



OPEN

## Characterization of somatic mutation-associated microenvironment signatures in acute myeloid leukemia patients based on TCGA analysis

Jun Wang, Feng-Ting Dao, Lu Yang & Ya-Zhen Qin

Recurrent genetic mutations occur in acute myeloid leukemia (AML) and have been incorporated into risk stratification to predict the prognoses of AML patients. The bone marrow microenvironment plays a critical role in the development and progression of AML. However, the characteristics of the genetic mutation-associated microenvironment have not been comprehensively identified to date. In this study, we obtained the gene expression profiles of 173 AML patients from The Cancer Genome Atlas (TCGA) database and calculated their immune and stromal scores by applying the ESTIMATE algorithm. Immune scores were significantly associated with OS and cytogenetic risk. Next, we categorized the intermediate and poor cytogenetic risk patients into individual-mutation and wild-type groups according to RUNX1, ASXL1, TP53, FLT3-ITD, NPM1 and biallelic CEBPA mutation status. The relationships between the immune microenvironment and each genetic mutation were investigated by identifying differentially expressed genes (DEGs) and conducting functional enrichment analyses of them. Significant immune- and stromal-relevant DEGs associated with each mutation were identified, and most of the DEGs (from the FLT3-ITD, NPM1 and biallelic CEBPA mutation groups) were validated in the GSE14468 cohort downloaded from the Gene Expression Omnibus (GEO) database. In summary, we identified key immune- and stromal-relevant gene signatures associated with genetic mutations in AML, which may provide new biomarkers for risk stratification and personalized immunotherapy.

Acute myeloid leukemia (AML) is a hematological malignancy characterized by the clonal expansion of myeloid blasts, resulting in impaired hematopoiesis and bone marrow failure<sup>1,2</sup>. The outcomes of AML patients are highly heterogeneous, and analysis of cytogenetic abnormalities is the backbone of risk stratification in AML<sup>3</sup>. In recent years, several somatic mutations, such as FLT3 and NPM1, were shown to be strongly prognostic and have been incorporated into risk categories of AML in both NCCN guidelines and ELN recommendations<sup>2,3</sup>.

Tumor microenvironment (TME) infrastructure, which comprises a variety of immune and stromal cell types (e.g., endothelial cells and fibroblasts) and extracellular components they secrete (e.g., cytokines, growth factors, hormones, and extracellular matrix), represents a chronic inflammatory, immunosuppressive, and proangiogenic intratumoral environment<sup>4-7</sup>. The TME not only plays an important role during tumor initiation, progression, and metastasis but also has profound implications for therapeutic efficacy and specificity<sup>8-13</sup>. AML myeloid blasts are able to adapt and grow in bone marrow environments with a significantly lower likelihood of detection and eradication by host immunosurveillance compared with other environments. Recent evidence has highlighted the importance of the bone marrow microenvironment in protecting leukemic stem cells (LSCs) from chemotherapy-induced cell death<sup>14</sup>. Therefore, efforts to characterize the TME signatures have drawn considerable attention in the field of solid tumors, as well as leukemia.

Estimation of 'Stromal and Immune cells in Malignant Tumors using Expression data' (ESTIMATE) is a method that uses gene expression signatures to infer the fraction of stromal and immune cells in tumor samples<sup>15</sup>. This algorithm has been employed to investigate the microenvironment of several solid tumors, such as gastric

Peking University People's Hospital, Peking University Institute of Hematology, National Clinical Research Center for Hematologic Disease, No. 11 Xizhimen South Street, Xicheng District, Beijing 100044, China. ✉email: qin2000@aliyun.com

cancer<sup>16</sup>, breast cancer<sup>17</sup> and glioblastoma<sup>18</sup>, and it has also been applied to estimate immune and stromal scores in AML patients<sup>19–21</sup>. Since gene mutations are important prognostic factors for AML, whether they individually have unique microenvironment features has not been determined to date.

In the current study, by downloading gene expression profiles for AML cohorts from The Cancer Genome Atlas (TCGA) database and analyzing the immune/stromal scores of patients based on the ESTIMATE algorithm, we characterized the gene mutation-associated microenvironment. Moreover, the identified immune- and stromal-relevant DEGs associated with some mutations were verified using the Gene Expression Omnibus (GEO) database.

## Results

**OS and impact of immune and stromal scores in AML patients.** The complete gene expression profiles and clinical information of 173 AML patients were retrieved from the TCGA database for this study. The median follow-up period was 304 days (range, 28–2861 days) for the entire cohort, and the 2-year OS rate was 44.4% (95% confidence interval (CI), 35.6–52.8%). According to the ESTIMATE algorithm, immune scores ranged from 1329.53 to 3971.97, whereas stromal scores varied from -1888.81 to 435.75. Then, patients were divided into high- and low-score groups according to the median immune and stromal scores, respectively. Patients with high immune scores had significantly lower 2-year OS rates than did those with low immune scores (32.7% [95% CI 21.9–43.8%] vs 58.1% [95% CI 44.6–69.5%],  $P=0.026$ , log-rank test, Fig. 1a). However, patients in the high stromal score group had similar 2-year OS rates to those in the low stromal score group (41.4% [95% CI 29.9–52.4%] vs 48.6% [95% CI 35.1–60.9%],  $P=0.58$ , log-rank test, Fig. 1b).

**Relationship between immune/stromal scores and cytogenetic risk.** Of the 173 patients, 32 (18.5%) were in the favorable cytogenetic risk group, 101 (58.4%) were in the intermediate cytogenetic risk group, 37 (21.4%) were in the poor cytogenetic risk group, and the remaining 3 (1.7%) belonged to the unknown group. As shown in Fig. 1c, patients in the intermediate cytogenetic risk group had a similar 2-year OS rate as those in the poor cytogenetic risk group (39.7% [95% CI 28.8–50.4%] vs 24.6% [95% CI 9.7–43.0%],  $P=0.16$ , log-rank test), and they both had significantly lower 2-year OS rates compared with the favorable cytogenetic risk group (82.0% [95% CI 61.7–92.2%],  $P=0.0009$  and  $0.0002$ , log-rank test).

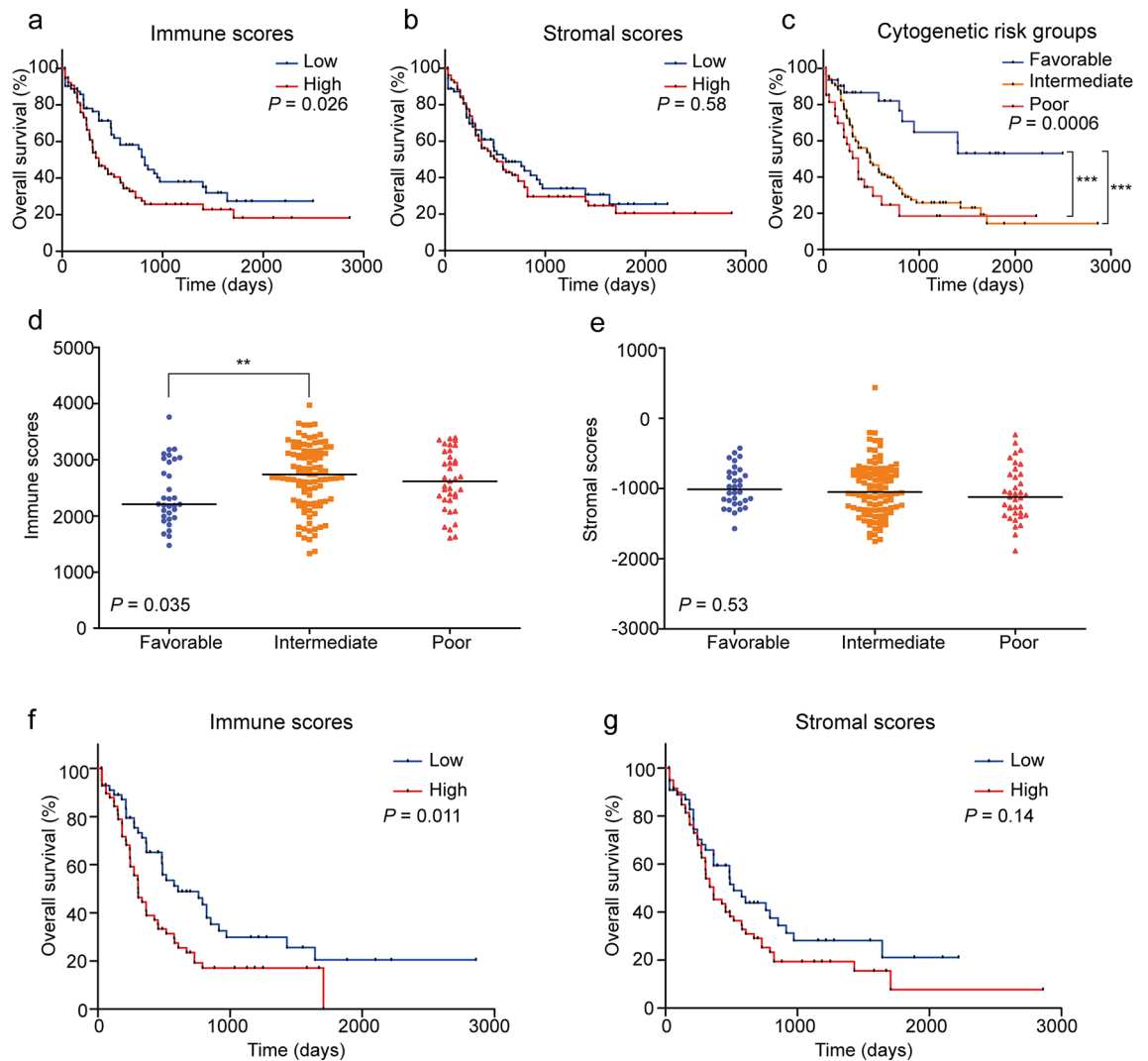
The median immune scores in the favorable, intermediate and poor cytogenetic risk groups were 2209.75, 2738.40 and 2616.32, respectively ( $P=0.035$ , ANOVA test). As shown in Fig. 1d, the immune score of the favorable cytogenetic risk group was significantly lower than those of the intermediate and poor cytogenetic risk groups ( $P=0.0084$  and  $0.067$ , Mann–Whitney test, two-sided), and the immune score of the intermediate risk group was similar to that of the poor risk group ( $P=0.44$ , Mann–Whitney test, two-sided). The median stromal scores in the favorable, intermediate and poor risk groups ( $-1011.98$ ,  $-1050.87$  and  $-1122.03$ ) were similar ( $P=0.53$ , ANOVA test, Fig. 1e).

Overall, patients in the favorable cytogenetic risk group had significantly lower immune scores and a higher 2-year OS rate. However, there were no significant differences in the 2-year OS rate, immune scores or stromal scores between the intermediate and poor cytogenetic risk groups. Therefore, the intermediate and poor cytogenetic risk groups ( $n=137$ ) were grouped together for the subsequent analysis. After grouping AML patients with intermediate and poor cytogenetic risk by the median score, those with high immune scores still had significantly lower 2-year OS rates than those with the low immune scores (23.5% [95% CI 13.2–35.5%] vs 48.8% [95% CI 34.0–62.1%],  $P=0.011$ , log-rank test, Fig. 1f). Furthermore, patients in the high stromal score group tended to have lower 2-year OS rates compared with patients in the low stromal score group (29.1% [95% CI 18.0–41.1%] vs 43.8% [95% CI 28.9–57.8%],  $P=0.14$ , log-rank test, Fig. 1g).

**Somatic mutation-associated immune/stromal scores in the intermediate and poor cytogenetic risk groups.** To further explore the association between immune/stromal scores and mutations, patients in the intermediate and poor cytogenetic risk groups were divided into two subgroups based on whether the individual somatic mutation existed, and their immune and stromal scores are presented in Fig. 2g,h.

**Poor prognostic mutations (RUNX1, TP53, ASXL1 and FLT3-ITD).** As shown in Fig. 2a–d, patients with the individual gene mutation had or tended to have lower 2-year OS rates than WT patients (RUNX1: 14.3% [95% CI 2.3–36.6%] vs 39.2% [95% CI 28.9–49.2%],  $P=0.11$ , Log-rank test; TP53: 0% [95% CI 0–0%] vs 39.7% [95% CI 29.6–49.6%],  $P=0.0001$ , Log-rank test; ASXL1: 33.3% [95% CI 0.9–77.4%] vs 35.5% [95% CI 26.1–45.0%],  $P=0.25$ , Log-rank test; FLT3-ITD: 30.2% [95% CI 10.5–52.9%] vs 36.6% [95% CI 26.4–46.8%],  $P=0.41$ , Log-rank test), respectively.

Patients in the RUNX1 and TP53 mutation groups had or tended to have higher immune scores and stromal scores than did those in the corresponding WT groups (RUNX1: immune score 2947.2 vs 2669.7,  $P=0.15$ , Mann–Whitney test, two-sided; stromal score  $-794.34$  vs  $-1098.615$ ,  $P=0.0008$ , Mann–Whitney test, two-sided. TP53: immune score 2993.49 vs 2670.13,  $P=0.35$ , Mann–Whitney test, two-sided; stromal score  $-923.14$  vs  $-1073.74$ ,  $P=0.089$ , Mann–Whitney test, two-sided). ASXL1 mutation patients had similar immune scores compared with WT patients (2642.07 vs 2685.88,  $P=0.91$ , Mann–Whitney test, two-sided), but they tended to have higher stromal scores than WT patients ( $-806.38$  vs  $-1072.71$ ,  $P=0.14$ , Mann–Whitney test, two-sided). However, both the immune scores and stromal scores of patients with FLT3-ITD mutation were lower or tended to be lower than those of WT patients (immune score 2565.07 vs 2762.73,  $P=0.061$ , Mann–Whitney test, two-sided; stromal score  $-1255.38$  vs  $-1010.38$ ,  $P=0.022$ , Mann–Whitney test, two-sided) (Fig. 2g,h).

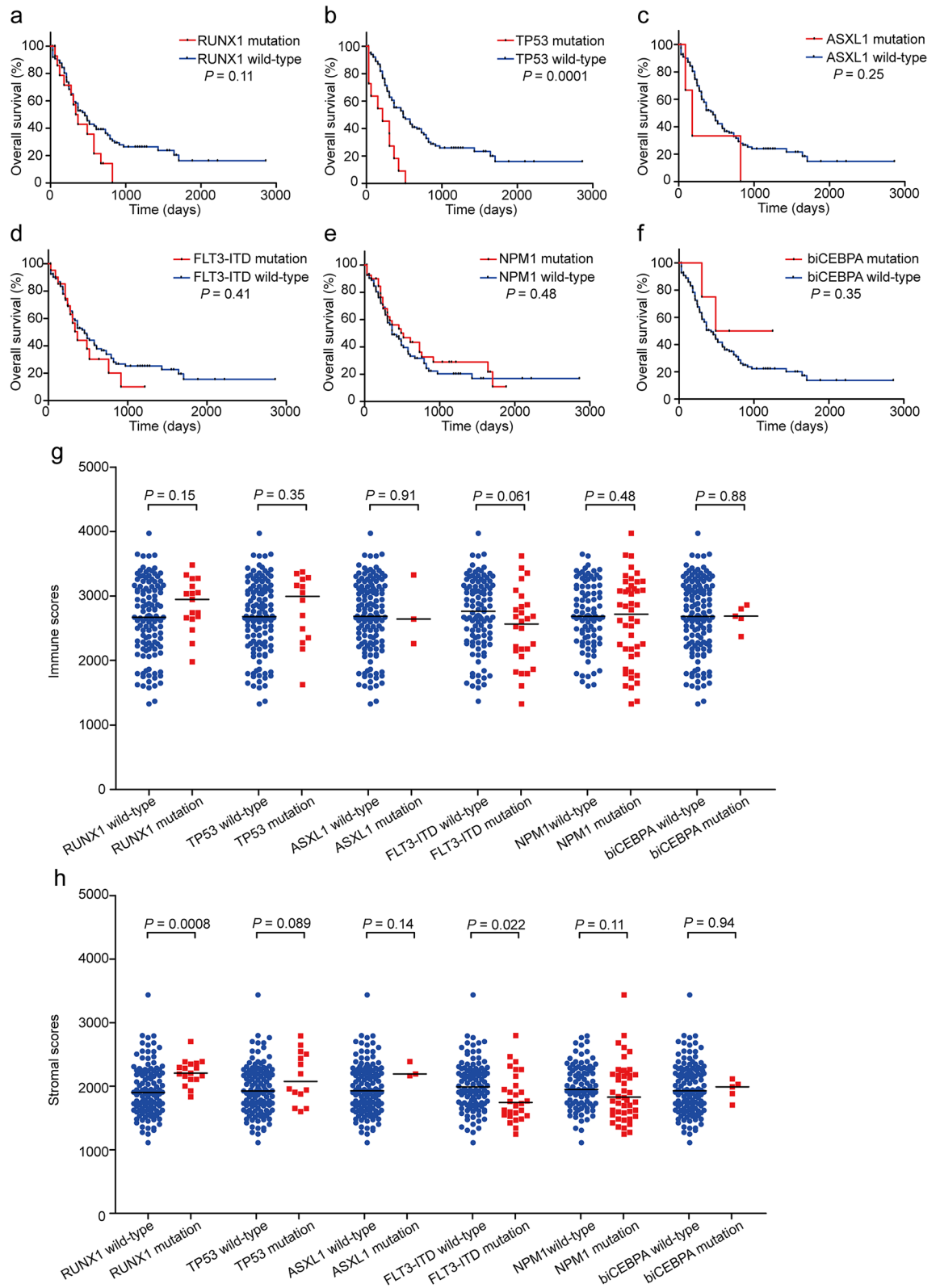


**Figure 1.** Immune scores and stromal scores are associated with AML OS and cytogenetic risk. **(a)** Kaplan-Meier survival analysis of high versus low immune score groups (log-rank test,  $P=0.026$ ). **(b)** Kaplan-Meier survival analysis of high versus low stromal score groups (log-rank test,  $P=0.58$ ). **(c)** Kaplan-Meier survival analysis of cytogenetic risk groups (log-rank test,  $P=0.0006$ ). **(d)** Distribution of immune scores within cytogenetic risk groups ( $P=0.035$ ). **(e)** Distribution of stromal scores within cytogenetic risk groups ( $P=0.53$ ). **(f)** Kaplan-Meier survival analysis of high versus low immune score groups in the intermediate and poor cytogenetic risk patients (log-rank test,  $P=0.011$ ). **(g)** Kaplan-Meier survival analysis of high versus low stromal score groups in the intermediate and poor cytogenetic risk patients (log-rank test,  $P=0.14$ ).

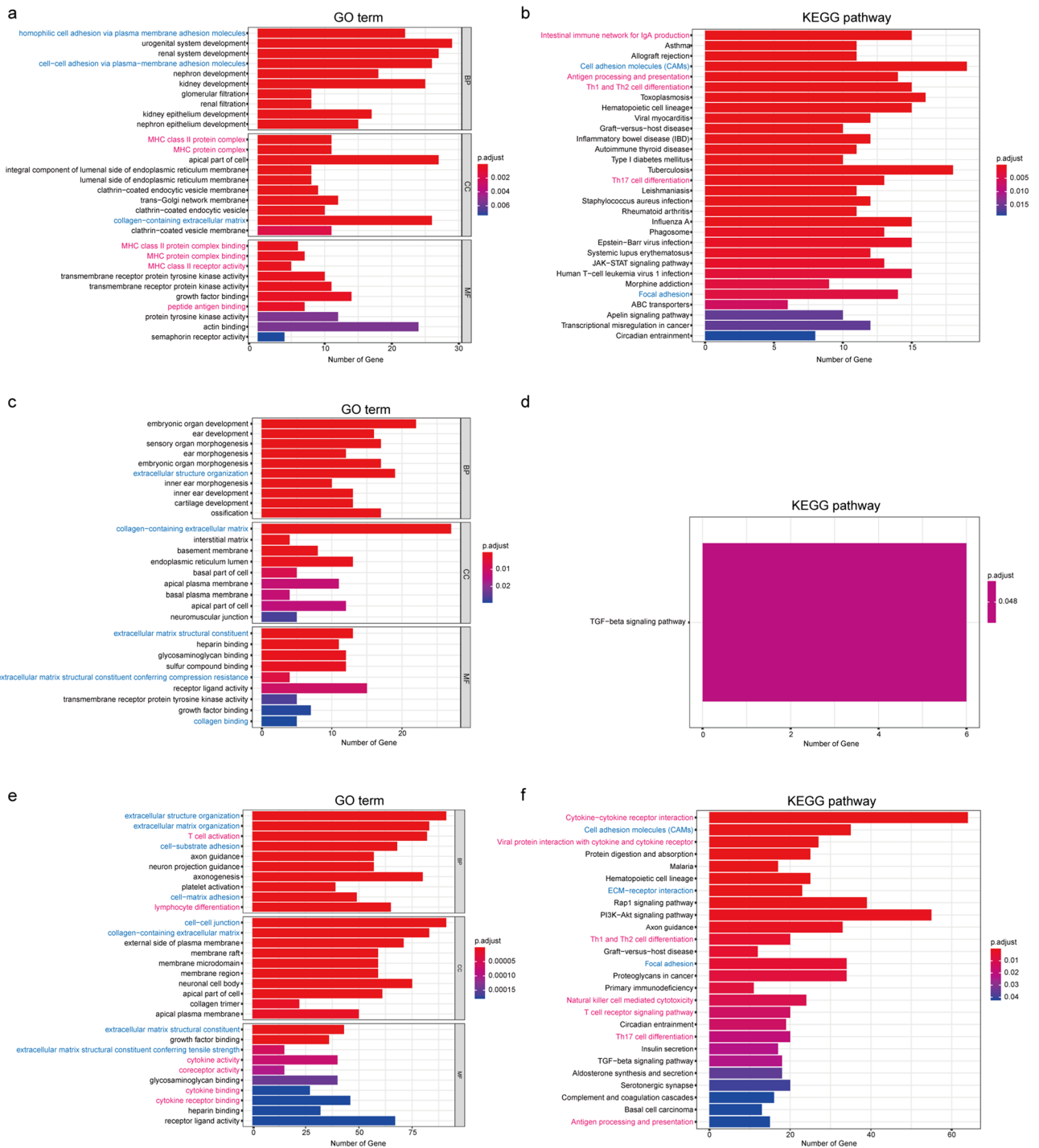
**Favorable prognostic mutations (*NPM1* and *biCEBPA*).** Patients in the *NPM1*/*biCEBPA* mutation groups individually tended to have higher 2-year OS rates than those in the WT groups (*NPM1*: 43.3% [95% CI 26.6–59.1%] vs 31.7% [95% CI 20.9–42.9%],  $P=0.48$ , Log-rank test, Fig. 2e; *biCEBPA*: 50.0% [95% CI 5.8–84.5%] vs 35.0% [95% CI 25.6–44.5%],  $P=0.35$ , Log-rank test, Fig. 2f).

*NPM1* mutation patients tended to have lower stromal scores than WT patients ( $-1172.08$  vs  $-1050.87$ ,  $P=0.11$ , Mann-Whitney test, two-sided), but their immune scores were similar ( $2716.05$  vs  $2683.65$ ,  $P=0.48$ , Mann-Whitney test, two-sided). Patients with *biCEBPA* mutation had similar immune scores and stromal scores compared with WT patients (immune score  $2688.10$  vs  $2683.34$ ,  $P=0.88$ , Mann-Whitney test, two-sided; stromal score  $-1010.38$  vs  $-1070.96$ ,  $P=0.94$ , Mann-Whitney test, two-sided) (Fig. 2g,h).

The above analysis reflected that there were distinct relationships between somatic mutations and immune/stromal scores. *RUNX1* and *TP53* mutations were related to both higher immune scores and higher stromal scores, *FLT3-ITD* was related to both lower immune scores and lower stromal scores, *ASXL1* mutation was only related to higher stromal scores, *NPM1* was only related to lower stromal scores, and *biCEBPA* was not related to immune scores and stromal scores.



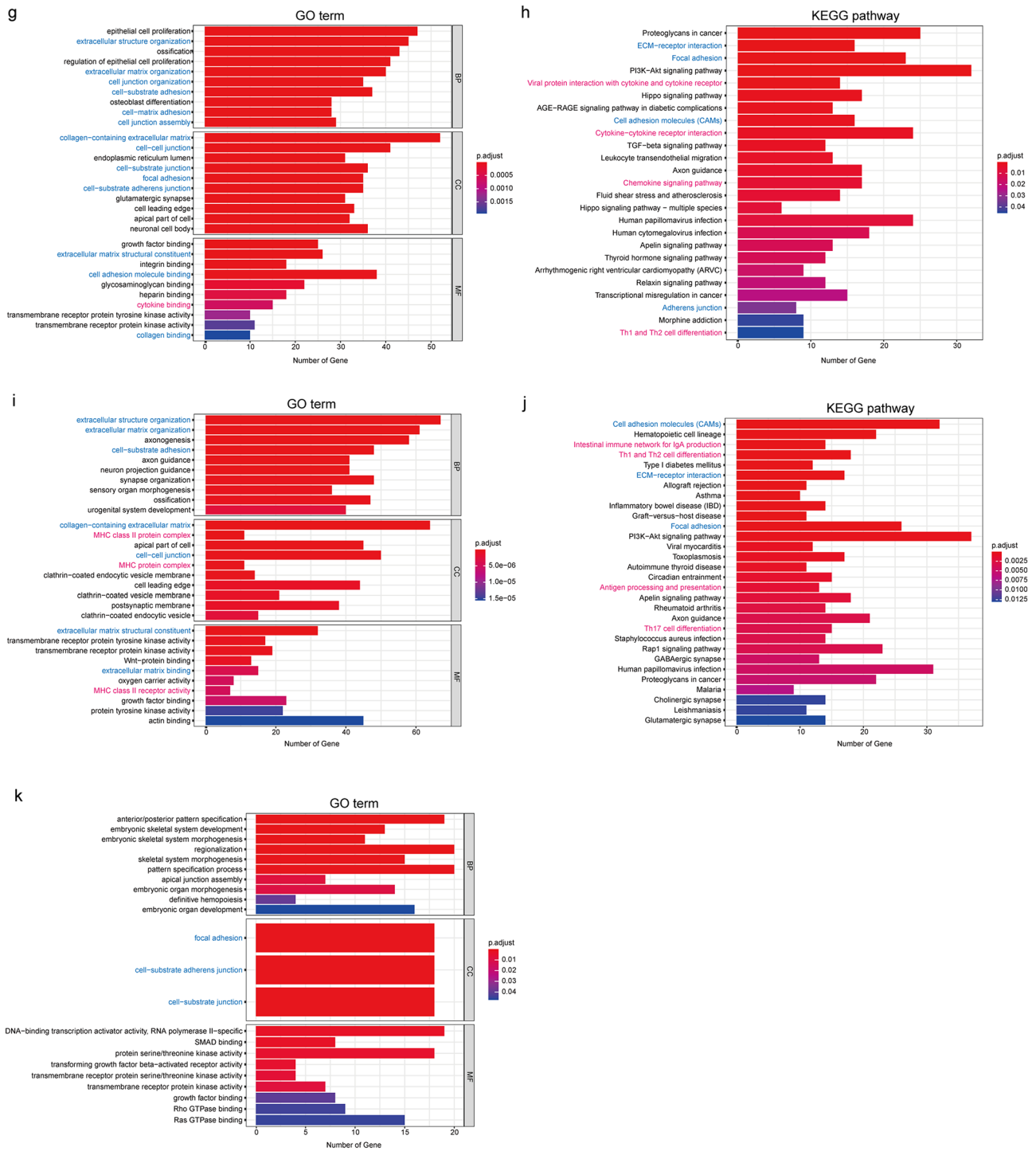
**Figure 2.** Mutations associated immune and stromal scores in the intermediate and poor cytogenetic risk groups. **(a–f)** Kaplan–Meier survival analysis of genetic mutation versus WT groups. **(g,h)** Immune and stromal scores of the genetic mutation versus WT groups.



**Figure 3.** GO term and KEGG pathway enrichment analysis of DEGs. **(a,b)** GO terms and KEGG pathways of RUNX1 upregulated genes. **(c,d)** GO terms and KEGG pathways of ASXL1 upregulated genes. **(e,f)** GO terms and KEGG pathways of TP53 upregulated genes. **(g,h)** GO terms and KEGG pathways of FLT3-ITD downregulated genes. **(i,j)** GO terms and KEGG pathways of NPM1 downregulated genes. **(k)** GO terms of biCEBPA downregulated genes. Note: immune-related pathways are marked *red*, and stromal-related pathways are marked *blue*.

**Identification of differentially expressed genes (DEGs) and functional enrichment analysis.** As shown in Supplementary Figure S1, after the gene expression data of patients with intermediate and poor cytogenetic risk was analyzed, a certain number of upregulated and downregulated genes were identified for the individual somatic mutations RUNX1, TP53, ASXL1, FLT3-ITD, NPM1 and biCEBPA.

The results of GO term and KEGG pathway enrichment analysis are shown in Fig. 3. In general, the enrichment analysis results were consistent with the immune/stromal scores. In other words, both immune- and



**Figure 3.** (continued)

stromal-related GO terms and KEGG pathways were enriched for RUNX1, TP53 and FLT3-ITD mutations, which corresponded to their association with immune and stromal scores. Only stromal-related GO terms were enriched for ASXL1 mutation, which corresponded to its only association with stromal scores. However, discrepancies were observed for NPM1 and biCEBPA mutations: both immune- and stromal-related GO terms and KEGG pathways were enriched for NPM1 mutation, although it was only related to lower stromal scores; stromal-related GO terms were enriched for biCEBPA mutation, despite its lack of a relationship to immune/stromal scores.

**Characteristics and prognostic significance of mutation-associated immune/stromal cell-relevant DEGs.** Mutation-associated immune/stromal score-relevant DEGs were selected according to ESTIMATE algorithm gene lists and are shown in Supplementary Table S1. Overall, the number of immune/stromal

cell-relevant DEGs was consistent with immune/stromal scores (Table 1): for RUNX1 and TP53 mutations, the majority of both immune and stromal cell-relevant DEGs were upregulated, which corresponded to their associated higher immune and stromal scores; for the ASXL1 mutation, only several stromal cell-relevant DEGs were upregulated, which corresponded to its associated higher stromal scores; the majority of both immune and stromal cell-relevant DEGs for FLT3-ITD and the majority of stromal cell-relevant DEGs for NPM1 mutation were downregulated, which corresponded to their associated lower scores; for biCEBPA mutation, almost no immune and stromal cell-relevant DEGs were observed, which corresponded to its lack of a relationship with immune and stromal scores. The only exception was that the majority of immune cell-relevant DEGs for NPM1 mutations were downregulated, despite their lack of association with immune scores.

There were overlaps among the individual mutation-associated immune/stromal cell-relevant DEGs (Supplementary Fig. S2). The common upregulated stromal cell-relevant DEGs among RUNX1, TP53 and ASXL1 mutations included DDR2 and FRZB. The common downregulated immune cell-relevant DEGs between FLT3-ITD and NPM1 mutations included CD3D, CD48, GBP1, and IL18RAP. The common downregulated stromal cell-relevant DEGs between FLT3-ITD and NPM1 mutations included BGN, CDH5, COL1A2, COL6A3, CXCL12, DCN, FRZB, ISLR, ITIH5, MXRA5, TRAT1, and VCAM1 (n = 12). Furthermore, FRZB was the only DEG associated with all 5 mutations (RUNX1, TP53, ASXL1, FLT3-ITD and NPM1), ITIH5 was associated with RUNX1, ASXL1, FLT3-ITD and NPM1 mutations, and ISLR was associated with TP53, ASXL1, FLT3-ITD and NPM1 mutations.

The intermediate and poor cytogenetic risk patients were grouped by the median transcription levels of the individual mutation-associated immune/stromal cell-relevant DEGs to evaluate their effects on OS. The following genes were found to have prognostic significance, which was consistent with that of the corresponding mutation: high expression of SCUBE2 (RUNX1 mutation-associated upregulation) was shown to be related to lower OS, high expression of SPON1 (TP53 mutation-associated upregulation) was shown to be related to lower OS, high expression of GREM1 (ASXL1 mutation-associated upregulation) was shown to be related to lower OS, low expression of COL3A1, CXCL12, EMCN, FRZB, ITIH5, KDR, MXRA5, TRAT1, and VCAM1 (FLT3-ITD mutation-associated downregulation) were shown to be related to lower OS, high expression of ADAMTS5 (NPM1 mutation-associated upregulation) was shown to be related to higher OS, and low expression of SPON1 (NPM1 mutation-associated downregulation) was shown to be related to higher OS (Supplementary Fig. S3).

**Validation in the GEO database and identification of hub genes.** To verify whether these immune and stromal cell-relevant DEGs identified from TCGA AML patients are also associated with mutations in an independent AML cohort, we analyzed the gene expression levels of 524 AML cases from GSE14468, for which FLT3-ITD, NPM1 and biCEBPA mutation data were available. For immune score-involved genes, 7/7 of FLT3-ITD-associated, 13/13 of NPM1 mutation-associated and 3/3 biCEBPA mutation-associated DEGs were confirmed to be significantly related to the individual gene mutation by GSE14468. Similarly, for stromal score-involved genes, 7/22 FLT3-ITD-associated, 20/32 NPM1 mutation-associated and 2/2 biCEBPA mutation-associated DEGs were also confirmed. The confirmed mutation-associated immune and stromal cell-relevant DEGs are shown in Table 2.

To further investigate the interaction among FLT3-ITD/NPM1-associated immune and stromal cell-relevant DEGs validated by GSE14468, we constructed PPI networks based on the STRING database and Cytoscape software. Then, we identified the top 10 FLT3-ITD/NPM1-associated hub genes by applying cytoHubba (Fig. 4). Node degree of importance is represented by circle color. As a result, the FLT3-ITD-associated hub genes were as follows: LCK, CD48, CD3D, IL2RB, CXCL12, VCAM1, DCN, IL10RA, TRAT1 and IL18RAP. NPM1-associated hub genes included VCAM1, CD3D, ZAP70, HLA-DRA, HLA-DPB1, HLA-DPA1, IRF8, CD48, GZMB and GBP1. LCK and VCAM1 were determined to be the most important hub genes in the networks associated with the FLT3-ITD mutation and the NPM1 mutation, respectively.

## Discussion

A number of genetic mutations have presented immune microenvironment modulatory properties in solid tumors: EGFR mutations correlate with an immunosuppressive TME and may impact the antitumor immune response in NSCLC<sup>22,23</sup>; TP53 and KRAS mutations in lung adenocarcinoma can regulate the immune microenvironment to affect PD-1 blockade immunotherapy<sup>24,25</sup>; JAK1 or JAK2 mutations may lead to acquired resistance to PD-1 blockade immunotherapy in patients with melanoma<sup>26</sup>. Recurrent genetic mutations found in AML have been heavily studied to classify and predict the risk of relapse after treatment. According to the ELN and NCCN guidelines, RUNX1, ASXL1, TP53, FLT3-ITD, NPM1 and biCEBPA mutations have been involved in AML prognostic stratification<sup>2,3</sup>. Interactions between leukemic stem cells and other cells in the BM microenvironment are known to be vital for the maintenance and progression of chemotherapy-resistant AML<sup>14</sup>. LSCs can remodel the BM niche into a favorable environment for expansion or even induce leukemic transformation. Nonetheless, the relationships between recurrent genetic mutations and the immune microenvironment in AML have not been comprehensively described<sup>27</sup>.

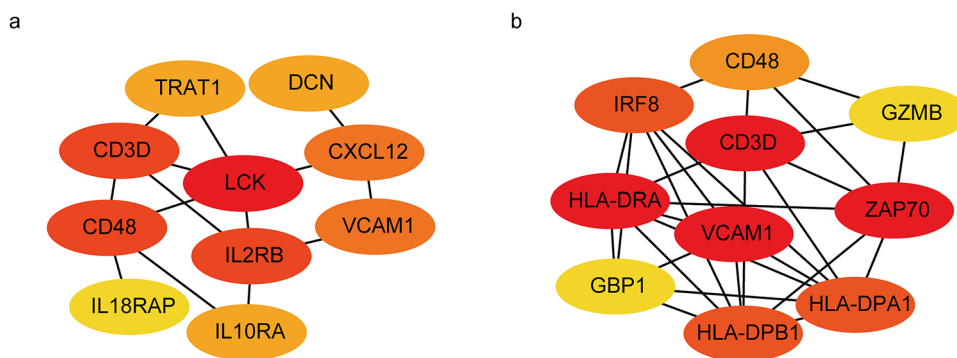
In our study, we calculated the immune and stromal scores for AML patients from the TCGA database based on the ESTIMATE algorithm. Our results showed that immune scores were significantly associated with OS and cytogenetic risk; a high immune score was a significantly poor prognostic factor for both the entire cohort and patients in the intermediate and poor cytogenetic risk groups, and a high stromal score tended to be correlated with poor OS in the intermediate and poor cytogenetic risk groups. The prognostic significance of immune and stromal scores was different in solid tumors: high immune and stromal scores correlated with poor survival in glioblastoma<sup>18</sup>, clear cell renal cell carcinoma (ccRCC)<sup>28</sup> and gastric cancer<sup>16</sup>, whereas high immune and stromal scores correlated with better survival in cervical squamous cell carcinoma (CSCC)<sup>29</sup> and pancreatic ductal

Gene	Immune		Stroma	
	Score (P values)	Number of DEGs* (Mu vs WT)	Score (P values)	Number of DEGs* (Mu vs WT)
RUNX1	Mu > WT (0.15)	Up: 8 down: 0	Mu > WT (0.0008)	Up: 17 down: 2
TP53	Mu > WT (0.35)	Up: 24 down: 2	Mu > WT (0.089)	UP: 41 down: 0
ASXL1	Mu ≈ WT (0.91)	Up: 0 down: 0	Mu > WT (0.014)	Up: 7 down: 1
FLT3-ITD	Mu < WT (0.061)	Up: 0 down: 7	Mu < WT (0.022)	Up: 2 down: 20
NPM1	Mu ≈ WT (0.48)	Up: 2 down: 11	Mu < WT (0.11)	Up: 5 down: 27
biCEBPA	Mu ≈ WT (0.88)	Up: 1 down: 2	Mu ≈ WT (0.94)	Up: 1 down: 1

**Table 1.** Characteristics of somatic mutation-associated immune/stromal cell-relevant DEGs. \*Only these involved in the ESTIMATE algorithm gene list were considered.

Mutation	Status of DEGs	Immune cell-relevant DEGs	Stromal cell-relevant DEGs
FLT3-ITD mutation	Upregulation	None	<i>ENPP2</i> ; <i>PTGIS</i> (n = 2)
	Downregulation	<i>CD3D</i> ; <i>CD48</i> ; <i>GBP1</i> ; <i>IL10RA</i> ; <i>IL18RAP</i> ; <i>IL2RB</i> ; <i>LCK</i> (n = 7)	<i>AOC3</i> ; <i>BGN</i> ; <i>C1QB</i> ; <i>CDH5</i> ; <i>COL15A1</i> ; <i>COL1A2</i> ; <i>COL3A1*</i> ; <i>COL6A3</i> ; <i>CXCL12*</i> ; <i>DCN</i> ; <i>EMCN*</i> ; <i>FRZB*</i> ; <i>ISLR</i> ; <i>ITIH5*</i> ; <i>KDR*</i> ; <i>MXRA5*</i> ; <i>PDE2A</i> ; <i>RARRES2</i> ; <i>TRAT1*</i> ; <i>VCAM1*</i> (n = 20)
NPM1 mutation	Upregulation	<i>CSTA</i> ; <i>SRGN</i> (n = 2)	<i>ADAMTS5*</i> ; <i>C3AR1</i> ; <i>F13A1</i> ; <i>TNN</i> ; <i>RASGRP3</i> (n = 5)
	Downregulation	<i>CD3D</i> ; <i>CD48</i> ; <i>GBP1</i> ; <i>GZMB</i> ; <i>HLA-DPA1</i> ; <i>HLA-DPB1</i> ; <i>IRF8</i> ; <i>HLA-DRA</i> ; <i>IL18RAP</i> ; <i>RGS1</i> ; <i>ZAP70</i> (n = 11)	<i>BGN</i> ; <i>CD200</i> ; <i>CD248</i> ; <i>CDH5</i> ; <i>COL1A2</i> ; <i>COL6A3</i> ; <i>COL8A2</i> ; <i>CXCL12</i> ; <i>DCN</i> ; <i>EDIL3</i> ; <i>ENPEP</i> ; <i>ERG</i> ; <i>FBLN2</i> ; <i>FRZB</i> ; <i>HGF</i> ; <i>IL18R1</i> ; <i>ISLR</i> ; <i>ITIH5</i> ; <i>ITM2A</i> ; <i>MXRA5</i> ; <i>PAPPA</i> ; <i>PRKG1</i> ; <i>SIGLEC1</i> ; <i>SPON1*</i> ; <i>SULF1</i> ; <i>TRAT1</i> ; <i>VCAM1</i> (n = 27)
biCEBPA mutation	Upregulation	<i>IL2RG</i> (n = 1)	<i>ITM2A</i> (n = 1)
	Downregulation	<i>GBP2</i> ; <i>IL4R</i> (n = 2)	<i>F13A1</i> (n = 1)

**Table 2.** FLT3-ITD, NPM1 and biCEBPA mutations-associated immune and stromal cell-relevant DEGs. Genes that displayed in *italics* were confirmed to be associated with mutations by GSE14468 data. \*These mutation-associated genes have prognostic significance in the intermediate and poor cytogenetic risk patients.



**Figure 4.** Results of degree algorithms from cytoHubba. (a) The PPI network of the top 10 hub genes associated with the FLT3-ITD mutation. (b) The PPI network of the top 10 hub genes associated with NPM1 mutations.

adenocarcinoma (PDAC)<sup>30</sup>. These studies demonstrated the varied effects of immune and stromal scores on prognosis, and these effects are related to tumor type.

Due to the similar OS rate, immune scores and stromal scores between the intermediate and poor cytogenetic risk groups, we considered these patients as a single group to explore the characteristics of the somatic mutation-associated immune microenvironment in AML. We compared the immune/stromal scores, identified DEGs between the mutation and WT groups, conducted functional enrichment analysis of DEGs and selected somatic mutation-associated immune/stromal cell-relevant DEGs. We found that similar to the impact of immune and stromal scores on prognosis, distinct relationships existed between somatic mutations and immune/stromal scores. In other words, RUNX1, ASXL1 and TP53 mutations were related to higher immune or stromal scores, whereas FLT3-ITD mutation was related to lower immune and stromal scores, although they were all poor



prognostic mutations. Furthermore, patients with NPM1 mutation had lower stromal scores, while patients with biCEBPA mutation showed similar immune and stromal scores, despite their favorable prognostic risks.

The functional enrichment analysis and immune/stromal cell-relevant DEGs were generally consistent with the immune/stromal scores for individual genetic mutations. There are unique and common genes among mutations associated with immune/stromal cell-relevant DEGs. The results obtained in this study demonstrated that RUNX1, TP53 and ASXL1 mutation-associated characteristics of the microenvironment are similar. Reports have revealed the pro-inflammatory impact of RUNX1, TP53 and ASXL1 mutations on the immune microenvironment. RUNX1 mutation has been shown to activate NF- $\kappa$ B signaling and has been proposed to promote inflammatory signaling pathways in the bone marrow microenvironment<sup>31</sup>. Previous reports have shown that TP53 mutations induce pro-inflammatory effects on epithelial cells through NF- $\kappa$ B-mediated production of inflammatory cytokines. Moreover, TP53 mutations in CAFs are associated with pro-tumor and pro-inflammatory effects through enhanced production of cytokines and chemokines, including CXCL12, SDF-1 and IL-6, which notably affect the immune microenvironment<sup>32,33</sup>. ASXL1 mutation is one of the most frequently observed mutations leading to clonal hematopoiesis (CH), which have been known to show elevated inflammation, impaired tumor suppressor function, and risk of eventual hematological malignancy (HM)<sup>34–36</sup>. Patients with NPM1 mutation and FLT3-ITD mutation not only had similar lower scores but also had multiple common immune/stromal cell-relevant DEGs, which were not consistent with their opposite prognostic significance. Our results indicated that there might be a common mechanism on the impact of NPM1 and FLT3-ITD mutations on the bone marrow microenvironment, which remains to be explored.

Notably, several mutations associated with immune/stromal cell-relevant DEGs were observed to have prognostic significance in intermediate and poor cytogenetic risk patients. The results implied that these specific genes may play an important role in the formation of a mutation-associated microenvironment and may affect the survival of AML patients. Moreover, the majority of immune/stromal cell-relevant DEGs were confirmed to be significantly correlated with FLT3-ITD, NPM1, and biCEBPA mutations in the GSE14468 database. PPI networks were subsequently built based on the verified FLT3-ITD/NPM1-associated immune and stromal cell-relevant DEGs, and the top 10 hub genes were subsequently identified by the degree of interaction.

LCK (lymphocyte-specific protein tyrosine kinase) was the most significant hub gene associated with the FLT3-ITD mutation. LCK plays an essential role in the selection and maturation of developing T cells in the thymus, the activation of mature T-cells and the initiation of T cell antigen receptor (TCR) signal transduction pathways<sup>37</sup>. Studies have indicated higher expression of LCK in leukemic cells from less differentiated cases of AML (AML-0 and AML-1)<sup>38</sup>. A recent report found that LCK is overexpressed and mutated in CTV-1 cells (AML-M5 cell lines)<sup>39</sup>. Nonetheless, the expression of LCK in FLT3-ITD-mutated cells has not been studied to date. In the present study, the downregulation of LCK was correlated with the FLT3-ITD mutation. In a study of a zebrafish model<sup>40</sup>, FLT3 was found to initiate definitive hematopoietic stem cells, and the knockdown of FLT3 reduced hematopoiesis. The expression of the FLT3-ITD mutation resulted in the expansion of myeloid cells and the reduction of T cells. These results suggest that the FLT3-ITD mutation decreases the expression of LCK and reduces the production of functional T cells.

VCAM1 (vascular cell adhesion molecule-1) was shown to be the most significant hub gene associated with NPM1 mutation. VCAM1 is a cell adhesion molecule primarily expressed on endothelial cells, and its expression is induced by pro-inflammatory cytokines, such as TNF $\alpha$ <sup>41,42</sup>. VCAM1 has been identified to regulate vascular adhesion and transendothelial migration by binding to VLA-4 (very late antigen-4, an  $\alpha 4\beta 1$  integrin) on leukocytes. VCAM1 binding to VLA-4 confers AML blast cell protection from chemotherapy-induced apoptosis<sup>43,44</sup>. In our study, the downregulation of VCAM1 was confirmed to be correlated with NPM1 mutation. Although no study to date has explored the function of VCAM1 in NPM1-mutated AML patients, we speculated that the downregulation of VCAM1 may reduce the stroma-mediated protection of leukemic cells, which might confer favorable outcomes to AML patients with NPM1 mutations.

In conclusion, we focused on the relationship between recurrent genetic mutations and the immune microenvironment in AML patients based on TCGA database by integrated bioinformatic approaches. Important immune and stromal cell-relevant DEGs that affected the immune landscape of patients with individual gene mutations were identified and validated. Considering the specific properties of the hematopoietic microenvironment of leukemia<sup>15</sup>, ESTIMATE may not accurately predict infiltrating stromal and immune cells for the AML microenvironment, and we need to develop a more suitable and accurate algorithm. Due to the limited patient numbers in mutation subgroups in the TCGA database, further investigation of these mutation-associated stromal and immune signatures in large clinical AML patient cohorts is warranted, which may provide new prognostic biomarkers to achieve precision tumor therapy. Our results may help to elucidate how AML genetic mutations modulate the immune microenvironment to better guide personalized immunotherapy in the era of precision medicine<sup>45</sup>.

## Materials and methods

**Database.** The transcriptional profiles and clinical and overall survival (OS) data of 173 AML patients were downloaded from the TCGA database (<https://portal.gdc.cancer.gov/>). The gene expression profile was measured experimentally using the Illumina HiSeq 2000 RNA Sequencing platform. Log<sub>2</sub> transformations were performed for all gene expression data. Immune and stromal scores were calculated by applying the ESTIMATE algorithm<sup>15</sup> to the mRNA expression data (<https://bioinformatics.mdanderson.org/estimate/>). The definitions of cytogenetic risk and risk-related somatic mutations were based on NCCN guidelines<sup>2</sup>.

For validation of the mutation-associated microenvironment signatures obtained from TCGA data, GSE14468 based on GPL570 (Affymetrix Human Genome U133 Plus 2.0 Array) were downloaded from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>), including cytogenetic risk, FLT3-ITD, NPM1 and CEBPA mutations.

**Identification of differentially expressed genes.** AML patients in the intermediate and poor cytogenetics risk categories were divided into mutation and wild-type (WT) groups according to the individual somatic mutation status (RUNX1, TP53, ASXL1, FLT3-ITD, and NPM1). For CEBPA, patients with biallelic CEBPA mutation were classified as the biCEBPA group, and patients with monoallelic mutation or wild-type CEBPA were classified as the WT group. Differentially expressed genes (DEGs) were identified using the limma package in R software (version 3.6.2; <https://www.r-project.org/>). Genes with  $|\log_2FC| > 1.0$  and adjusted  $P$  values ( $q$  values)  $< 0.05$  were selected as DEGs. Volcano plots were generated using the ggplot2 package in R software.

**Functional enrichment analysis of DEGs.** Functional enrichment analysis of DEGs was performed based on clusterProfiler, enrichplot, org.Hs.eg.db, and ggplot2 packages to identify the Gene Ontology (GO) categories, including biological processes (BP), cellular components (CC), and molecular functions (MF). Pathway enrichment analysis based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database<sup>46,47</sup> was also conducted using these packages. Upregulated and downregulated DEGs were annotated by functional enrichment analyses, and FDR (false discovery rate)  $< 0.05$  was considered to be significant. The top 10 GO terms in each of the BP, CC, MF and top 30 KEGG pathways are presented using bar plots.

**Immune/stromal cell-relevant DEGs and overall survival analysis.** According to genes selected by the ESTIMATE algorithm, immune/stromal cell-relevant DEGs of each mutation group were identified. To explore the prognostic value of these immune/stromal cell-relevant DEGs in predicting the overall survival of AML patients in the intermediate and poor cytogenetic risk groups, Kaplan–Meier survival curves were generated by the “survival” package in R software using the log-rank test.  $P$  values  $< 0.05$  were considered to be significant.

**Protein–protein interaction (PPI) network and hub genes.** Protein–protein interaction (PPI) network construction of FLT3-ITD/NPM1-associated immune and stromal cell-relevant DEGs validated by GSE14468 was based on the STRING online database (version 11.0; <https://string-db.org/>) and Cytoscape software (version 3.6.0; <https://cytoscape.org/>). We used cytoHubba to identify the top 10 hub genes according to the degree algorithm.

**Statistical analysis.** Comparisons of immune and stromal scores among cytogenetic risk groups were performed using one-way analysis of variance. Comparisons of immune and stromal scores between the mutation and WT groups were performed using the Mann–Whitney U test. Survival functions were estimated using the Kaplan–Meier method and were compared using the log-rank test. The SPSS Statistics 22.0 (IBM Corp. in Armonk, NY; <https://www.ibm.com/>) and GraphPad Prism 5.0 (GraphPad Software, La Jolla California USA, [www.graphpad.com](http://www.graphpad.com)) were used for the data analysis.

Received: 29 July 2020; Accepted: 21 October 2020

Published online: 04 November 2020

## References

1. Ferrara, F. & Schiffer, C. A. Acute myeloid leukaemia in adults. *Lancet* **381**, 484–495. [https://doi.org/10.1016/s0140-6736\(12\)61727-9](https://doi.org/10.1016/s0140-6736(12)61727-9) (2013).
2. Tallman, M. S. *et al.* Acute myeloid leukemia, version 3.2019, NCCN clinical practice guidelines in oncology. *J. Natl. Compr. Cancer Netw. JNCCN* **17**, 721–749. <https://doi.org/10.6004/jnccn.2019.0028> (2019).
3. Döhner, H. *et al.* Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* **129**, 424–447. <https://doi.org/10.1182/blood-2016-08-733196> (2017).
4. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674. <https://doi.org/10.1016/j.cell.2011.02.013> (2011).
5. Hanahan, D. & Coussens, L. M. Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell* **21**, 309–322. <https://doi.org/10.1016/j.ccr.2012.02.022> (2012).
6. Pitt, J. M. *et al.* Targeting the tumor microenvironment: removing obstruction to anticancer immune responses and immunotherapy. *Ann. Oncol.* **27**, 1482–1492. <https://doi.org/10.1093/annonc/mdw168> (2016).
7. Wu, T. & Dai, Y. Tumor microenvironment and therapeutic response. *Cancer Lett.* **387**, 61–68. <https://doi.org/10.1016/j.canlet.2016.01.043> (2017).
8. Achyut, B. R. & Arbab, A. S. Myeloid cell signatures in tumor microenvironment predicts therapeutic response in cancer. *Oncotargets and therapy* **9**, 1047–1055. <https://doi.org/10.2147/ott.S102907> (2016).
9. Gbolahan, O. B. *et al.* Immunotherapeutic concepts to target acute myeloid leukemia: focusing on the role of monoclonal antibodies, hypomethylating agents and the leukemic microenvironment. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms18081660> (2017).
10. Binnewies, M. *et al.* Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat. Med.* **24**, 541–550. <https://doi.org/10.1038/s41591-018-0014-x> (2018).
11. Lamballe, A. J. & Lind, E. F. Targeting the immune microenvironment in acute myeloid leukemia: a focus on t cell immunity. *Front. Oncol.* **8**, 213. <https://doi.org/10.3389/fonc.2018.00213> (2018).
12. Shafi, A. A. *et al.* Patient-derived models reveal impact of the tumor microenvironment on therapeutic response. *Eur. Urol. Oncol.* **1**, 325–337. <https://doi.org/10.1016/j.euo.2018.04.019> (2018).
13. Taube, J. M. *et al.* Implications of the tumor immune microenvironment for staging and therapeutics. *Mod. Pathol.* **31**, 214–234. <https://doi.org/10.1038/modpathol.2017.156> (2018).
14. Ladikou, E. E., Sivaloganathan, H., Pepper, A. & Chevassut, T. Acute myeloid leukaemia in its niche: the bone marrow microenvironment in acute myeloid leukaemia. *Curr. Oncol. Rep.* **22**, 27. <https://doi.org/10.1007/s11912-020-0885-0> (2020).
15. Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612. <https://doi.org/10.1038/ncomms3612> (2013).

16. Wang, H., Wu, X. & Chen, Y. Stromal-immune score-based gene signature: a prognosis stratification tool in gastric cancer. *Front. Oncol.* **9**, 1212. <https://doi.org/10.3389/fonc.2019.01212> (2019).
17. Xie, P. *et al.* Development of an immune-related prognostic signature in breast cancer. *Front. Genet.* **10**, 1390. <https://doi.org/10.3389/fgene.2019.01390> (2019).
18. Jia, D. *et al.* Mining TCGA database for genes of prognostic value in glioblastoma microenvironment. *Aging* **10**, 592–605. <https://doi.org/10.18632/aging.101415> (2018).
19. Huang, S. *et al.* Identification of prognostic genes in the acute myeloid leukemia microenvironment. *Aging* **11**, 10557–10580. <https://doi.org/10.18632/aging.102477> (2019).
20. Ni, J. *et al.* Screening the cancer genome atlas database for genes of prognostic value in acute myeloid leukemia. *Front. Oncol.* **9**, 1509. <https://doi.org/10.3389/fonc.2019.01509> (2019).
21. Yan, H. *et al.* Identification of prognostic genes in the acute myeloid leukemia immune microenvironment based on TCGA data analysis. *Cancer Immunol. Immunother. CII* **68**, 1971–1978. <https://doi.org/10.1007/s00262-019-02408-7> (2019).
22. Lin, A., Wei, T., Meng, H., Luo, P. & Zhang, J. Role of the dynamic tumor microenvironment in controversies regarding immune checkpoint inhibitors for the treatment of non-small cell lung cancer (NSCLC) with EGFR mutations. *Mol. Cancer* **18**, 139. <https://doi.org/10.1186/s12943-019-1062-7> (2019).
23. Jia, Y. *et al.* EGFR-targeted therapy alters the tumor microenvironment in EGFR-driven lung tumors: Implications for combination therapies. *Int. J. Cancer* **145**, 1432–1444. <https://doi.org/10.1002/ijc.32191> (2019).
24. Biton, J. *et al.* TP53, STK11, and EGFR mutations predict tumor immune profile and the response to anti-PD-1 in lung adenocarcinoma. *Clin. Cancer Res.* **24**, 5710–5723. <https://doi.org/10.1158/1078-0432.Ccr-18-0163> (2018).
25. Dong, Z. Y. *et al.* Potential predictive value of TP53 and KRAS mutation status for response to PD-1 blockade immunotherapy in lung adenocarcinoma. *Clin. Cancer Res.* **23**, 3012–3024. <https://doi.org/10.1158/1078-0432.Ccr-16-2554> (2017).
26. Zaretsky, J. M. *et al.* Mutations associated with acquired resistance to PD-1 blockade in melanoma. *N. Engl. J. Med.* **375**, 819–829. <https://doi.org/10.1056/NEJMoal604958> (2016).
27. Mendez, L. M., Posey, R. R. & Pandolfi, P. P. The interplay between the genetic and immune landscapes of aml: mechanisms and implications for risk stratification and therapy. *Front. Oncol.* **9**, 1162. <https://doi.org/10.3389/fonc.2019.01162> (2019).
28. Xu, W. H. *et al.* Prognostic value and immune infiltration of novel signatures in clear cell renal cell carcinoma microenvironment. *Aging* **11**, 6999–7020. <https://doi.org/10.18632/aging.102233> (2019).
29. Pan, X. B., Lu, Y., Huang, J. L., Long, Y. & Yao, D. S. Prognostic genes in the tumor microenvironment in cervical squamous cell carcinoma. *Aging* **11**, 10154–10166. <https://doi.org/10.18632/aging.102429> (2019).
30. Pu, N. *et al.* Genetic landscape of prognostic value in pancreatic ductal adenocarcinoma microenvironment. *Ann. Transl. Med.* **7**, 645. <https://doi.org/10.21037/atm.2019.10.91> (2019).
31. Nakagawa, M. *et al.* AML1/RUNX1 functions as a cytoplasmic attenuator of NF- $\kappa$ B signaling in the repression of myeloid tumors. *Blood* **118**, 6626–6637. <https://doi.org/10.1182/blood-2010-12-326710> (2011).
32. Taura, M. *et al.* p53 regulates Toll-like receptor 3 expression and function in human epithelial cell lines. *Mol. Cell. Biol.* **28**, 6557–6567. <https://doi.org/10.1128/mcb.01202-08> (2008).
33. Cui, Y. & Guo, G. Immunomodulatory function of the tumor suppressor p53 in host immune response and the tumor microenvironment. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms17111942> (2016).
34. Genovese, G. *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* **371**, 2477–2487. <https://doi.org/10.1056/NEJMoal409405> (2014).
35. Jaiswal, S. *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.* **371**, 2488–2498. <https://doi.org/10.1056/NEJMoal408617> (2014).
36. Kaner, J. *et al.* Clonal hematopoiesis and premalignant diseases. *Cold Spring Harbor Perspect. Med.* <https://doi.org/10.1101/cshperspect.a035675> (2020).
37. Alarcón, B. & van Santen, H. M. Two receptors, two kinases, and T cell lineage determination. *Sci. Signal.* **3**, pe11. <https://doi.org/10.1126/scisignal.3114pe11> (2010).
38. Rouer, E., Dreyfus, F., Melle, J. & Benarous, R. Pattern of expression of five alternative transcripts of the lck gene in different hematopoietic malignancies: correlation of the level of lck messenger RNA I B with the immature phenotype of the malignancy. *Cell Growth Differ* **5**, 659–666 (1994).
39. Li, L., Cui, Y., Shen, J., Dobson, H. & Sun, G. Evidence for activated Lck protein tyrosine kinase as the driver of proliferation in acute myeloid leukemia cell, CTV-1. *Leuk. Res.* **78**, 12–20. <https://doi.org/10.1016/j.leukres.2019.01.006> (2019).
40. He, B. L. *et al.* Functions of flt3 in zebrafish hematopoiesis and its relevance to human acute myeloid leukemia. *Blood* **123**, 2518–2529. <https://doi.org/10.1182/blood-2013-02-486688> (2014).
41. Osborn, L. *et al.* Direct expression cloning of vascular cell adhesion molecule 1, a cytokine-induced endothelial protein that binds to lymphocytes. *Cell* **59**, 1203–1211. [https://doi.org/10.1016/0092-8674\(89\)90775-7](https://doi.org/10.1016/0092-8674(89)90775-7) (1989).
42. Kong, D. H., Kim, Y. K., Kim, M. R., Jang, J. H. & Lee, S. Emerging roles of vascular cell adhesion molecule-1 (VCAM-1) in immunological disorders and cancer. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms19041057> (2018).
43. Cavenagh, J. D., Gordon-Smith, E. C., Gibson, F. M. & Gordon, M. Y. Acute myeloid leukaemia blast cells bind to human endothelium in vitro utilizing E-selectin and vascular cell adhesion molecule-1 (VCAM-1). *Br. J. Haematol.* **85**, 285–291. <https://doi.org/10.1111/j.1365-2141.1993.tb03168.x> (1993).
44. Lévesque, J. P., Helwani, F. M. & Winkler, I. G. The endosteal “osteoblastic” niche and its role in hematopoietic stem cell homing and mobilization. *Leukemia* **24**, 1979–1992. <https://doi.org/10.1038/leu.2010.214> (2010).
45. Scheetz, L. *et al.* Engineering patient-specific cancer immunotherapies. *Nat. Biomed. Eng.* **3**, 768–782. <https://doi.org/10.1038/s41551-019-0436-x> (2019).
46. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30. <https://doi.org/10.1093/nar/28.1.27> (2000).
47. Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K. & Tanabe, M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* **47**, D590–D595. <https://doi.org/10.1093/nar/gky962> (2018).

## Author contributions

Y.-Z.Q. designed the study. Material preparation, data collection and analysis were performed by J.W. F.-T.D. and L.Y. J.W. wrote the manuscript. Y.-Z.Q. revised the manuscript. All authors read and approved the final manuscript.

## Funding

This work was supported by the National Nature Science Foundation of China (81870125).

## Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-76048-8>.

**Correspondence** and requests for materials should be addressed to Y.-Z.Q.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**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/>.

© The Author(s) 2020