Prognostic value and potential mechanism of long non-coding RNA Lnc-SMIM20-1 in acute myeloid leukemia

Sha Liu*, Ziyi Sun*, Mengyuan Zhu*, Minling Liu, Min Wei, Xiaofen Pan and Shan Huang

Department of Oncology, Jingzhou Central Hospital, The Second Clinical Medical College, Yangtze University, Jingzhou, Hubei, China;
*Department of Oncology, Taikang Tongji (Wuhan) Hospital, Wuhan, Hubei, China; Department of Oncology, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, Guangdong, China

ABSTRACT
Objectives: Acute myeloid leukemia (AML) is a common hematologic malignancy with high heterogeneity and poor prognosis. Although long non-coding RNAs (lncRNAs) have been used as biomarkers for tumors, the clinical relevance of numerous lncRNAs in AML remains to be investigated.

Research design and methods: Differentially expressed lncRNAs between AML and normal peripheral blood samples were identified using DESeq2. Pan-cancer analysis was performed by GEPIA tool. Kaplan–Meier survival curve was applied for prognosis analysis. KEGG pathway analysis and GSEA were used for functional enrichment. The ceRNA network was constructed by GDCRNAtools.

Results: Lnc-SMIM20-1 was most highly expressed in AML and up-regulated in the TCGA-AML cohort compared to normal tissues. Patients with high expression of Lnc-SMIM20-1 had poor overall prognosis both in the TCGA adult AML cohort and the TARGET pediatric AML cohort, no matter whether they were treated with chemotherapy or allo-HSCT. Lnc-SMIM20-1 might participate in cancer-associated signaling pathways and immune-related signaling pathways by interacting with four microRNAs and 20 mRNAs.

Conclusion: Lnc-SMIM20-1 was up-regulated in AML acting as a stable poor prognostic factor. The prognostic impact of Lnc-SMIM20-1 cannot be overcome by allo-HSCT. Our findings provide insight into the clinical relevance of Lnc-SMIM20-1 in AML; aiming to progress the development of novel therapeutics.

1. Background
Acute myeloid leukemia (AML) is a clonal and malignant proliferative disease of the hematopoietic system with poor prognosis and high mortality, and it is the most common acute leukemia in adults. Currently, many prognostic molecular markers for AML have been well established [1]. For example, mutations of the NPM1 gene, encoding for a nucleolar multifunctional protein, occur in over 30% of patients with AML. The European Leukemia Net (ELN) have declared that AML with NPM1 mutations was associated with good prognosis [2,3]. FLT3-ITD is a common driver mutation that presents with a high leukemic burden and confers a poor prognosis in patients with AML [4,5]. AML patients with biallelic mutations of CEBPA display a favorable clinical outcome and are defined as a unique entity in the 2016 World Health Organization classification [6,7]. Besides, Somatic or germline mutant RUNX1 is associated with poorer outcomes in AML [8–10]. In recent years, various studies demonstrated that long non-coding RNAs (lncRNAs) act as oncogenes or tumor suppressors in cancer [11,12]. In leukemia, numerous studies have illuminated that lncRNA PVT1 played a crucial role in hematologic malignancy cell proliferation and apoptosis [13]. LncRNA CCAT1 was a regulator in AML by modulating miR-155 availability [14]. LncRNA HOTAIR was overexpressed in AML and associated with poor prognosis [15,16]. LncRNA MALAT1 regulates migration, proliferation, and apoptosis by sponging miR-146a to regulate CXCR4 expression in AML cells [17]. However, the oncogenic roles and the potential mechanism of plenty of lncRNAs remain to be elucidated. In this study, we found that the expression of LncRNA Lnc-SMIM20-1 was dysregulated in AML and associated with the poor prognosis of AML. We preliminary unveiled that Lnc-SMIM20-1 might be related to multiple clinically actionable signaling pathways in AML, including the PI3K-Akt signaling pathway, JAK-STAT signaling pathway, and MAPK signaling pathway. Our findings provide insightful understanding of Lnc-SMIM20-1 in AML and may contribute to AML diagnosis and the development of therapeutics.

2. Methods
2.1. Data collection
The complete transcriptome profile, microRNA expression profile, and clinical data of 151 AML patients were downloaded from The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov) Data released in 2019. Besides, the
transcriptome profiles of other 32 types of cancer were also downloaded from the TCGA database. As verification, 187 patients with pediatric Leukemia from the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) database were also enrolled in this study, and the transcriptome profile and clinical data were available on the TARGET website (https://ocg.cancer.gov/programs/target Data released in 2019). The transcriptome profile of 70 normal peripheral blood samples was downloaded from the Genotype-Tissue Expression Project (GTEx) database (https://gtexportal.org/ Data released in 2016).

2.2. Selection of prognostic lncRNA in CN-AML

Firstly, DESeq2 [18] was used to perform differentially expressed genes analyses between TCGA-AML cohort and the normal peripheral blood samples in the GTEx dataset [19–21], and 8594 up-regulated lncRNAs were identified for further analysis (Supplementary Table S1). Secondly, the expression of lncRNA was performed in pan-cancer, and 6847 lncRNAs were selected as the most highly expressed lncRNAs in AML across 33 types of cancer (Supplementary Table S2). And then, we overlapped the lncRNAs between two sets of lncRNAs mentioned above, and 4194 lncRNAs were extracted (Supplementary Table S3). Thirdly, univariate analysis was performed in both the TCGA-AML cohort and the TARGET-AML cohort to identify the prognostic lncRNAs and 102 lncRNAs were identified (Supplementary Table S4). Subsequently, we divided the TCGA-AML cohort into two groups based on the different treatments (chemotherapy and allo-HSCT). And then, a univariate analysis was performed to identify the prognostic lncRNAs in both patients who received chemotherapy and underwent allo-HSCT. Finally, only Lnc-SMIM20-1 (ENSEMBL: ENSG00000250541) was selected (Supplementary Table S5). Therefore, we considered that Lnc-SMIM20-1 could be the potential prognostic indicator in AML. A flow chart of this process was displayed in Supplementary Figure S1.

2.3. Gene Expression Profiling Interactive Analysis (GEPIA)

The GEPIA database (https://gepia.cancer-pku.cn) is a comprehensive resource for the systematic analysis of gene expression [22]. This database includes 9736 tumor and 8587 normal tissue samples from the TCGA and the Genotype-Tissue Expression (GTEx) projects. In the function module, we used the Ensembl Gene ID of Lnc-SMIM20-1: ENSG00000250541 to perform the single-gene analysis and the threshold was selected as the default value.

2.4. Gene expression analysis

The log counts per million (logCPM) that were transformed from raw counts by R voom package were used to measure the expression of genes. Comparison of Lnc-SMIM20-1 expression between the TCGA-AML cohort and GTEx cohort was analyzed by DESeq2(|FC|≥1.5, p ≤ 0.01). All statistical analyses were performed using R version 4.0.

2.5. Prognosis analysis

After rejecting 15 M3 patients, a total of 136 patients were divided into two groups (Lnc-SMIM20-1high and Lnc-SMIM20-1low) based on the median expression of Lnc-SMIM20-1. Kaplan–Meier analysis and log-rank test were performed to compare the overall survival and the event-free survival between the two groups. Multivariable analysis was conducted to demonstrate whether the expression of Lnc-SMIM20-1 was an independent factor for prognosis in AML. The p-value <0.05 was considered significant.

2.6. Differentially expressed gene analysis and function enrichment analysis

Differentially expressed genes (DEGs) were identified by differentially expressed gene analysis using DESeq2 [18], the threshold value of differential gene analysis was false discovery rate (FDR)<0.05 and |FC|>1.5. Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were conducted using R package ‘clusterProfiler’ [23]. Gene Set Enrichment Analysis (GSEA) was conducted using, gseKEGG and gsePathway function in R package ‘clusterProfiler,’ with the parameters: nPerm = 1000, minGSSize = 10, maxGSSize = 1000, pvalueCutoff = 0.05.

2.7. Competing endogenous RNAs (ceRNA) network construction and analysis

Differentially expressed microRNAs and mRNAs selected from the DEGs were used to construct the Lnc-SMIM20-1 ceRNA network, which was constructed by GDCRNAtools [24] and visualized by ‘igraph’ R package. In the ceRNA network, three criteria are used to determine the competing endogenous interactions between lncRNA-mRNA pairs by GDCRNAtools: the lncRNA and mRNA must share a significant number of miRNAs, expression of lncRNA and mRNA must be positively correlated, and those common miRNAs should play similar roles in regulating the expression of lncRNA and mRNA. So ceRNA network was constructed satisfying the following requirements: hyperPValue > 0.01, corPValue > 0.01, regSim = 0. miRcode was used to collect predicted and experimentally validated lncRNA targets. StarBase v2.0 was used to predict miRNA–mRNA interactions. Next, functional analysis based on the ceRNA was performed and noted using Metascape (https://metascape.org/) [25]. The least absolute shrinkage and selection operator (LASSO) Cox regression was conducted to establish the prognostic model based on the Lnc-SMIM20-1 related ceRNA network by using the R package glmnet [26,27]. The performance of the prognostic model was evaluated in the TCGA-AML cohort by Kaplan-Meier OS curve analysis and receiver operating characteristic (ROC) analysis.

2.8. Immune infiltrate analysis

The CIBERSORT tool (https://cibersort.stanford.edu/) is an analytical tool to provide an estimation of the abundances of member cell types in a mixed cell population by using gene
expression data. We estimated the immune cell infiltration in the TCGA-AML cohort by using the CIBERSORT. Pearson’s correlation analysis was used to evaluate the relationship between Lnc-SMIM20-1 and immune cells.

3. Results

3.1. Lnc-SMIM20-1 overexpression in AML

To figure out whether Lnc-SMIM20-1 play important role in AML specifically, GEPIA online tool was applied to perform the pan-cancer analysis. The results showed that Lnc-SMIM20-1 was majorly over-expressed in Esophageal carcinoma (ESCA) tumor tissues, Stomach adenocarcinoma (STAD) tumor tissues, and Acute Myeloid Leukemia (LAML). Notably, Lnc-SMIM20-1 was the most highly expressed in AML (Figure 1(a)). Subsequently, the expression of Lnc-SMIM20-1 was up-regulated in the TCGA-AML cohort compared with the normal peripheral blood samples in the GTEx cohort (Figure 1(b), p < 0.0001). And then, AML patients were stratified into favorable, intermediate, and adverse groups using the ELN stratification system (2017 version). The significant difference of expression level of Lnc-SMIM20-1 between these three groups was found by ANOVA (Figure 1(c), p=3e-06). And then, we further identified that the favorable group was associated with comparatively low expression of Lnc-SMIM20-1 using the Tukey HSD test method, which was validated using the Least – Significant Difference (LSD) test method (Supplementary Table S6). Next, we evaluated Lnc-SMIM20-1 expression among TCGA-AML patients with cytogenetics subtypes including RUNX1-RUNX1T1, CBFB-MYH11, KMT2A rearrangements, PML-RARA, BCR-ABL1, and complex, as well as the normal karyotype. The significant difference of expression level of Lnc-SMIM20-1 between these seven groups was also found by ANOVA (Figure 1(d), p = 0.0016). And then, we further found that expression of Lnc-SMIM20-1 was significantly low in patients with CBFB-MYH11 comparing to those in the other subgroup, while patients with KMT2A rearrangements showed significantly high expression of Lnc-SMIM20-1 comparing to those with RUNX1-RUNX1T1, CBFB-MYH11, PML-RARA, or normal karyotype using the Tukey HSD test method, which were also validated using the LSD test (Supplementary Table S7). In addition, we performed the same process in the TARGET pediatric AML cohort. The result showed that there is significant difference between the expression level of Lnc-SMIM20-1 in patients with different cytogenetics subtype (Supplementary Figure S2, p = 2.6e-06). Pediatric AML patients with CBFB-MYH11 had significantly lower expression of Lnc-SMIM20-1 comparing to those with normal karyotype or other, while patients with KMT2A rearrangements had significantly higher expression of Lnc-SMIM20-1 compared to those with CBFB-MYH11 or t(8;21) (Supplementary Table S8), which was consistent with the results in the TCGA-AML cohort.

3.2. The prognostic value of Lnc-SMIM20-1 in AML

AML patients were separated into Lnc-SMIM20-1high group and Lnc-SMIM20-1low group based on the median expression of Lnc-SMIM20-1. Subsequently, Kaplan–Meier survival curves showed that Lnc-SMIM20-1high group had to decrease OS (Figure 2(a), p = 0.000628) and EFS (Figure 2(b), p = 0.0022) compared with Lnc-SMIM20-1low group. As validation, another independent cohort including 187 patients with pediatric Leukemia in TARGET was employed. The survival curve showed no difference in the outcome between the TARGET-AML cohort and the TCGA-AML dataset. The OS and EFS of the TARGET AML patients with high expression of Lnc-SMIM20-1 were worse than that of patients with low expression of Lnc-SMIM20-1 (Figure 2(c,d); p = 0.00204, p = 0.0469).

3.3. Evaluation of the expression of Lnc-SMIM20-1 and clinical features of AML

According to the median of Lnc-SMIM20-1 expression, 136 AML patients were divided into Lnc-SMIM20-1high group and Lnc-SMIM20-1low group, each group included 68 patients. The significance of the difference in various clinical and molecular features between Lnc-SMIM20-1high and Lnc-SMIM20-1low groups indicated that AML Patients with CBFB-MYH11 or RUNX1-RUNX1T1 were more significantly clustered in Lnc-SMIM20-1low group (Figure 3(a), p = 0.001, p = 0.013). Besides, Lnc-SMIM20-1low group had more AML patients with CEBPA double mutation (Figure 3(a), p = 0.034); inversely, more AML patients with NPM1 or DNMT3A mutation were enrolled in Lnc-SMIM20-1high group (Figure 3(a), p = 0.046, p = 0.017). No statistical significance was found between two groups in treatment (chemotherapy versus allo-HSCT), age (<60 years versus ≥60 years), gender (male versus female), white blood cell (WBC) counts (>15 × 10^9/L versus ≥15 × 10^9/L), bone marrow (BM) blast counts (<20% versus ≥20%), peripheral blood (PB) blast counts (<70% versus ≥70%), and frequency of other recurrent genetic mutations (FLT3-ITD, IDH1/2, etc., wild type versus mutation). The details were showed in Supplementary Table S9.

Afterward, multivariate Cox regression analysis showed that the HR of the expression of Lnc-SMIM20-1 and 95% confidence interval (CI) were 1.68 and 1.07–2.65 (Figure 3(b), p = 0.025); meanwhile, CBFB-MYH11 mutation was also associated with prognosis in AML (Figure 3(b), HR = 0.30, 95%CI = 0.09–0.99, p = 0.049). The details were showed in Supplementary Table S10. Besides, owing to the above result indicating that AML with KMT2A rearrangement had higher expression level of Lnc-SMIM20-1 compared to other subtypes of AML, we included KMT2A rearrangement in the multivariate analysis to further validate the independence of Lnc-SMIM20-1 on the prognosis in AML and the result also indicated that expression level of Lnc-SMIM20-1 was an independent poor prognostic factor in AML (Supplementary Figure S3).

3.4. Comparison of the Lnc-SMIM20-1 expression in AML patients received different treatment

Subsequently, the TCGA-AML cohort was divided into the chemotherapy-only group including 70 AML patients who only received chemotherapy, and the allo-HSCT group including 66 AML patients who underwent allo-HSCT. In the chemotherapy-only group, AML patients with high Lnc-SMIM20-1 expression were associated with worse OS (Figure 4(b), p = 0.00841) and EFS (Figure 4(b), p = 0.00107). In the allo-HSCT group, the OS of AML patients with high expression of Lnc-SMIM20-1 was shorter than
Figure 1. Lnc-SMIM20-1 expression in AML. 

(a) Lnc-SMIM20-1 expression in various cancer tissues and normal tissues analyzed by GEPIA; red dots represent the expression of a certain tumor type and green dots represent the expression of normal tissue. 

(b) Lnc-SMIM20-1 expression of AML and normal peripheral blood samples analyzed by TCGA and GTEx database. 

(c) Lnc-SMIM20-1 expression of AML patients stratified by ELN stratification system in TCGA-AML cohort. 

(d) Lnc-SMIM20-1 expression of AML patients with various chromosomal abnormalities TCGA-AML cohort.
that of AML patients with low expression of Lnc-SMIM20-1 (Figure 4(c), \( p = 0.0118 \)); while the EFS of AML patients had no significant difference between high- and low-expression of Lnc-SMIM20-1 (Figure 4(d), \( p = 0.115 \)). In addition, we found 16 patients in the TCGA-AML cohort that received decitabine or azacitidine which were hypomethylating agents. Hypomethylating agent (HMA) strategies are also often recommended and used as frontline treatment approaches in elderly patients [28,29], so we divided the AML patients that received hypomethylating agents into high and low-expression groups based on the median expression level of Inc-smim20-1 after excluding 2 patients without expression data. And then, we compared the overall survival between the two groups using Kaplan–Meier survival analysis. The result showed no significant difference between the overall survivals of the two groups (Supplementary Figure S4). The possible reason for this result was the small size of patients (\( n = 14 \)).

### 3.5. Potential Mechanism of Lnc-SMIM20-1 in AML

Differentially expressed genes (DEGs) analysis was performed between Lnc-SMIM20-1\(^{\text{high}}\) group and Lnc-SMIM20-1\(^{\text{low}}\) group and 361 up-regulated genes and 157 down-regulated genes were found (Figure 5(a), Supplementary Table S11). KEGG pathway analysis showed that the DEGs were enriched in Hematopoietic cell lineage, T cell receptor signaling pathway, Th1 and Th2 cell differentiation, Focal adhesion, Rap1 signaling pathway, Th17 cell differentiation, Cell adhesion molecules, Primary immunodeficiency, and NF-kappa B signaling pathway (Figure 5(b)). Besides, GSEA results further indicated that high expression of Inc-SMIM20-1 was associated with the activation of PI3K-Akt signaling pathway, JAK-STAT signaling pathway, Cytokine–cytokine receptor interaction, T cell receptor signaling pathway, Th1 and Th2 cell differentiation, and Th17 cell differentiation (Figure 5(c–h)).
Supplementary Table S12). Notably, many enriched signaling pathways were closely related to tumor occurrence and development, such as NF-kappa B signaling pathway, PI3K-Akt signaling pathway, and JAK-STAT signaling pathway. Besides, some enriched pathways were related to immunoregulation, such as the T cell receptor signaling pathway, Th1 and Th2 cell differentiation, Th17 cell differentiation, and primary immunodeficiency. Moreover, immune infiltrate analysis was also performed to further indicate the association between Lnc-SMIM20-1 and immune cell infiltration in AML. The result showed that high expression of Lnc-SMIM20-1 was associated with less infiltration of B cell memory, T cell gamma delta, and myeloid dendritic resting cell. Besides, the expression of Lnc-SMIM20-1 was positively correlated with T cell CD4 memory activated (Supplementary Figure S5).

3.6. Construction and analyses of ceRNA network

Differentially expressed microRNAs and mRNAs were selected to construct the ceRNA network (Figure 6(a)). In the ceRNA network, we found that Lnc-SMIM20-1 bound to four microRNAs (miR-190b, miR-20b, miR-150, and let-7b) directly and was associated with 20 mRNAs, including SLC29A2, PRR34, NCR3LG1, ALDH1L2, FCRL2, IL2RB, ZNF831, ST8SLA6, TNFSF14, PCDH12, HIVEP3, PIK2, ENO2, ABCB6, TMEM144, RBFOX2, EPGN, CAMK2D, BICDL1, and PLEKHA6 (Figure 6(b), Supplementary Table S13). Subsequently, functional enrichment analysis was processed based on genes in the ceRNA showed that RAF/MAP kinase cascade was the main signaling pathway of the Lnc-SMIM20-1-related ceRNA (Figure 6(c)). Furthermore, five genes within the Lnc-SMIM20-1-related ceRNA network were selected by LASSO Cox regression to establish a prognostic model and the results showed that the prognostic model based on the Lnc-SMIM20-1-related ceRNA network performed great predictive efficiency in the TCGA-AML cohort (Supplementary Figure S6).

4. Discussion

LncRNAs are members of non-coding RNA longer than 200 nt in length, which is present in large numbers in the genome. LncRNAs have been thought to be only ‘transcriptional noise’ because many LncRNAs are expressed very low [30]. Along with increasing studies about the function of LncRNAs, LncRNAs were found to act as activators, decoys, guides, or scaffolds for their interacting protein, DNA, and RNA to participate in almost all the biological processes. Besides, LncRNAs may be involved in almost all human cancers by impacting cancer cell proliferation, invasion, and metastasis [31]. Hence, LncRNAs have the potential of bringing new insights in the
In this study, we found that a novel lncRNA Lnc-SMIM20-1 that had not been investigated before was up-regulated in AML, indicating inferior prognosis in both the TCGA adult AML cohort and the TARGET pediatric AML cohort. Besides, we also found that Lnc-SMIM20-1 was the most highly expressed in AML across 33 types of cancer, which indicated that Lnc-SMIM20-1 might be a unique indicator for the prognosis in AML.

Nowadays, some clinical characteristics have been clarified as prognostic indicators clinically. For example, RUNX1-RUNX1T1 is one of the most common genetic abnormalities in AML that have been clarified as a good prognostic factor for AML [32,33]. AML with inv(16)(p13.1q22) resulting in CBF8-MYH11 fusion is associated with a favorable prognosis [34]. Rearrangements of the KMT2A gene are found in 10% of adult AML which had poor prognosis [35]. Besides, multiple risk stratification systems have been established [36]. ELN stratification system is the latest and widely accepted system for AML [37]. Our result showed that AML patients classified as ‘favorable’ by ELN, AML with RUNX1-RUNX1T1 and AML with CBF8-MYH11 had lower expression of Lnc-SMIM20-1 while AML with KMT2A rearrangements had higher expression of Lnc-SMIM20-1, which further indicated that Lnc-SMIM20-1 could be a promising prognostic indicator in AML. Moreover, the multivariable analysis also indicated that Lnc-SMIM20-1 expression was an independent unfavorable prognosis factor in AML.

It is well acknowledged that chemotherapy and allo-HSCT are two major therapies for patients with AML. Chemotherapy is the main treatment for most of AML patients and allo-HSCT remains the first-line treatment for poor and very-poor-risk patients and yields a high rate of curability for AML [38,39]. Therefore, our further analysis showed that the prognosis impact of Lnc-SMIM20-1 cannot be eliminated by allo-HSCT, which indicated that the expression level of Lnc-SMIM20-1 could be a stable prognostic indicator for AML even after treatment with allo-HSCT.
To further explore the potential function of lnc-SMIM20-1 in AML, we performed functional analysis and we found that lnc-SMIM20-1 may participate in the PI3K-Akt signaling pathway, the JAK/STAT signaling pathway, and the NF-kappa B signaling pathway in AML. Numerous studies have demonstrated that the PI3K-Akt signaling pathway [40], the JAK/STAT signaling pathway [41], and the NF-kappa B signaling pathway [42] are frequently activated in acute myeloid leukemia (AML) patient blasts and strongly contribute to proliferation, survival, and drug resistance of these cells. These three signaling pathways...
have been the target for AML treatment and corresponding inhibitors, such as JAK inhibitors, can be the potential candidate combination therapies for precision medicine in AML [43]. Therefore, our finding suggested that the expression level of Lnc-SMIM20-1 might be the potential predictor for the response of some inhibitors, such as JAK inhibitors in AML.

In addition, immunotherapy is a promising therapy for cancer patients, but the clinical activity for immunotherapy in AML is very limited. Thus, many immunotherapy-based combined therapies have been developed. In this study, we found that Lnc-SMIM20-1 might participate in multiple immune-related signaling pathways, such as Th1 and Th2 cell differentiation and Th17 cell differentiation. Th1, Th2, and Th17 are three important CD4 + T helper cells in immune regulation. CD4+ helper T cells promote priming and both the effector functions and the memory functions of cytotoxic T lymphocytes (CTLs) and help CTLs to overcome negative regulation, so Enhancing CD4 + T cell help to improve CTL responses in cancer immunotherapy [44]. Therefore, our findings suggested that the expression of Lnc-SMIM20-1 might be associated with the activation of immune-related pathways, which further indicated that Lnc-SMIM20-1 might be the potential target for combined therapy helping the improvement of the clinical application of immunotherapy in AML.

LncRNAs regulate gene expression through various mechanisms including sponging microRNA. To further explore how Lnc-SMIM20-1 plays its role in AML, we constructed the Lnc-SMIM20-1 centric ceRNA network. We found that Lnc-SMIM20-1 interacted with four microRNAs (miR-190b, miR-20b, miR-150, let-7b) directly and was associated with 20 mRNAs, such as RBFOX2, ZNF831, PTK2, and HIVEP3. A study has demonstrated that high expression of miR-20b was associated with poor prognosis in AML [45]. Numerous studies have suggested that downregulation of miR-150 was required for leukemogenesis [46,47]. Additionally, a study has revealed that let-7b played important role in t(8;21) leukemogenesis and maintenance by downregulating AML1-ETO oncoprotein expression [48]. Although miR-190b and the mRNAs in the network have been reported as the promoter for the proliferation and progression in multiple tumors, the role of these genes in AML has not been reported yet. Our findings indicated that these genes might also play important role in AML, which provided ideas for further study of the potential role of these genes in AML. Besides, according to the ceRNA hypothesis, our findings indicated that Lnc-SMIM20-1 might be the target for miR-190b, miR-20b, miR-150, and let-7b, which provided new clues for the molecular mechanism of these microRNAs in AML as well.

Even so, there are some limitations in our study. For example, this study is retrospective research and still needs prospective analysis. In addition, we preliminary unveiled the potential mechanism of Lnc-SMIM20-1 in AML, but the biological mechanism has not been fully elucidated. Therefore, these shortcomings may be overcome through external experiments in the future. In conclusion, we found that LncRNA Lnc-SMIM20-1 that was highly expressed in AML can be a potential biomarker of the prognosis of AML. We also found that Lnc-SMIM20-1 was associated with clinically actionable signaling pathways and immune-related pathways, providing new ideas for the diagnosis and treatment of AML.
5. Conclusion
High expression of Lnc-SMIM20-1 was found in AML and associated with poor prognosis in AML. The prognostic impact of Lnc-SMIM20-1 cannot be overridden by allo-HCT, which provides potential new directions for AML therapy.

Acknowledgments
The authors would like to acknowledge the Second Clinical Medical College of Yangtze University, Taikang Tongji (Wuhan) hospital, and the Seventh Affiliated Hospital of Sun Yat-sen University, for providing support during the project.

Funding
The authors have no funding to report.

Declaration of interest
The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Reviewer disclosures
Peer reviewers on this manuscript have received no honorarium from Expert Review of Anticancer Therapy for their review work but have no other relevant financial relationships to disclose.

Author contribution statement
Study concept and design: Sha Liu, Ziyi Sun, Xiaofen Pan, and Shan Huang;
Acquisition of data: Sha Liu, Ziyi Sun, Xiaofen Pan and Shan Huang;
Analysis and interpretation of data: Sha Liu, Ziyi Sun, Mengyuan Zhu, Minling Liu, Min Wei, Shan Huang and Xiaofen Pan;
Drafting of the manuscript: Sha Liu and Ziyi Sun;
Statistical analysis: Sha Liu and Ziyi Sun;
Administrative, technical or material support: Sha Liu, Ziyi Sun, Mengyuan Zhu, Minling Liu, Min Wei, Shan Huang and Xiaofen Pan;
Supervision: Shan Huang and Xiaofen Pan.
All authors read and approved the final manuscript.

Data availability statement
The datasets generated and/or analyzed during the current study are available in The Cancer Genome Atlas (TCGA), Therapeutically Applicable Research to Generate Effective Treatments (TARGET), and Genotype-Tissue Expression Project (GTEx) repository, [https://cancergenome.nih.gov/; https://cancer.gov/projects/tARGET; https://www.genome.gov/Funded-Programs-Projects/Genotype-Tissue-Expression-Project]

References
Papers of special note have been highlighted as either of interest (+) or of considerable interest (+++) to readers.
1. DiNardo C, Lachowicz C. Acute myeloid leukemia: from mutation profiling to treatment decisions. Curr Hematol Malig Rep. 2019 Oct:14 (5):386–394.
2. Hindley A, Catherwood MA, McMullin MF, et al. Significance of NPM1 gene mutations in AML. Int J Mol Sci. 2021 Sep 17;22(18):10040.
3. Heath EM, Cham SM, Minden MD, et al. Biological and clinical consequences of NPM1 mutations in AML. Leukemia. 2017 Apr;31(4):798–807.
4. Daver N, Schlenk RF, Russell NH, et al. Targeting FLT3 mutations in AML: review of current knowledge and evidence. Leukemia. 2019 Feb;33(2):299–312.
5. Laguna-Rangel FA, Chavez-Valencia V, FLT3-ITD and its current role in acute myeloid leukaemia. Med Oncol. 2017 Jun;34(6):114.
6. Mannelli F, Ponziani V, Bencini S, et al. CEBPA-double-mutated acute myeloid leukemia displays a unique phenotypic profile: a reliable screening method and insight into biological features. Haematologica. 2017 Mar;102(3):529–540.
7. Wouters BJ, Lowenberg B, Erpelink-Verschueren CA, et al. Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome. Blood. 2009 Mar 26;113(13):3088–3091.
8. Tang JL, Hou HA, Chen CY, et al. AML1/RUNX1 mutations in 470 adult patients with de novo acute myeloid leukemia: prognostic implication and interaction with other gene alterations. Blood. 2009 Dec 17;114(26):5352–5361.
9. Greif PA, Konstandin NP, Metzeler KH, et al. RUNX1 mutations in cytogenetically normal acute myeloid leukemia are associated with a poor prognosis and up-regulation of lymphoid genes. Haematologica. 2012 Dec;97(12):1909–1915.
10. Schnittger S, Dicker F, Kern W, et al. RUNX1 mutations are frequent in de novo AML with noncomplex karyotype and confer an unfavorable prognosis. Blood. 2011 Feb 24;117(8):2348–2357.
11. Garzon R, Vaino S, and Papaioannou D, et al. Expression and prognostic impact of IncRNAs in acute myeloid leukemia. Proc Natl Acad Sci U S A. 2014 Dec 30;111(52):18679–18684.
+- LncRNA expression in AML is closely associated with recurrent mutations. A small subset of IncRNAs is correlated strongly with treatment response and survival.
12. Ghetti M, Vannini I, Storlazzi CT, et al. Linear and circular PVT1 in hematological malignancies and immune response: two faces of the same coin. Mol Cancer. 2020 Mar 30;19(1):69.
13. Tolomeo D, Agostini A, Vischi G, et al. PVT1: a long non-coding RNA recurrently involved in neoplasia-associated fusion transcripts. Gene. 2021 May 5;779:145497.
14. Chen L, Wang W, Cao L, et al. Long non-coding RNA CAT1 acts as a competing endogenous RNA to regulate cell growth and differentiation in acute myeloid leukemia. Mol Cells. 2016 Apr 30;35(4):330–336.
15. Wu S, Zheng C, Chen S, et al. Overexpression of long non-coding RNA HOTAIR predicts a poor prognosis in patients with acute myeloid leukemia. Oncol Lett. 2015 Oct:10(4):2410–2414.
16. Hao S, Shao Z. HOTAIR is upregulated in acute myeloid leukemia and that indicates a poor prognosis. Int J Clin Exp Pathol. 2015;8 (6):7223–7228.
17. Sheng XF, Hong LL, Li H, et al. Long non-coding RNA MALAT1 modulate cell migration, proliferation and apoptosis by sponging microRNA-146a to regulate CXCR4 expression in acute myeloid leukemia. Hematol. 2021 Dec;26(1):43–52.
18. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11(10):R106.
19. Cheng Y, Su Y, Wang S, et al. Identification of circRNA-lncRNA-miRNA-mRNA competitive endogenous RNA network as novel prognostic markers for acute myeloid leukemia. Genes (Basel). 2020 Jul 31;11(8):868.
20. Liang Q, Zhang L, Wang W, et al. High expression of DC-Stamp gene predicts adverse outcomes in AML. Front Genet. 2022;13:876689.
21. Sun X, Huang S, Wang X, et al. CD300A promotes tumor progression by PECAM1, ADCC7 and AKT pathway in acute myeloid leukemia. Oncotarget. 2018 Jun 8;9(44):27574–27584.
22. Tang Z, Li C, Kang B, et al. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. 2017 Jul 3;45(W1):W98–W102.
23. Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012 May;16(5):284–287.
This review summarizes the characteristics of IncRNAs, including their roles, functions, and working mechanisms, describes methods for identifying and annotating IncRNAs, and discusses future opportunities for IncRNA-based therapies using antisense oligonucleotides.