studies are essential for a better understanding of how
84   various factors contribute to the disease progression.
85   Chemokines and their receptors mediate crucial biological functions of tumor progression
86   including leukocyte recruitment and function, cellular senescence, tumor cell proliferation and
87   survival, and invasion and metastasis (Mantovani et al. 2010). *CXC* chemokine ligand-12
88   (*CXCL12*), also known as stromal cell-derived factor-1 (*SDF-1*), is a *CXC* chemokine that
89   belongs to the large family of chemotactic cytokines. The protein, functioning through its
90   receptor *CXC* chemokine receptor 4 (*CXCR4*), plays an important role in hematopoietic cell
91   development, immune system foundation, and retention of hematopoietic stem and progenitor
92   cells in the BM microenvironment (Janssens et al. 2018). *CXCL12* expression has been found to
93   be related to at least 23 different types of tumors (Balkwill 2004) and the gene's critical
94   influences on solid tumors have been illustrated in prostate cancer (Taichman et al. 2002), non-
95   small cell lung cancer (Phillips et al. 2003), renal cell carcinoma (Pan et al. 2006) and so on. As
96   for hematological malignancies, *CXCL12* is reported to be overexpressed in chronic lymphocytic
97   leukemia (CLL) (Möhle et al. 1999), as well as in childhood acute lymphoblastic leukemia
98   (Crazzolara et al. 2001). Also, Spoo AC et al. have proved that elevated *CXCR4* expression is
99   relevant to inferior prognosis of AML in adult patients (Konoplev et al. 2007; Rombouts et al.
100   2004; Spoo et al. 2007). All these findings may contribute to a better understanding of *CXCL12*'s
101   potential mechanism and clinical relevance in both solid tumors and blood cancers.
102   Despite enormous progress having been made to understand the relationship between *CXCR4*
103   and AML, studies concentrating on *CXCL12* and AML remain insufficient, and large-scale
104   analysis of the potential association between the gene and AML hasn't been performed before.

105 Herein, we adopted similar methods of our previous research (Xu et al. 2019) and made

106 comprehensive analyses of the gene's possible interaction with AML using public datasets,

107 trying to reveal the implication of *CXCL12* expression in AML and providing potential treatment

108 strategies for the clinic.

109

Materials & Methods

111 **Patients and Database**

112 We used six publicly available datasets from Gene Expression Omnibus (GEO,

113 http://www.ncbi.nlm.nih.gov/geo/) and The Cancer Genome Atlas (TCGA,

114 https://cancergenome.nih.gov/) in this study. Of these datasets, three of them consist of

115 expression data for bulk primary AML samples (the TCGA dataset, GSE6891, and GSE10358),

116 two involve expression data for healthy and AML bone marrow samples (GSE30029 and

117 GSE63270) and one includes expression data for major hematopoietic lineages during the

118 differentiation (GSE42519), respectively. Expression data for FAB AML subtypes and patient

119 characteristics were acquired from TCGA sequencing data. Normal mRNA expression data from

120 Genotype-Tissue Expression (GTEx) database and pan-cancer cell line sequencing data from

121 Broad Institute Cancer Cell Line Encyclopedia (CCLE) database were extracted through their

122 portal websites for analysis.

123 **Estimation of immune cell fractions**

124 CIBERSORT algorithm (https://cibersort.stanford.edu/index.php) was used to quantify immune

125 cell fractions according to the expression pattern of *CXCL12*. More detailed information for this

126 method is available in Newman AM et al.'s introduction (Newman et al. 2015). Immune cells

127 calculated in this study contains B cells (including B cells memory and B cells naive), T cells

128 (including CD8+ T cells, CD4+ naive T cells, CD4+ memory resting T cells, CD4+ memory

129 activated T cells, follicular helper T cells, regulatory T cells and T cells gamma delta),

130 macrophage (including M0 macrophage, M1 macrophage, and M2 macrophage), NK cells

131 (including NK cells resting and NK cells activated), dendritic cells (including dendritic cells

132 resting and dendritic cells activated), mast cells (including mast cells resting and mast cells

133 activated), plasma cells, monocyte, eosinophils, and neutrophils. Multiple immune infiltration

134 deconvolution methods including TIMER, quanTIseq, xCell, MCP-counter, and EPIC were

135 achieved from their websites.

136 **Statistical analysis and bioinformatics**

137 Low and high *CXCL12* expressers were discriminated against based on the median expression

138 level of the gene. Analyses of patient characteristics and survival were described previously (Xu

139 et al. 2019). Differential expression was analyzed using 'limma' or 'edgeR' R package.

140   The Pearson Correlation coefficient (denoted as Cor) was calculated to assess whether there was
141   a correlation between cell infiltration and *CXCL12* expression. Moreover, 'ggplot2' and 'ggpubr'
142   R packages were applied for visualizing the results of data analysis. The single-cell correlation
143   analysis of *CXCL12* was conducted using single-cell data sequenced by van Galen, P. et al (van
144   Galen et al. 2019), and the published scRNA-seq dataset was downloaded from
145   GEO: GSE116256. And gene ontology (GO) enrichment and Kyoto encyclopedia of genes and
146   genomes (KEGG) analysis were conducted using the online website of STRING (http://string-
147   db.org). One-way ANOVA was used to test for differences among at least three groups and the
148   Wilcoxon test was used to determine differences in each two-group comparison. All statistical
149   analyses and figure plotting were performed using the R software version 3.5.3 (https://www.r-
150   project.org/). All differences were considered statistically significant at the level of $P < 0.05$.
151

Results

A systematic overview of differential genes and cytokines in AML

154   Initially, we systematically analyzed a total of 261 cytokine and cytokine receptor genes
155   (Supplementary Data 1) in AML patients and normal controls utilizing GSE30029, and all genes
156   were screened in Supplementary Data 2 with the difference considered significant if adjusted *p*-
157   value < 0.05 and $|logFC| > 0$. As obviously displayed in Fig. 1A, the *CXCL12* gene was one
158   of the most significantly different genes between AML patients and normal controls. Then, we
159   further verified the differences in GSE63270 (Fig. 1B, Supplementary Data 3). The top 50 genes
160   with significant differences were extracted from two analytic results respectively and then
161   intersected. The common 11 genes included *CXCL12, TNFRSF10B, IL1RAP, IL12RB1, IL3RA,*
162   *CSF2RA, EPOR, TNFSF14, IL18BP, TNFRSF4,* and *IL2RG*. Furthermore, considering the
163   essential functions of *CXCL12* in AML, we ultimately decided to study *CXCL12* in detail.

CXCL12 expression is different among normal tissues and various carcinomas

165   To explore how *CXCL12* is expressed across various tissues in healthy people, we analyzed the
166   mRNA expression status utilizing the GTEx database (Fig. 2A, Supplementary Data 4).
167   Apparently, *CXCL12* expression was extremely low in blood and bone marrow while it was
168   much higher among other normal tissues. It was also worth mentioning that the brain, stomach,
169   and testis showed a relatively low level of *CXCL12* expression compared with other tissues.
170   More importantly, when analyzing different cancer cell lines (Supplementary Data 5), we found
171   *CXCL12* was lowly expressed among most hematopoietic neoplasms, especially in AML and
172   chronic myelocytic leukemia (CML). Interestingly, cervix cancer cells showed exclusively lower
173   *CXCL12* expression levels than AML and CML, as shown in Fig. 2B.

CXCL12 expression is markedly reduced in AML and single cells

175 Considering our purpose, we studied the *CXCL12* expression difference between healthy people

176 and AML patients. A significant decrease of *CXCL12* expression was identified in AML cells

177 (GSE30029, n $=$ 90) in comparison to normal cells (GSE30029, n $=$ 31) (P $=$ 3.2e-16, Fig.

178 3A). We detected a similar result in the GSE63270 cohort (n_{AML} $=$ 62, n_{normal} $=$ 42, P $=$

179 7.4e-06), as indicated in Fig. 3B. Additionally, we analyzed *CXCL12* expression difference in

180 eight FAB AML subtypes from M0 to M7 (Fig. S1) and *CXCL12* transcription during

181 hematopoiesis differentiation (Fig. S2), although the difference was of no statistical significance.

182   For further exploration, we delineated *CXCL12* expression patterns in several hematological

183 cells including hematopoietic stem cells (HSCs), multipotent progenitor cells (MPPs), granulo-

184 monocyte progenitor cells (GMPs), common myeloid progenitor cells (CMPs), megakaryocyte-

185 erythroid progenitor cells (MEPs), monocytes, and erythrocytes using the single-cell data

186 sequenced by van Galen, P. et al (van Galen et al. 2019), as shown in Fig. S3A. Interestingly,

187 unlike the former result, MPPs and MEPs showed exclusively high expression of *CXCL12*. Also,

188 we studied *CXCL12* and *CXCR4* expression status in the single-cell sequenced data and found

189 *CXCR4* expression was abundant in single cells from all patients while *CXCL12* was rarely

190 expressed (Fig. S3B).

191 **Clinical features and molecular insights of *CXCL12***

192 The patient characteristics with respect to *CXCL12* expression from TCGA were summarized

193 (Table 1). Low *CXCL12* expressed patients tended to have higher white blood cell (WBC) counts

194 (P $<$ 0.0001) and more percentages of blasts in BM (P $<$ 0.001) and peripheral blood (PB)

195 (P $<$ 0.0001). With regard to frequently mutated genes in AML, low *CXCL12* expressers were

196 more likely to have *FLT3*-internal tandem duplications (*FLT3*-ITD) (P $=$ 0.010) and *NPM1*

197 mutations (P $=$ 0.015) and less likely to have mutated *TP53* (P $=$ 0.046) than those with high

198 *CXCL12* expression.

199 **Decreased *CXCL12* expression predicts adverse prognosis**

200 Survival analyses were further performed using three independent cohorts (TCGA, GSE6891,

201 and GSE10358), and all cohorts were divided base on optional cut-off values determined by the

202 x-tile method. In the TCGA whole-AML cohort, *CXCL12*^{low} patients showed significantly

203 decreased overall survival (OS) (P $=$ 0.013, Fig. 4A) and event-free survival (EFS) (P $=$

204 0.0001, Fig. 4D). The impact of *CXCL12* expression on outcome was also observed in the non-

205 M3-AML (OS, P $=$ 0.002, Fig. 4B; EFS, P $=$ 0.0007, Fig. 4E) and cytogenetically normal

206 (CN)-AML (OS, P $=$ 0.002, Fig. 4C; EFS, P $=$ 0.002, Fig. 4F) subgroups. This finding was

207 further validated using GSE6891 (Fig. S4) and GSE10358 (Fig. S5) and the results were

208 identical to the former.

209   The potential relationship between low *CXCL12* expression and the existence of *FLT3*-ITD

210 and *NPM1* mutations led us to further explore the additional prognostic value of the gene in more

211   subdivided groups. As expected, low *CXCL12* expression patients showed significantly inferior

212   survival probability in the *FLT3*-ITD absent (Fig. 5A, $P = 0.003$) and *NPM1* mutated (Fig. 5B,

213   $P = 0.068$, a trend for significance actually) patients, which could provide more subdivision

214   proofs for AML based on the dichotomous stratification of *CXCL12* expression.

215   **Multivariate analysis of *CXCL12* expression for OS in the TCGA cohorts**

216   To evaluate whether there were potential interrelationships between *CXCL12* expression and

217   known clinical factors, we first conducted a univariate analysis in the TCGA dataset and then

218   covered variables with a univariable $P \leqslant 0.20$ in the multivariate Cox proportional hazard

219   model. The multivariate analysis result of *CXCL12* expression for OS was displayed in Table 2.

220   Low *CXCL12* expression maintained a high hazard rate for whole-AML patients (HR $= 2.25$,

221   95% CI $1.51-3.35$, $P < 0.0001$), after adjusting age ($P < 0.0001$), WBC count ($P <$

222   0.0001), cytogenetic risk ($P = 0.002$) and mutation status of *TP53* ($P = 0.004$), *RUNX1* (P

223   $= 0.023$) and *DNMT3A* ($P = 0.012$). The predictive significance also applied to CN-AML

224   patients (HR $= 2.14$, 95% CI $1.15-3.97$, $P = 0.016$), as indicated.

225   **Profile of relationship between *CXCL12* expression and immune cells infiltration**

226   The landscape of the relationship between *CXCL12* expression and immune cell infiltration was

227   systematically evaluated by the CIBERSORT algorithm (Fig. 6). We found that B cells memory,

228   plasma cells, T cells regulatory (Tregs), T cells gamma delta, NK cells activated, mast cells

229   activated, and eosinophils were significantly changed between high *CXCL12* and low *CXCL12*

230   groups, while B cell naive, T cells CD8, T cells CD4 naive, T cells CD4 memory resting, T cells

231   CD4 memory activated, T cell follicular helper, NK cells resting, monocyte, macrophages M0,

232   macrophages M1, macrophages M2, dendritic cells resting, dendritic cells activated, mast cells

233   resting, and neutrophils were not obviously altered between groups. More specifically, high

234   *CXCL12* expressed groups tended to harbor more memory B cells and plasma cells infiltration

235   while low *CXCL12* expressed groups exhibited more eosinophils infiltration. Less importantly, T

236   cells regulatory (Tregs), T cells gamma delta, and NK cells were rarely infiltrated in both high

237   and low *CXCL12* expressed groups.

238   To verify the signatures of immune infiltration, we adopted multiple deconvolution methods

239   including TIMER, quanTIseq, xCell, MCP-counter, and EPIC to cross-check the analytic results

240   of CIBERSORT. Then, we made a Pearson correlation analysis between all infiltration results

241   and *CXCL12* expression, and totally 16 significantly correlated results were displayed in Fig. S6.

242   Plasma cells (xCell, Cor $= 0.273$, $P < 0.001$) and B cells (quanTIseq, Cor $= 0.246$, $P <$

243   0.001; EPIC, Cor $= 0.198$, $P = 0.007$; MCP-counter, Cor $= 0.145$, $P = 0.050$) had a

244   positive correlation with *CXCL12* expression.

245   **GO analysis and KEGG pathways analysis of *CXCL12***

246   GO analyses of *CXCL12* for biological process (BP), cellular component (CC) and molecular
247   function (MF) were performed on the online website of STRING (http://string-db.org) (Fig. 7),
248   using genes differentially expressed between high and low *CXCL12* expressers (|log FC| $>$ 1.5,
249   adjusted *p*-value $<$ 0.05, Supplementary Data 6). GO term annotation showed that *CXCL12*
250   was correlated with anatomical structure morphogenesis, extracellular structure organization,
251   animal organ morphogenesis, biological adhesion, system development, multicellular organismal
252   process, animal organ development, cell adhesion, extracellular matrix organization, and
253   multicellular organism development (BP); extracellular region, extracellular space, extracellular
254   matrix, intrinsic component of plasma membrane, integral component of plasma membrane, cell
255   periphery, plasma membrane, collagen-containing extracellular matrix, basement membrane, and
256   collagen trimer (CC); binding, growth factor binding, signaling receptor binding, calcium ion
257   binding, cell adhesion molecule binding, glycosaminoglycan binding, chemokine activity,
258   heparin binding, receptor ligand activity, and receptor regulator activity (MF) (Fig. 7,
259   Supplementary Data 7). GO analysis showed significantly high enrichment in anatomical
260   structure morphogenesis for biological process, extracellular region for cellular component, and
261   binding for molecular function. Moreover, *CXCL12* was involved directly or indirectly in the
262   protein digestion and absorption pathway according to the KEGG pathways analysis (Fig. 7,
263   Supplementary Data 7).

264

265   Discussion

266   Our study showed *CXCL12* expression was extremely low in blood and bone marrow compared
267   with other normal tissues and also downgraded in hematological neoplasms among pan-
268   carcinomas. For our purpose, we then validated the expression differences between AML and
269   their normal contrasts in two independent cohorts and found a consistent result. Herein, we
270   concluded that *CXCL12* was notably down-regulated in AML. Hitherto, there hasn't been
271   enough researches to investigate *CXCL12* expression in AML, although its receptor *CXCR4* has
272   been demonstrated to be overexpressed in AML (Rombouts et al. 2004). To our knowledge,
273   *CXCL12* plays a prominent role in the interaction between AML cells and the BM
274   microenvironment, and the *CXCL12*/*CXCR4*-axis is importantly involved in the migration,
275   infiltration, and overproliferation of AML cells (Ladikou et al. 2020).
276   The clinical implications of *CXCL12* were further investigated. *CXCL12*[low] patients tended to
277   have higher WBC counts and more percentages of blasts, implying more active proliferation of
278   *CXCL12*-reduced AML cells. Moreover, *CXCL12*[low] patients were more likely to harbor *FLT*-
279   ITD and *NPM1* mutations and less often to have *TP53* mutation.

280   Concerning prognostic relevance, we identified and validated low *CXCL12* expression as a
281   prognostic indicator for inferior OS and EFS in all AML, non-M3-AML, and CN-AML patients.
282   In fact, high *CXCR4* expression level has been associated with inferior prognosis of AML
283   (Burger & Bürkle 2007; Spoo et al. 2007) and *CXCR4* antagonists such as AMD3100 have been
284   proven to enhance chemosensitization of AML (Nervi et al. 2009). However, it was interesting to
285   discover the predictive value of *CXCL12* for AML prognosis exactly opposite to *CXCR4*, which
286   was beyond our expectation. Previously, Petit I et al. demonstrated that up-regulation of *CXCR4*
287   could serve to increase the sensitivity of cells to lower *CXCL12* signals (Petit et al. 2002), and
288   Mandawat A et al. found that exposure to *CXCL12* caused the lowering of the surface expression
289   of *CXCR4* on AML cells (Mandawat et al. 2010), which might provide an explanation for this
290   phenomenon. Moreover, we pointed out that *FLT3*-ITD absent patients and *NPM1* mutated
291   patients could be further subdivided into high-risk groups (with *CXCL12* down-regulation) and
292   low-risk groups (with *CXCL12* up-regulation) respectively based on the expression level of
293   *CXCL12*.
294   GO analysis helped us understand the concrete functions of *CXCL12* and the result showed
295   that the gene dynamically participated in anatomical structure morphogenesis for biological
296   process, extracellular region for cellular component, and binding for molecular function. It was
297   also reasonable to extrapolate protein digestion and absorption pathway was curbed molecularly
298   in some way during the development of AML. It has been widely evidenced that *CXCL12* plays
299   an essential role in stem cell anchorage to the BM microenvironment and reduction of *CXCL12*
300   concentrations within the BM may interfere with retention and facilitate the egress of cells (Petit
301   et al. 2002), which could promote the progress of AML. As previously demonstrated, *CXCL12*
302   promoted glycolytic reprogramming in AML cells (Braun et al. 2016) whereas the malignant
303   cells could benefit from aerobic glycometabolism, which endowed them with better capacity for
304   cell proliferation, immune evasion, and chemotherapeutics resistance (Vander Heiden et al.
305   2009). Low *CXCL12* expression restrained glycolytic reprogramming and indirectly promoted
306   aerobic glycometabolism might become one of the possible mechanisms for low *CXCL12*
307   expression inducing poor AML prognosis. When analyzing the relationship between *CXCL12*
308   and immune cell infiltration, we found high *CXCL12* expressed groups tended to harbor more
309   memory B cells and plasma cells infiltration while low *CXCL12* expressed groups exhibited
310   more eosinophils infiltration. However, the pathological role of *CXCL12* in AML is still largely
311   unknown and needs to be elucidated in future studies.
312   Although we had been committing ourselves to make the research as comprehensive and
313   credible as possible, there were still several drawbacks that couldn't be ignored. Firstly, the
314   collections were entirely acquired from the public database due to the quantity of clinical data
315   was too limited to conduct such a large-scale analysis. We attempted to mend it by analyzing

316 multi-group independent datasets to guarantee our results convincing. And secondly, we did not
317 carry out subsequent experiments to authenticate our results, so more laboratory proofs such as
318 protein level verification are required to be offered for our research.
319

Conclusions

321 In summary, we carry out a large-scale data analysis of the potential relationship between
322 *CXCL12* and AML for the first time. *CXCL12* is significantly down-regulated in AML and low
323 *CXCL12* expression is an independent and poor predictor of clinical outcome of AML. The gene
324 could provide additional prognostic value in *FLT3*-ITD absent or *NPM1* mutated AML patients.
325 Furthermore, our research replenishes the understanding of the immune and clinical association
326 between *CXCL12* and AML, which could help clinicians to explore new targeted drugs and
327 immunotherapy for AML. Finally, further experiments are warranted to support our findings.
328

Acknowledgements

330 None.

331

References

333 Ayala F, Dewar R, Kieran M, and Kalluri R. 2009. Contribution of bone microenvironment to
334   leukemogenesis and leukemia progression. *Leukemia* 23:2233-2241.
335   10.1038/leu.2009.175
336 Balkwill F. 2004. The significance of cancer cell expression of the chemokine receptor CXCR4.
337   *Seminars in Cancer Biology* 14:171-179.
338 Braun M, Qorraj M, Büttner M, Klein FA, Saul D, Aigner M, Huber W, Mackensen A, Jitschin R,
339   and Mougiakakos D. 2016. CXCL12 promotes glycolytic reprogramming in acute myeloid
340   leukemia cells via the CXCR4/mTOR axis. *Leukemia* 30:1788-1792. 10.1038/leu.2016.58
341 Burger JA, and Bürkle A. 2007. The CXCR4 chemokine receptor in acute and chronic leukaemia:
342   a marrow homing receptor and potential therapeutic target. *British Journal of Haematology*
343   137:288-296.
344 Crazzolara R, Kreczy A, Mann G, Heitger A, Eibl G, Fink FM, Möhle R, and Meister B. 2001. High
345   expression of the chemokine receptor CXCR4 predicts extramedullary organ infiltration in
346   childhood acute lymphoblastic leukaemia. *British Journal of Haematology* 115:545-553.
347 Döhner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Büchner T, Dombret H, Ebert BL,
348   Fenaux P, Larson RA, Levine RL, Lo-Coco F, Naoe T, Niederwieser D, Ossenkoppele GJ,
349   Sanz M, Sierra J, Tallman MS, Tien H-F, Wei AH, Löwenberg B, and Bloomfield CD. 2017.
350   Diagnosis and management of AML in adults: 2017 ELN recommendations from an
351   international expert panel. *Blood* 129:424-447. 10.1182/blood-2016-08-733196

352   Döhner H, Weisdorf DJ, and Bloomfield CD. 2015. Acute Myeloid Leukemia. *The New England*
353   *journal of medicine* 373:1136-1152. 10.1056/NEJMra1406184
354   Grove CS, and Vassiliou GS. 2014. Acute myeloid leukaemia: a paradigm for the clonal evolution
355   of cancer? *Disease Models & Mechanisms* 7:941-951. 10.1242/dmm.015974
356   Janssens R, Struyf S, and Proost P. 2018. Pathological roles of the homeostatic chemokine
357   CXCL12. *Cytokine and Growth Factor Reviews* 44:51-68. 10.1016/j.cytogfr.2018.10.004
358   Kell J. 2004. Emerging treatments in acute myeloid leukaemia. *Expert opinion on emerging drugs*
359   9:55-71.
360   Konoplev S, Rassidakis GZ, Estey E, Kantarjian H, Liakou CI, Huang X, Xiao L, Andreeff M,
361   Konopleva M, and Medeiros LJ. 2007. Overexpression of CXCR4 predicts adverse overall
362   and event-free survival in patients with unmutated FLT3 acute myeloid leukemia with
363   normal karyotype. *Cancer* 109:1152-1156.
364   Ladikou EE, Chevassut T, Pepper CJ, and Pepper AG. 2020. Dissecting the role of the
365   CXCL12/CXCR4 axis in acute myeloid leukaemia. *British Journal of Haematology*
366   189:815-825. 10.1111/bjh.16456
367   Mandawat A, Fiskus W, Buckley KM, Robbins K, Rao R, Balusu R, Navenot J-M, Wang Z-X,
368   Ustun C, Chong DG, Atadja P, Fujii N, Peiper SC, and Bhalla K. 2010. Pan-histone
369   deacetylase inhibitor panobinostat depletes CXCR4 levels and signaling and exerts
370   synergistic antimyeloid activity in combination with CXCR4 antagonists. *Blood* 116:5306-
371   5315. 10.1182/blood-2010-05-284414
372   Mantovani A, Savino B, Locati M, Zammataro L, Allavena P, and Bonecchi R. 2010. The
373   chemokine system in cancer biology and therapy. *Cytokine and Growth Factor Reviews*
374   21:27-39. 10.1016/j.cytogfr.2009.11.007
375   Meads MB, Hazlehurst LA, and Dalton WS. 2008. The bone marrow microenvironment as a tumor
376   sanctuary and contributor to drug resistance. *Clinical cancer research : an official journal*
377   *of the American Association for Cancer Research* 14:2519-2526. 10.1158/1078-
378   0432.CCR-07-2223
379   Meyer SC, and Levine RL. 2014. Translational implications of somatic genomics in acute myeloid
380   leukaemia. *The Lancet Oncology* 15:e382-e394. 10.1016/S1470-2045(14)70008-7
381   Möhle R, Failenschmid C, Bautz F, and Kanz L. 1999. Overexpression of the chemokine receptor
382   CXCR4 in B cell chronic lymphocytic leukemia is associated with increased functional
383   response to stromal cell-derived factor-1 (SDF-1). *Leukemia* 13:1954-1959.
384   Mrózek K, Heerema NA, and Bloomfield CD. 2004. Cytogenetics in acute leukemia. *Blood*
385   *Reviews* 18:115-136.
386   Nervi B, Ramirez P, Rettig MP, Uy GL, Holt MS, Ritchey JK, Prior JL, Piwnica-Worms D, Bridger
387   G, Ley TJ, and DiPersio JF. 2009. Chemosensitization of acute myeloid leukemia (AML)
388   following mobilization by the CXCR4 antagonist AMD3100. *Blood* 113:6206-6214.
389   10.1182/blood-2008-06-162123
390   Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, and Alizadeh
391   AA. 2015. Robust enumeration of cell subsets from tissue expression profiles. *Nature*
392   *methods* 12:453-457. 10.1038/nmeth.3337

393 Pan J, Mestas J, Burdick MD, Phillips RJ, Thomas GV, Reckamp K, Belperio JA, and Strieter RM.
394   2006. Stromal derived factor-1 (SDF-1/CXCL12) and CXCR4 in renal cell carcinoma
395   metastasis. *Molecular Cancer* 5:56.
396 Petit I, Szyper-Kravitz M, Nagler A, Lahav M, Peled A, Habler L, Ponomaryov T, Taichman RS,
397   Arenzana-Seisdedos F, Fujii N, Sandbank J, Zipori D, and Lapidot T. 2002. G-CSF
398   induces stem cell mobilization by decreasing bone marrow SDF-1 and up-regulating
399   CXCR4. *Nature Immunology* 3:687-694.
400 Phillips RJ, Burdick MD, Lutz M, Belperio JA, Keane MP, and Strieter RM. 2003. The stromal
401   derived factor-1/CXCL12-CXC chemokine receptor 4 biological axis in non-small cell lung
402   cancer metastases. *American Journal of Respiratory and Critical Care Medicine*
403   167:1676-1686.
404 Prada-Arismendy J, Arroyave JC, and Röthlisberger S. 2017. Molecular biomarkers in acute
405   myeloid leukemia. *Blood Reviews* 31:63-76. 10.1016/j.blre.2016.08.005
406 Rombouts EJC, Pavic B, Löwenberg B, and Ploemacher RE. 2004. Relation between CXCR-4
407   expression, Flt3 mutations, and unfavorable prognosis of adult acute myeloid leukemia.
408   *Blood* 104:550-557.
409 Spoo AC, Lübbert M, Wierda WG, and Burger JA. 2007. CXCR4 is a prognostic marker in acute
410   myelogenous leukemia. *Blood* 109:786-791.
411 Taichman RS, Cooper C, Keller ET, Pienta KJ, Taichman NS, and McCauley LK. 2002. Use of
412   the stromal cell-derived factor-1/CXCR4 pathway in prostate cancer metastasis to bone.
413   *Cancer Research* 62:1832-1837.
414 van Galen P, Hovestadt V, Wadsworth Ii MH, Hughes TK, Griffin GK, Battaglia S, Verga JA,
415   Stephansky J, Pastika TJ, Lombardi Story J, Pinkus GS, Pozdnyakova O, Galinsky I,
416   Stone RM, Graubert TA, Shalek AK, Aster JC, Lane AA, and Bernstein BE. 2019. Single-
417   Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity.
418   *Cell* 176. 10.1016/j.cell.2019.01.031
419 Vander Heiden MG, Cantley LC, and Thompson CB. 2009. Understanding the Warburg effect:
420   the metabolic requirements of cell proliferation. *Science (New York, NY)* 324:1029-1033.
421   10.1126/science.1160809
422 Xu Z-J, Ma J-C, Zhou J-D, Wen X-M, Yao D-M, Zhang W, Ji R-B, Wu D-H, Tang L-J, Deng Z-Q,
423   Qian J, and Lin J. 2019. Reduced protocadherin17 expression in leukemia stem cells: the
424   clinical and biological effect in acute myeloid leukemia. *Journal of Translational Medicine*
425   17:102. 10.1186/s12967-019-1851-1
426
427

Figure 1

Expression heat map of differentially expressed cytokine and cytokine receptor genes between AML patients and normal controls.

(A) Differentially expressed genes between AML patients and normal controls in GSE30029.

(B) Differentially expressed genes between AML patients and normal controls in GSE63270.

Only the top 50 genes were shown if adjusted *p*-value < 0.05 and |logFC| > 0.

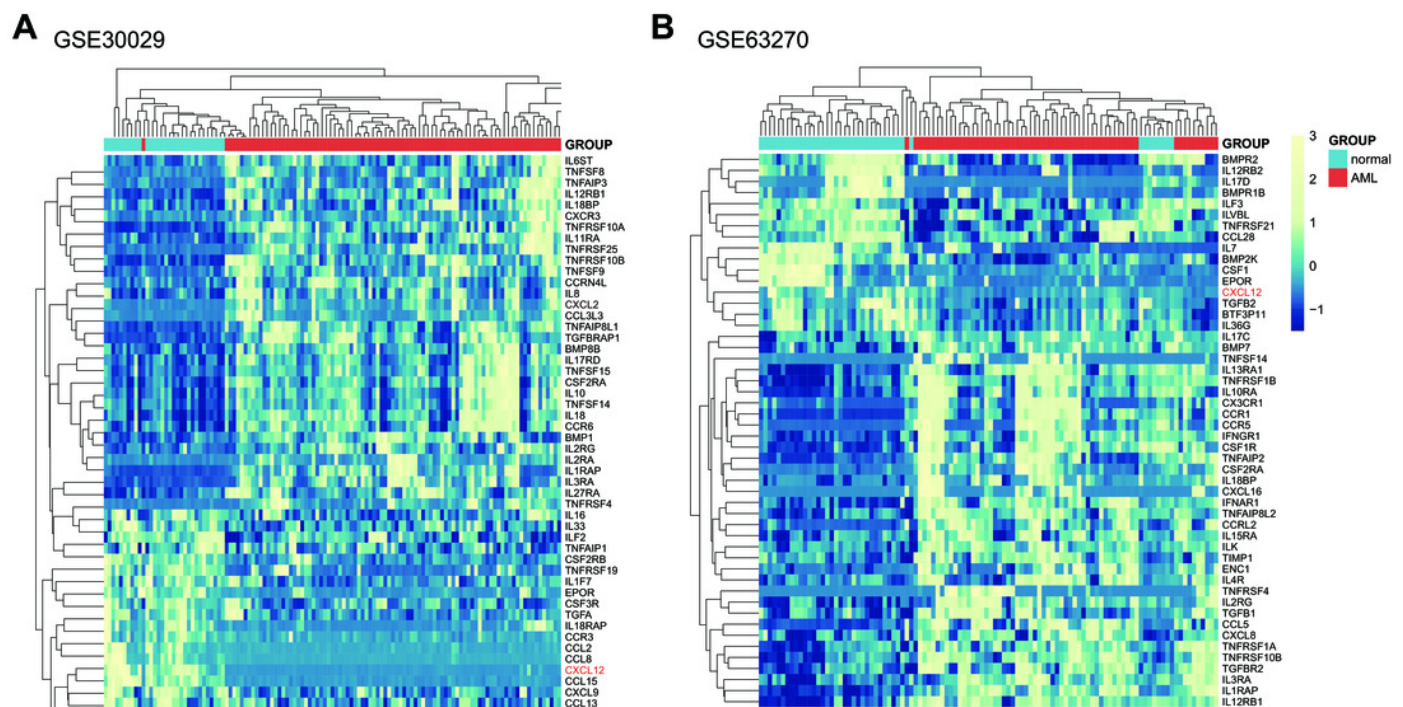



Figure 2

mRNA expression levels of *CXCL12* from different tissue origins and tumors.

(A) mRNA expression levels of *CXCL12* in different tissues from the GTEx database. (B) mRNA expression levels of *CXCL12* in various tumor cell lines from the CCLE database.

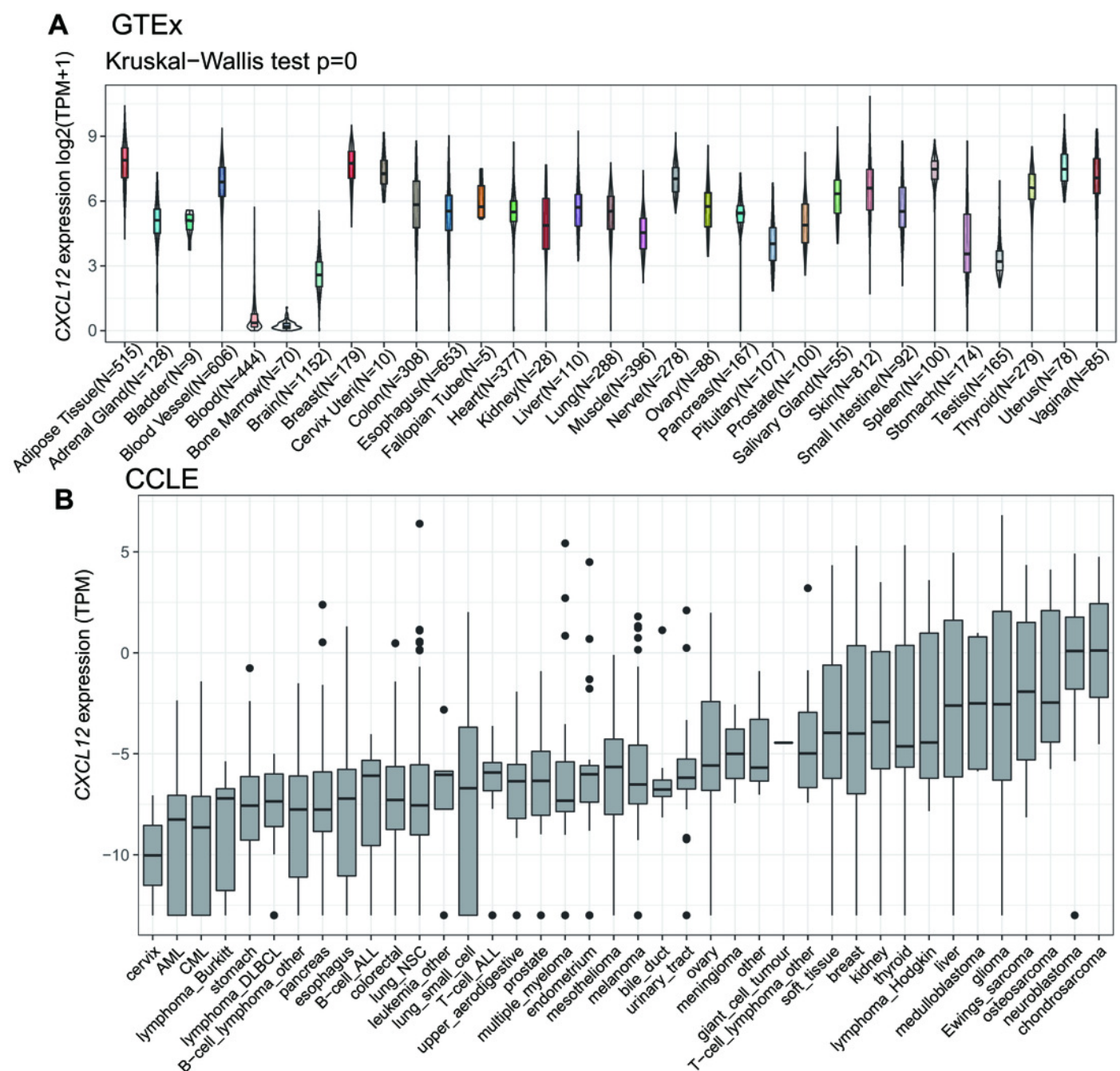



Figure 3

CXCL12 expression in normal hematopoiesis and AML cells.

The *CXCL12* expression level of normal controls and AML patients from GSE30029 (A) and GSE63270 (B) datasets.

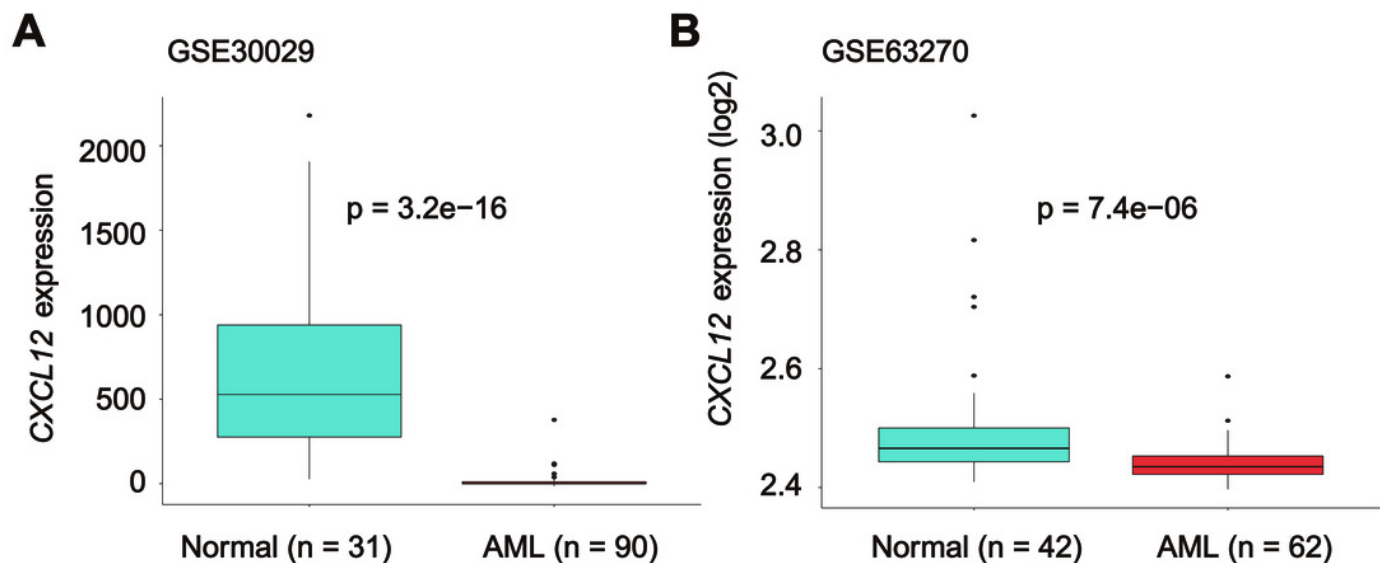



Figure 4

Survival analysis of *CXCL12* using the TCGA cohort.

(A, D) The correlation of *CXCL12* expression level with OS (A) and EFS (D) for the total AML patients. (B, E) The correlation of *CXCL12* expression level with OS (B) and EFS (E) for the non-M3-AML patients. (C, F) The correlation of *CXCL12* expression level with OS (C) and EFS (F) for the CN-AML patients.

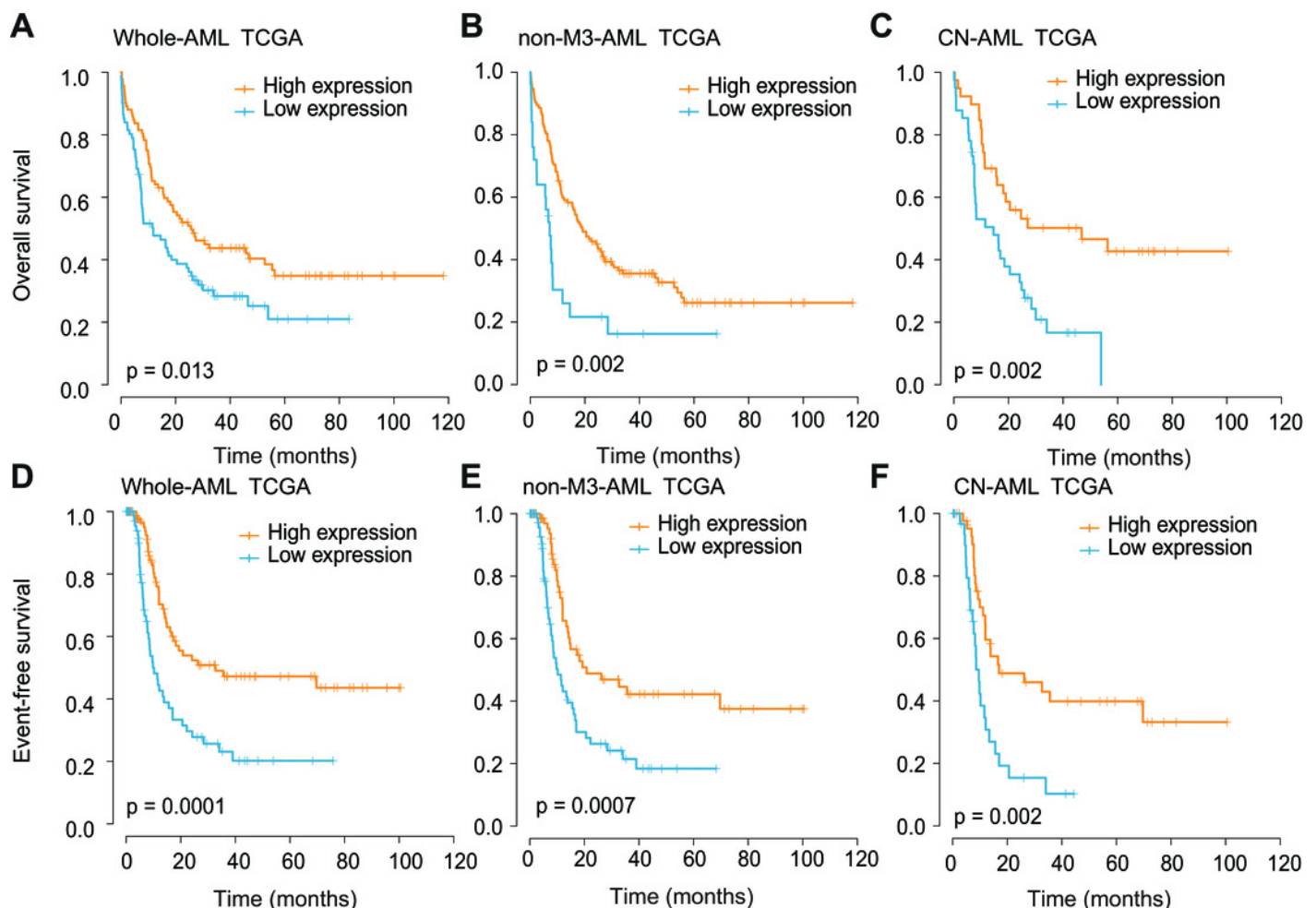



Figure 5

Prognostic significance of *CXCL12* in the molecularly defined subgroups in the TCGA cohort.

(A) Kaplan-Meier estimate for OS of *FLT3*-ITD-negative AML patients. (B) Kaplan-Meier estimate for OS of *NPM1*-mutated AML patients.

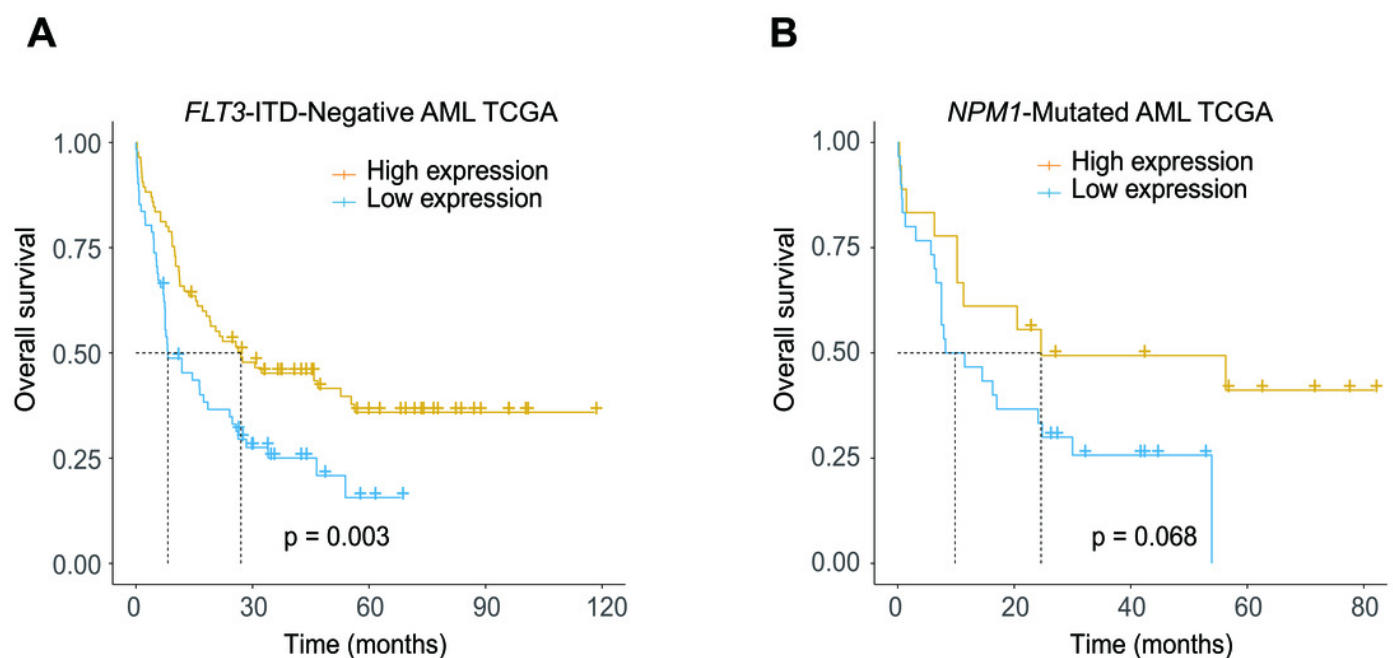



Figure 6

Immune cell infiltration difference between high and low *CXCL12* expressers.

The Violin plot exhibits the difference between 22 CIBERSORT immune cell fractions between high and low *CXCL12* expressers.

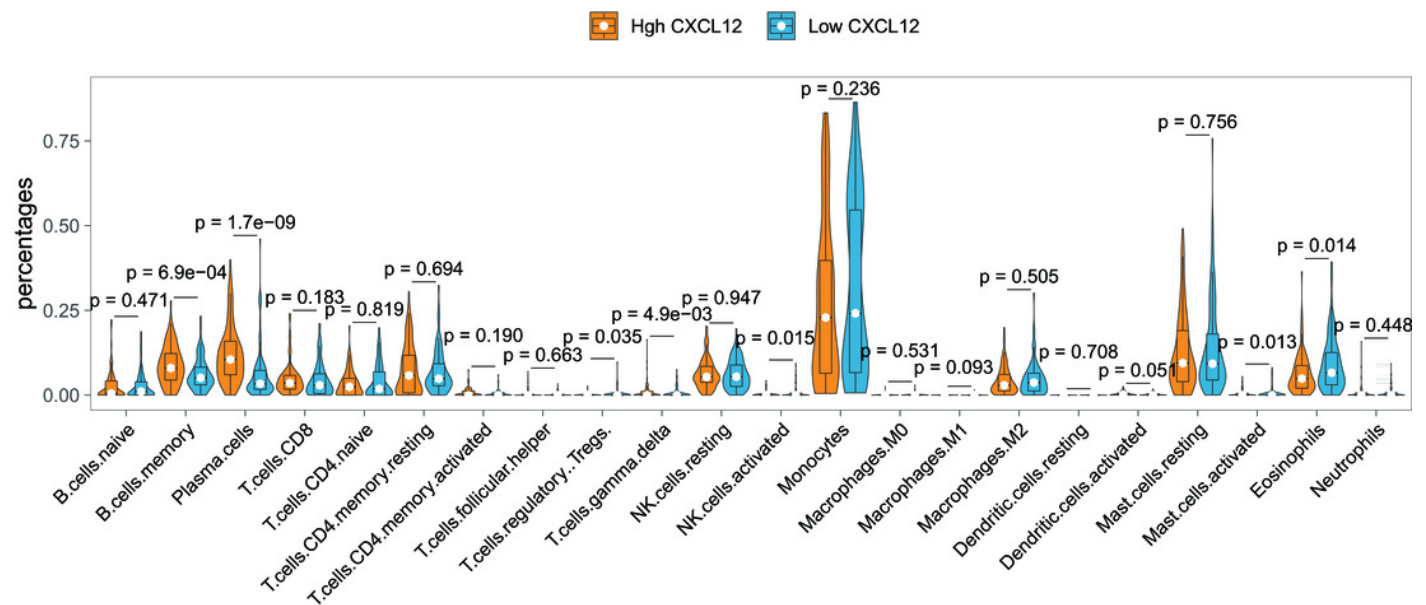
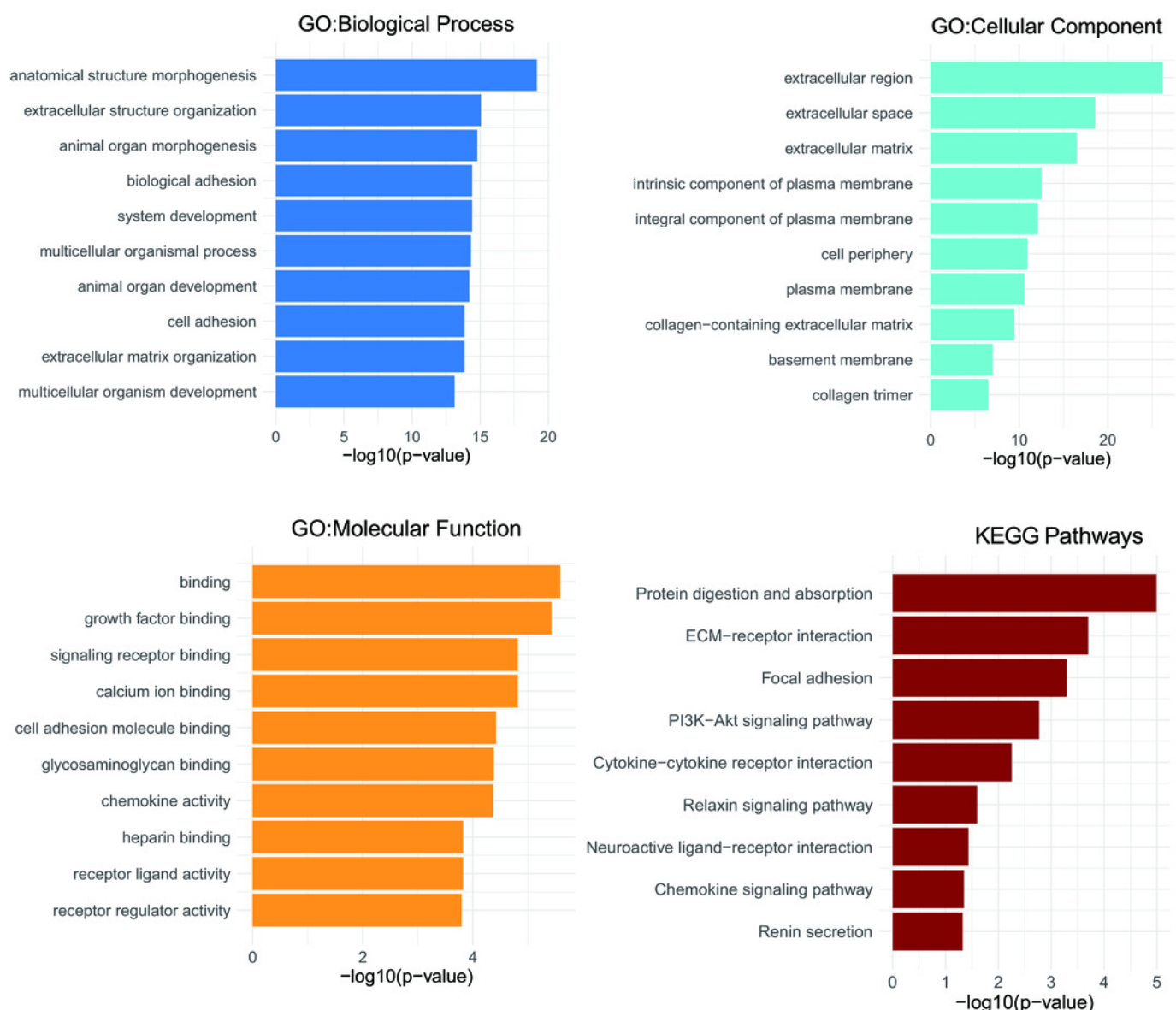



Figure 7

GO and KEGG pathways analysis of differentially expressed genes between high and low *CXCL12* expressers.

The top-ten statistically significant results identified including Biological Process (BP), Cellular Component (CC), Molecular Function (MF), and KEGG Pathways are listed according to their -log10 (*p*-value) (colored bars).

Patient characteristics with respect to *CXCL12* expression.

Abbreviations: AML, acute myeloid leukemia; TCGA, The Cancer Genome Atlas; WBC, white blood cells; BM, bone marrow; PB, peripheral blood; *ITD*, internal tandem duplication.

Patient characteristics	AML (TCGA dataset)		
	High *CXCL12* (n = 86)	Low *CXCL12* (n = 87)	*P*
Age, years			0.537
Median	60	57	
Range	21-88	18-81	
Sex (male/female)	52/34	40/47	0.056
WBC count, ×10^9/L			<0.0001
Median	6.4	42.1	
Range	0.4-116.2	0.9-297.4	
BM blasts, %			<0.001
Median	62	77	
Range	30-100	32-100	
PB blasts, %			<0.0001
Median	14	52	
Range	0-97	0-98	
Karyotype			0.872
Favorable	15	17	
Intermediate	50	51	
Adverse	20	17	
Unknown	1	2	
FLT3-ITD			0.010
Present	7	19	
Absent	79	66	
NPM1			0.015
Mutated	17	31	
Wild-type	69	54	
CEBPA			0.318
Single mutated	3	5	
Double mutated	4	1	
Wild-type	79	79	
IDH1			0.980
Mutated	8	8	
Wild-type	78	77	
IDH2			0.818
Mutated	9	8	
Wild-type	77	77	
RUNX1			0.062
Mutated	11	4	
Wild-type	75	81	

DNMT3A			0.965
Mutated	21	21	
Wild-type	65	64	

Table continued

Patient characteristics	AML (TCGA dataset)		
	High *CXCL12* **(n = 86)**	**Low *CXCL12*** **(n = 87)**	***P***
TP53			0.046
Mutated	10	3	
Wild-type	76	82	

Multivariate analysis of *CXCL12* expression for overall survival in the TCGA cohorts.

Abbreviations: TCGA, The Cancer Genome Atlas; CN-AML, cytogenetically normal AML; WBC, white blood cells; CI, confidence interval.

NOTE: Hazard Ratio > 1 or Hazard Ratio < 1 indicate a higher or lower risk. Only variables with a univariable *P* ≤ 0.20 were included in the multivariable models.

[a] Low vs high expression.

[b] > 60 vs ≤ 60 years.

[c] ≥ 30 vs < 30 $\times 10^9$/L.

[d] Adverse vs intermediate vs favorable.

[e] Mutated vs wild type.

Variables	Whole-AML (n = 173)		CN-AML (n = 80)	
	Hazard Ratio (95% CI)	*P*	Hazard Ratio (95% CI)	*P*
CXCL12[a]	2.25 (1.51-3.35)	< 0.0001	2.14 (1.15-3.97)	0.016
Age[b]	1.01 (1.00-1.03)	< 0.0001	2.37 (1.35-4.18)	0.003
WBC count[c]	3.14 (1.92-5.14)	< 0.0001	1.21 (0.67-2.19)	0.520
Cytogenetic risk[d]	1.79 (1.25-2.57)	0.002	-	-
TP53[e]	2.85 (1.41-5.75)	0.004	-	-
RUNX1[e]	2.08 (1.10-3.93)	0.023	-	-
DNMT3A[e]	1.74 (1.13-2.68)	0.012	1.95 (1.11-3.44)	0.020
IDH1 [e]	-	-	0.60 (0.21-1.75)	0.351