Received: 26 April 2022 | Revised: 1 July 2022 | Accepted: 22 August 2022

DOI: 10.1002/cncr.34481

ORIGINAL ARTICLE

Subclones of bone marrow CD34$^+$ cells in acute myeloid leukemia at diagnosis confer responses of patients to induction chemotherapy

Ruinan Jia MS[1] | Min Ji MD[1] | Guosheng Li MS[1,2] | Yuan Xia MS[1] | Shouhui Guo PhD[3,4] | Peng Li MD[1] | Yanping Sun MD[1] | Fei Lu MD[1] | Jingru Zhang MD[1] | Shaolei Zang MD[1] | Shuxin Yan MS[1] | Jingjing Ye MD[1,2] | Fuzhong Xue PhD[3,4] | Daoxin Ma PhD[1,2] | Tao Sun PhD[1,2] | Chunyan Ji MD, PhD[1,2]

[1]Department of Hematology, Qilu Hospital, Cheeloo College of Medicine, Shandong University, Jinan, China

[2]Shandong Key Laboratory of Immunohematology, Qilu Hospital, Shandong University, Jinan, China

[3]Department of Biostatistics, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan, China

[4]National Institute of Health Data Science of China, Shandong University, Jinan, China

Correspondence

Tao Sun and Chunyan Ji, Department of Hematology, Qilu Hospital, Cheeloo College of Medicine, Shandong University, Jinan, China.
Email: taosun@sdu.edu.cn and jichunyan@sdu.edu.cn

Funding information

Natural Science Foundation of Shandong Province, Grant/Award Numbers: ZR2020KH016, ZR2020MH118, ZR2020MH119; National Natural Science Foundation of China, Grant/Award Numbers: 81700143, 81770159, 81873425, 82070160; Taishan Scholar Foundation of Shandong Province, Grant/Award Numbers: tspd20210321, tsqn201812132

Abstract

Background: Acute myeloid leukemia (AML) is a hematopoietic malignancy with a prognosis that varies with genetic heterogeneity of hematopoietic stem/progenitor cells (HSPCs). Induction chemotherapy with cytarabine and anthracycline has been the standard care for newly diagnosed AML, but about 30% of patients have no response to this regimen. The resistance mechanisms require deeper understanding.

Methods: In our study, using single-cell RNA sequencing, we analyzed the heterogeneity of bone marrow CD34$^+$ cells from newly diagnosed patients with AML who were then divided into sensitive and resistant groups according to their responses to induction chemotherapy with cytarabine and anthracycline. We verified our findings by TCGA database, GEO datasets, and multiparameter flow cytometry.

Results: We established a landscape for AML CD34$^+$ cells and identified HSPC types based on the lineage signature genes. Interestingly, we found a cell population with *CRIP1*^{high}*LGALS1*^{high}*S100As*^{high} showing features of granulocyte-monocyte progenitors was associated with poor prognosis of AML. And two cell populations marked by CD34$^+$CD52$^+$ or CD34$^+$CD74$^+$DAP12$^+$ were related to good response to induction therapy, showing characteristics of hematopoietic stem cells.

Conclusion: Our study indicates the subclones of CD34$^+$ cells confers for outcomes of AML and provides biomarkers to predict the response of patients with AML to induction chemotherapy.

KEYWORDS

Acute myeloid leukemia, CD34$^+$cell, Heterogeneity, Induction chemotherapy, Single-cell RNA-sequencing

Ruinan Jia and Min Ji contributed equally to the work.

INTRODUCTION

Acute myeloid leukemia (AML) encompasses a group of heterogeneous disorders characterized by the expansion of malignant clones of hematopoietic progenitor cells blocked at various stages of differentiation.[1] Remission induction chemotherapy with cytarabine and anthracycline has been a standard treatment for newly diagnosed AML for more than 30 years.[2] Although 70% of patients with newly diagnosed AML attain morphologic complete remission (CR) with intensive induction chemotherapy, approximately 30% of adults with AML are not sensitive to chemotherapy, and at least 50% of those who achieve remission will relapse.[3–5] The different outcomes of AML patients are because AML is a heterogeneous and molecularly complex disease with variable hematologic phenotypes.[6]

AML hematopoietic stem/progenitor cells sustain the disease and display stem cell properties, such as self-renewal, quiescence, and heterogeneity.[7] Therefore, the AML hematopoietic stem/progenitor cells with heterogeneous properties enable their ability to generate the heterogeneity of intra- or interpatient, and then fuel different responses to induction chemotherapy during disease progression.[8] Identification of the heterogeneity in AML hematopoietic stem/progenitor cells is of vital importance for prediction of AML prognosis.

Recently, single-cell RNA-sequencing (scRNA-seq) technologies have matured such that one can sequence and analyze thousands of cells per tumor. At this scale, it can derive significant insights into the cellular heterogeneity, characteristics of the molecular diversity, and the biological features that distinguish different cell subpopulations.[9–11] Savas et al. found that $CD8^+$ tissue resident memory T cells contributed to breast cancer immunosurveillance and were the key targets of modulation by immune checkpoint inhibition through scRNA-seq of T cells isolated from human breast cancers.[12] Mathys et al. discovered disease-associated cellular subpopulations in Alzheimer's disease.[13] Witkowski et al. uncovered a role for nonclassical monocytes subcluster in bone marrow microenvironment supporting acute B lymphoblastic leukemia progression and treatment evasion.[14]

Here, we adapted 10X Genomics scRNA-seq technology[15] to acquire transcriptional data for thousands of single cells from bone marrow $CD34^+$ cells of newly diagnosed patients with AML. We showed a comprehensive and heterogeneous map of the $CD34^+$ cells in patients with AML and healthy controls. We leveraged transcriptome wide features to distinguish malignant-like cells from normal-like cells, and malignant-like clusters from each other via several analytical strategies. Furthermore, by comparing sensitive patients with AML with resistant patients, we discovered a poor-prognosis–related subpopulation with specific gene signatures in $CD34^+$ cells. Moreover, we anchored two subpopulations with specific markers in AML $CD34^+$ cells that were associated with good response of patients to induction chemotherapy. Our discovery created a chance for prognosis or prediction in newly diagnosed patients with AML.

MATERIALS AND METHODS

Full descriptions of experimental procedures and analytical methods are in the Supplementary information.

Patients and samples

Bone marrow samples were obtained from patients with newly diagnosed AML during routine diagnostic assessments before treatment at Qilu Hospital, Shandong University, Jinan, China. Informed consent was obtained in accordance with the Declaration of Helsinki. All primary samples procedures were reviewed and approved by the Medical Ethics Committee of Qilu Hospital of Shandong University. Patients were treated with induction chemotherapy according to the guideline for newly diagnosed AML in adults of China. Responses were assessed in accordance with the International Working Group Criteria[16] for AML. CR was defined as bone marrow blasts ≤5%, no circulating blasts, with complete blood count recovery (neutrophil count ≥1000/μL and platelet count ≥100 000/μL). Partial remission required all the hematologic values of CR with a decrease of ≥50% in the percentage of bone marrow blasts to 5% to 25%. No remission (NR) was defined as not up to the criteria of partial remission and CR. The 2017 European Leukemia Net (ELN) risk stratification by genetics was used to evaluate patients.

Single-cell RNA library construction and sequencing

We used the Single Cell 3′ Reagent kit (V2) and the Chromium instrument to prepare individually barcode scRNA-seq libraries following the manufacturer's protocols (10× Genomics). Sequencing with dual indexing was conducted on an Illumina NovaSeq PE150 machine using the 150-cycle High Output kit. The scRNA-seq data for normal $CD34^+$ cells from four healthy donors were obtained from the GEO database (GSE133181).[17] The Cell Ranger Single Cell Software Suite was used to perform sample demultiplexing, barcode processing, and single-cell 3' gene counting. Further analysis was performed using the Seurat R package and Loupe software. The integration analysis was used by CCA + MNN to correct the batch effect in Seurat. Genes with pct.1/pct.2 greater than 1.5 in the top 10 significantly high-expression genes were chosen to mark the cluster.

Statistical analysis

Differences between two groups were analyzed using an unpaired Student t test. Data are presented as mean ± SD. The Fisher exact test was used for comparing the CR rates of patients. $p < .05$ was considered statistically significant.

RESULTS

Identification of cell populations in CD34$^+$ cells from newly diagnosed patients with AML

To clarify cell diversity of CD34$^+$ cells in AML, we performed scRNA-seq on the Chromium platform (10× Genomics) of primary CD34$^+$ cells that were magnetic activated cell sorting–purified from the bone marrow of six newly diagnosed patients with AML before treatment. Normal CD34$^+$ cells from four healthy donors were used as controls; their scRNA-seq data were obtained from the GEO database. The patients' clinical information is shown in Table 1. We acquired high-quality data of 60 402 CD34$^+$ cells from patients with AML and healthy controls and analyzed the data by Seurat using its anchoring-based integration method to account for technical and biological variance between individual samples. This was followed by projection dimensionality reduction and t-distributed stochastic neighbor embedding (t-SNE) for visualization (Figure 1A).

We then performed extensive unbiased clustering of all the CD34$^+$ cells to identify transcriptionally distinct cell clusters. As a result, CD34$^+$ cells were classified into six hematopoietic stem/progenitor cell types (hematopoietic stem cell [HSC], multilymphoid progenitor [MLP], megakaryocyte-erythroid progenitor [MEP], granulocyte-monocyte progenitor [GMP], pro-B cell, and earliest thymic progenitors [ETP]) and 17 clusters (HSC-1 and 2, MLP, MEP-1 through 3, GMP-1 through 7, pro-B cell 1 and 2, and ETP-1 and 2) based on comparing cluster-specific genes with reported lineage signature genes (Figure 1B–D, Table S1).[18]

CD34$^+$ cells of patients with AML at diagnosis enriches GMP population

We observed obvious differences in the composition of the stem/progenitor lineage between newly diagnosed patients with AML and healthy controls. The results showed that the proportion of GMP-like population in patients with AML was significantly higher than that of GMP in healthy controls, whereas the proportion of HSC and pro-B populations were significantly lower (Figure S1). We then defined the cluster to be normal-like if the high percentage (>60%) of cells in the cluster was from healthy controls; the remaining clusters were defined as malignant-like. As a result, the clusters GMP-1 through 7, MEP-1, MLP, and ETP-1 were classified as malignant-like clusters, and the others were classified as normal-like clusters (Figure 1E). Obviously, CD34$^+$ cells from patients with AML were mainly composed of malignant-like clusters (Figure S2). Copy number variations were also highly consistent with the result and malignant-like cells were grouped with extensive copy number

TABLE 1 Clinical information of AML patients and healthy controls for scRNA-seq

	AML01	AML02	AML03	AML05	AML06	AML07
Sex	Female	Female	Female	Male	Male	Male
Age, y	32	63	36	37	61	20
WBC count, $\times 10^9$/L	2.44	11.77	21.94	86.33	0.68	83
Hemoglobin, g/L	115	92	84	89	67	127
Platelet count, $\times 10^9$/L	70	42	6	13	35	21
Peripheral blood blasts, %	92	50	91	50	66	66
Bone marrow blasts, %	30	45	70	42	77	90
FAB classification	NA	M4b	NA	M2	M5	M5
Fusion gene	*MLL-(AF17/AF1q/AF1p/AFX/SEPT6)*	*CBFB-MYH11*	*MLL-AF9*	Negative	Negative	Negative
Mutation	Negative	Negative	*NRAS*	*CEBPA* *KIT*	*IDH1* *WT1* *NRAS*	*CEBPA* *NRAS*
Chromosome karyotyping	46, XX[4]	NA	NA	46, XY[3]	NA	46, XY[20]
ELN risk stratification	Adverse	Favorable	Intermediate	Favorable	Intermediate	Favorable
Induction chemotherapy	IA	IA	IA	IA	DA	HAA
Response	NR	CR	NR	CR	CR	CR
	CTL01	**CTL02**	**CTL03**			**CTL04**
Age, y	53	21	30			41

Abbreviations: AML, acute myeloid leukemia; CR, complete remission; DA, daunorubicin + cytarabine; ELN, European LeukemiaNet; FAB, French-American-British; HAA, homoharringtonine + aclarubicin + cytarabine; IA, idarubicin + cytarabine; NA, not available; NR, none remission; WBC, white blood cell.

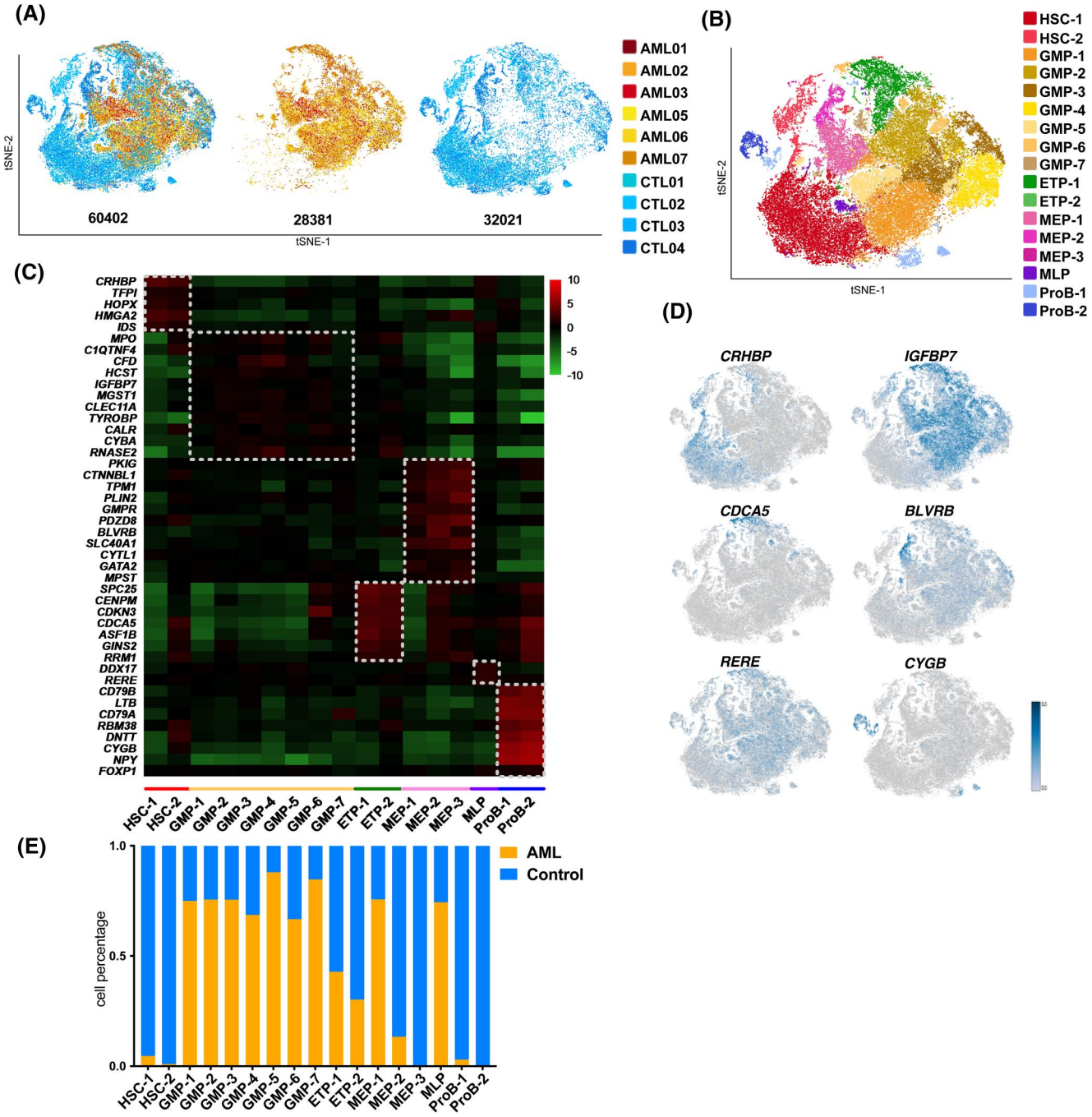

FIGURE 1 Identification of cell populations in CD34$^+$ cells from newly diagnosed patients with acute myeloid leukemia (AML). (A) t-distributed stochastic neighbor embedding (t-SNE) visualization of 60 402 cells from six newly diagnosed patients with AML and four healthy controls (CTL). The numbers in panel A represent cell counts. (B) Marker-based cell type identification analysis allowed prediction of 6 hematopoietic stem/progenitor cell types across 17 clusters. (C) Gene expression heatmap of the cell-type–specific marker genes measured. (D) Expression levels of the cell-type–specific marker genes overlaid on the t-SNE representation. (E) Histogram showing percentage of each cluster between patients with AML and controls after normalizing baseline to 100%. $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$, $^{****}p < .0001$, $^{*****}p < .00001$, $^{******}p < .000001$

variations across the whole genome (Figure S3). Gene set enrichment analysis using the KEGG and GO pathways showed that gene expression profiles of malignant-like clusters enriched for cell cycle, oxidative phosphorylation, cell apoptosis, and cancer-related pathways (Figures S4A and B).

Heterogeneity of malignant-like clusters in CD34$^+$ cells of AML

We hypothesized that the distinct response of patients with AML to chemotherapy was due to the heterogeneity of CD34$^+$ cells. To

address this, we divided the six patients with AML into two groups according to the response of patients with AML to the standard induction therapy with cytarabine and anthracycline. AML02, AML05, AML06, and AML07 achieved CR after the first course of induction therapy and were assigned to the "sensitive" group, whereas AML01 and AML03 were assigned to the "resistant" group because they did not respond to the first course of induction therapy (Figure 2A and Table 1). The six patients with AML were in different cytogenetic risk categories. We then analyzed the single-cell data of $CD34^+$ cells in the two groups. We found that both groups of $CD34^+$ cells were composed mainly of the 10 malignant-like clusters which belonged to four cell types (GMP, MEP, MLP, and ETP) (Figure 2A). The proportion of each cell cluster did not show significantly different between the sensitive and resistant groups (Figure 2B).

It seemed the modest increase in the proportion of malignant-like clusters was observed in the resistant group, when compared with the sensitive group (Figure 2C). However, there was no significant difference in the proportion of cells derived from the two groups in each malignant-like cluster (Figure 2D). Furthermore, to search for the differences between AML-sensitive and AML-resistant groups, we compared the transcriptional expression and enrichment genes in each malignant-like cluster. The results of KEGG, GO, and GSEA enrichment showed that the malignant-like clusters GMP-1, -2, -3, -5, and -7 and MLP in the AML-resistant group obviously enriched for AML, cell cycle, oxidative phosphorylation, and uncontrolled transcription compared with the sensitive group (Figure 2E–G). In addition, these clusters in the AML-resistant group expressed higher levels of genes that were reported to be associated with poor prognosis[19] than in the AML-sensitive group (Figure S5). The results indicated that although the composition of $CD34^+$ cells was almost the same between the AML-sensitive and AML-resistant groups, the gene expression profiles were quite different in the GMP-1, -2, -3, -5, and -7 and MLP clusters, suggesting these clusters were probably composed of different subclusters, which may confer the distinct response of AML patients to induction chemotherapy.

Subclusters with specific gene signatures (*CRIP1*^{high}*LGALS1*^{high}*S100As*^{high}) are associated with poor prognosis of AML

To figure out the specific subcluster of $CD34^+$ cells that are associated with no response to treatment of AML, we subgrouped these six cell clusters (GMP-1, -2, -3, -5, and -7 and MLP) and compared the proportion of subclusters in the AML-sensitive, AML-resistant, and control groups (Figure 3A).

In the GMP-1 cluster, we identified nine transcriptionally distinct cell subclusters, and the proportion of subcluster GMP-1-0 was markedly higher in the AML-resistant group than in the AML-sensitive group and controls (Figure 3A). Specific gene expression data analysis indicated the highly expressing markers of subcluster GMP-1-0, such as *CRIP1*, *S100A10*, *S100A6*, and *LGALS1*, which were all significantly related to poor prognosis in cancers (Figure 3B). Interestingly, in the GMP-3 cluster, we found the proportion of subcluster GMP-3-0 was also higher in the AML-resistant group than in the sensitive group and its specific gene signatures were almost the same as the subcluster GMP-1-0 (Figure 3A and B). Furthermore, we performed an enrichment analysis of transcriptional signatures in these two subclusters and found that the associated upregulated genes were involved in cell adhesion, cell apoptosis, and S100 protein binding (Figure 3C).

To further clarify the significance of these subclusters with specific gene signatures (*CRIP1*^{high}*LGALS1*^{high}*S100As*^{high}) for AML, we took advantage of RNA-seq data from 134 patients with AML in the TCGA database and calculated scores of the specific genes of these subclusters, then divided the 134 patients into two groups based on the scores (the high-score group [$n = 67$] and the low-score group [$n = 67$]; Figure 3D and Table S3). We noted a significantly inferior overall survival in the high-score group, when compared with the low-score group, suggesting that the gene profile of the subclusters with *CRIP1*^{high}*LGALS1*^{high}*S100As*^{high} plays a pernicious role in AML survival (Figure 3E). Therefore, we speculated that GMP-1-0 and GMP-3-0 may be the specific subpopulations associated with poor responses to induction chemotherapy in newly diagnosed patients with AML.

Another way of clustering also confirms the existence of the subcluster with specific gene signatures (*CRIP1*^{high}*LGALS1*^{high}*S100As*^{high})

To exclude the interference of healthy controls, we removed healthy controls and reanalyzed the data of $CD34^+$ cells from patients with AML only, which contained a total of 28 381 single-cell transcriptomes from AML-resistant ($n = 8935$) and AML-sensitive ($n = 19\,446$) groups (Figure 4A). The sample viability of AML01, 02, 03, 05, 06, and 07 was 96%, 99%, 95%, 84%, 88%, and 90%, respectively, which did not significantly correlate with the mean number of detected genes ($p = .6631$), the number of unique molecular identifiers ($p = .7495$), or the fraction of mitochondrial RNA transcripts ($p = .2356$). As shown in Figure 4B, $CD34^+$ cells from patients with AML were partitioned into 19 clusters. After analysis of the composition in each group, we observed increases in the proportion of clusters 2, 3, 5, and 6 in the AML-resistant group (Figure 4C).

In clusters 2, 3, 5, and 6, the high proportion (>60%) of cells was from the AML-resistant group (Figure 4D). Through matching the marker genes of clusters 2, 3, 5, and 6 with lineage signature genes, we found clusters 2, 5, and 6 were similar to GMP and cluster 3 was similar to MEP (Figure 4E). The cell types of these clusters were defined as GMP-like and MEP-like. Figure 4F shows the specific marker genes that were remarkably upregulated in clusters compared with the other clusters. Coincidentally, the genes program in cluster 2 perfectly matched the marker genes of the subcluster GMP-1-0 and GMP-3-0, which were related to poor survival of AML, suggesting cluster 2 was the same cluster as those two subclusters. To understand the function of the genes program in cluster 2, we

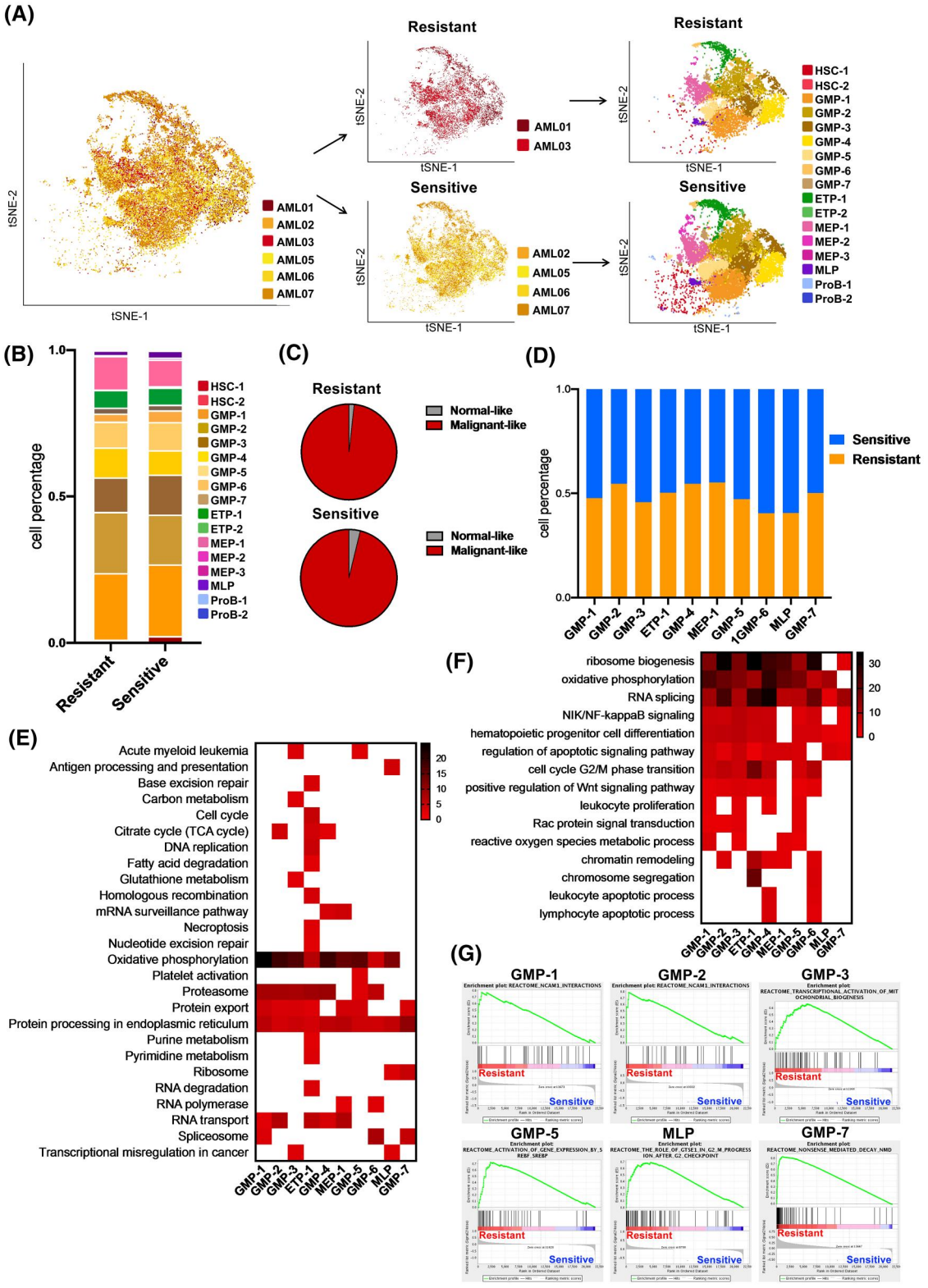

FIGURE 2 Heterogeneity of malignant-like clusters in CD34$^+$ cells of AML. (A) t-SNE visualization of CD34$^+$ cells from six newly diagnosed patients with AML (left), resistant and sensitive groups (medium), and marker-based cell type identification analysis including six cell types from 17 clusters (right). (B) Histogram showing cell percentage of 17 clusters between resistant and sensitive groups after normalizing baseline to 100%. (C) Pie chart showing cell percentage of malignant-like and normal-like clusters between the resistant and sensitive groups. (D) Histogram showing cell percentage of resistant and sensitive groups in 10 malignant-like clusters after normalizing baseline to 100%. (E) Significantly KEGG pathways enrichment analysis of differentially expressed genes (DEGs) comparing resistant with sensitive groups in each malignant-like cluster. (F) Significantly GO pathways enrichment analysis of DEGs comparing resistant with sensitive groups in each malignant-like cluster. (G) Significantly gene set enrichment analysis (GSEA) of genes comparing resistant with sensitive groups in GMP-1, -2, -3, -5, and -7 and MLP clusters. AML indicates acute myeloid leukemia; t-SNE, t-distributed stochastic neighbor embedding

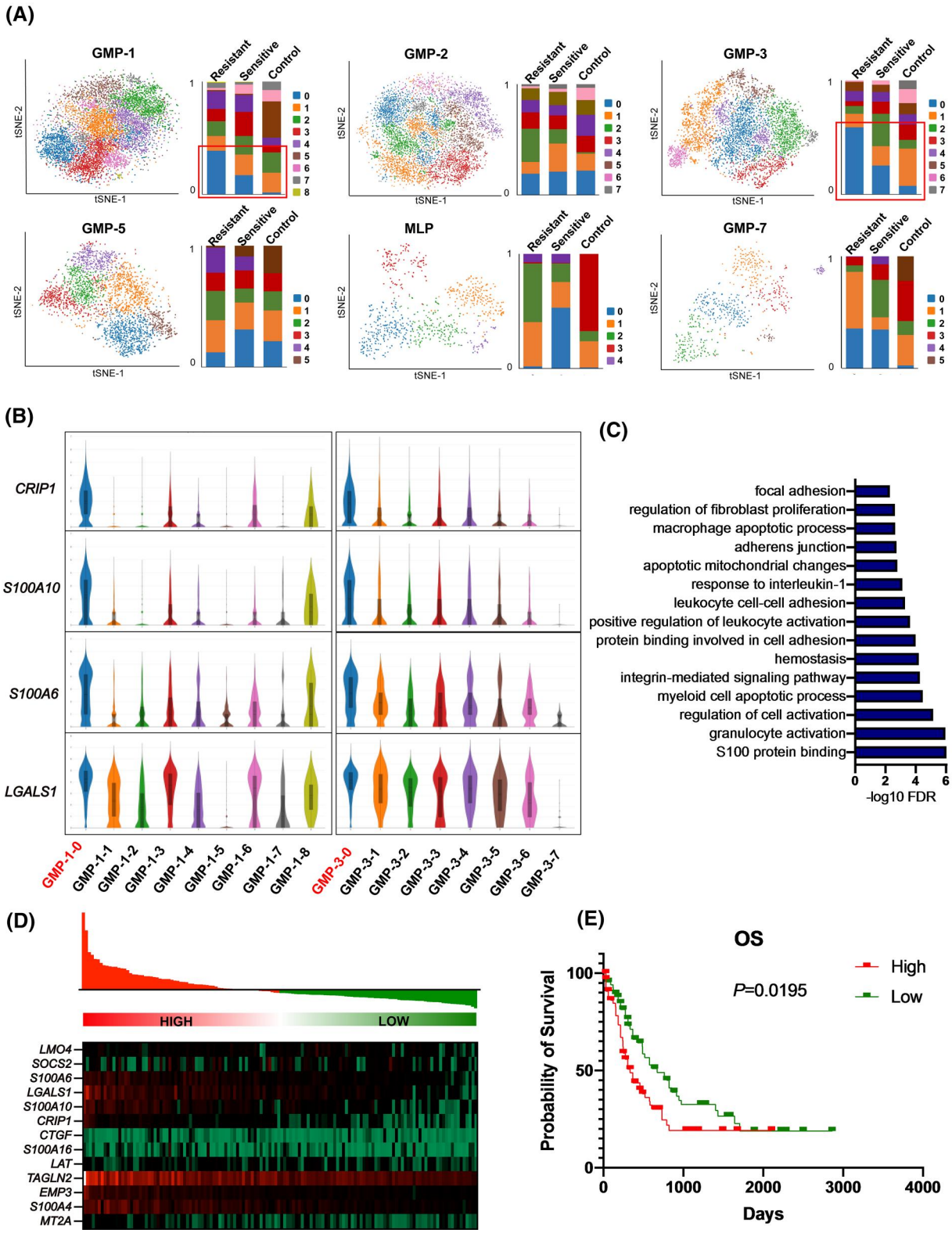

FIGURE 3 Subclusters of CD34^{+} cells associated with poor prognosis in AML. (A) t-SNE visualization of the subgrouped results in six alternative clusters (left). Boxplot showing cell percentage of corresponding subclusters between resistant, sensitive, and control groups after normalizing baseline to 100% (right). (B) Violin plot showing relative expression of top marker genes in GMP-1-0 and GMP-3-0. (C) Bar graph showing significantly KEGG pathways enrichment analysis of molecular signatures in GMP-1-0 and GMP-3-0. (D) Heatmap showing expression of signature genes (rows) in GMP-1-0 and GMP-3-0 across 134 bulk AML profiles from TCGA database (columns). (E) Kaplan–Meier curve showing the survival of 134 patients with AML with high and low scores of the signature genes. *p* value calculated by log-rank test. AML indicates acute myeloid leukemia; t-SNE, t-distributed stochastic neighbor embedding

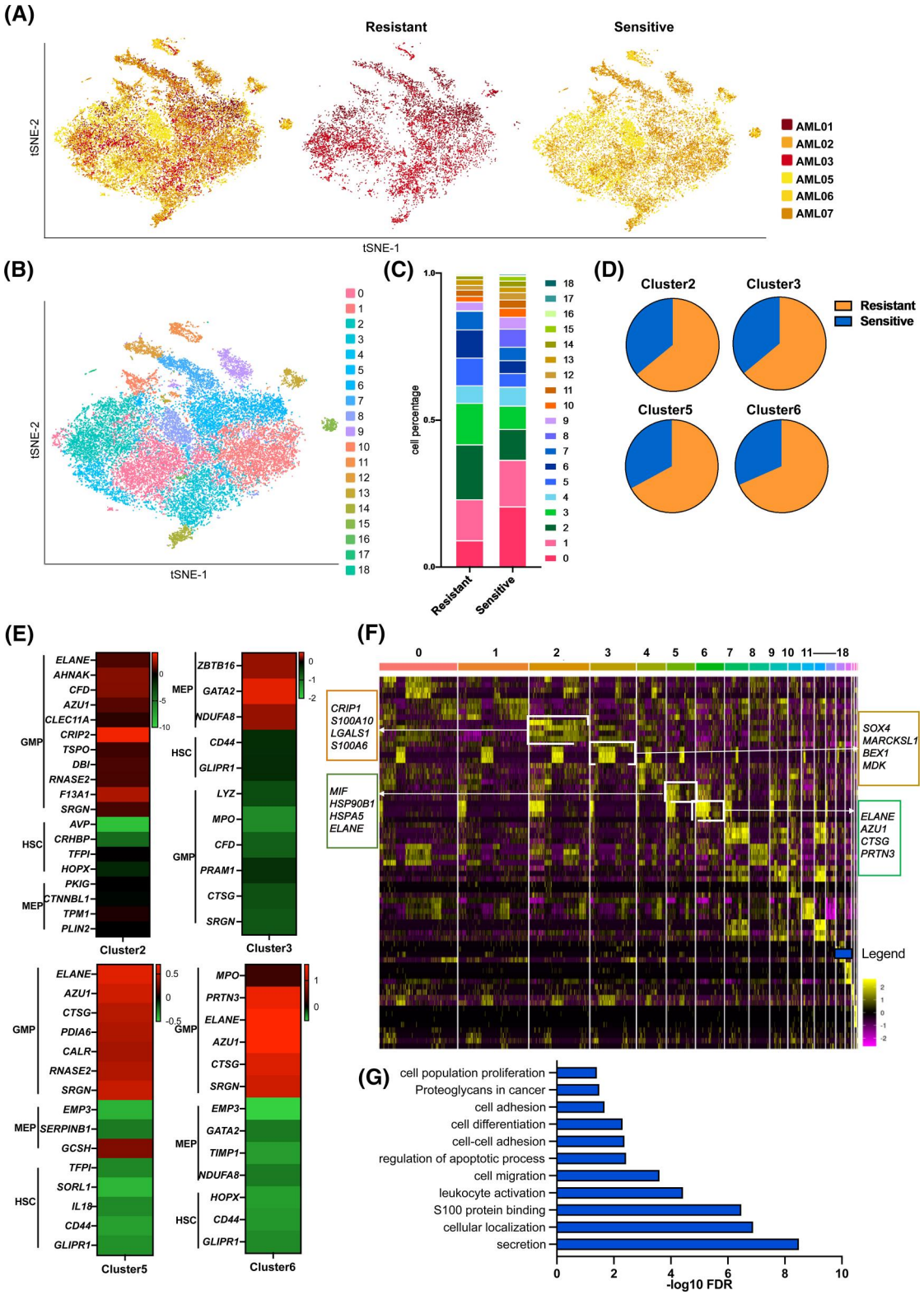

FIGURE 4 Regrouping of AML CD34^{+} cells and reappearance of the *CRIP1*^{high}*LGALS1*^{high}*S100As*^{high} cluster. (A) t-SNE visualization of 28 381 cells from six newly diagnosed patients with AML. Total (left), resistant group (medium), and sensitive group (right). (B) t-SNE visualization of the 28 381 cells regrouped result into 19 clusters. (C) Histogram showing cell percentage of 19 clusters between resistant and sensitive groups after normalizing baseline to 100%. (D) Pie chart showing cell percentage between resistant and sensitive groups in clusters 2, 3, 5, and 6. (E) Gene expression heatmap of the cell-type–specific marker genes in clusters 2, 3, 5, and 6. (F) Heatmap showing relative expression of top marker genes in 19 clusters and emphasized marker genes in clusters 2, 3, 5, and 6. (G) Bar graph showing significantly KEGG pathways enrichment analysis of molecular signatures in cluster 2. AML indicates acute myeloid leukemia; t-SNE, t-distributed stochastic neighbor embedding

performed gene expression enrichment analysis, revealing the upregulated genes were involved in cell adhesion, cell proliferation, and S100 protein (Figure 4G). Thus, the specific genes program of cluster 2, such as *CRIP1*, *S100A10*, *LGALS1*, and *S100A6*, may be a predictive factor for poor prognosis of AML.

Coincidentally, because the two patients (AML01 and AML03) in the resistant group both had MLL rearrangements, we tried to figure out whether the cluster (*CRIP1*^{high}*LGALS1*^{high}*S100As*^{high}) associated with poor responses to induction therapy had MLL rearrangement–related gene signatures. We analyzed MLL rearrangement–related genes in the cluster, but no prominent enrichment of MLL-related genes was observed in the cluster featured with *CRIP1*^{high}*LGALS1*^{high}*S100As*^{high}.

Cell populations associated with good response to induction chemotherapy in AML exhibit characteristics of HSC

Interestingly, when we compared the proportion of each cluster of cells in AML-resistant and AML-sensitive groups, we found that there was a significant increase in the proportion of cluster 0 and 8 in the sensitive group (Figure 5A and B). Through matching the marker genes of cluster 0 or 8 with lineage signature genes, we found cluster 0 and 8 exhibited characteristics of HSC (Figure 5C). The cell type of cluster 0 and 8 was defined as HSC-like. To understand the function of the two clusters, we performed gene expression enrichment analysis, which revealed that the upregulated genes in the two clusters were involved in helper T cell differentiation and immune response (Figure 5D).

We also recognized the surface markers of cluster 0 (CD52) and 8 (CD74 and DAP12) because they were in the top of the list of significantly high expression genes (Figure 6A) and were convenient for detection by flow cytometry in the clinic. To verify whether the distinct response to induction chemotherapy was due to the differences of clusters in CD34$^+$ cells, we used multiparameter flow cytometry to detect the HSC-like clusters in newly diagnosed patients with AML. Bone marrow mononuclear cells were isolated from samples of 32 newly diagnosed patients with AML before treatment. Seventeen patients achieved CR after induction chemotherapy and were assigned to the "sensitive" group, whereas the other 15 patients showed NR after induction chemotherapy and were assigned to the "resistant" group (Table S2). The result showed that the proportion of CD52$^+$ cells and CD74$^+$DAP12$^+$ cells in the CD34$^+$ population at diagnosis was significantly higher in patients with AML with CR (sensitive group) than those with NR (resistant group), which was consistent with the findings of scRNA-seq (Figure 6B and C). In patients with CD52$^+$ cells >10% of the CD34$^+$ population, the CR rate after induction chemotherapy was significantly higher than that in patients with CD52$^+$ cells ≤10% of the CD34$^+$ population (100% vs 11.8%, $p < .0001$). The CR rate was also higher in patients with CD74$^+$DAP12$^+$ cells >30% than ≤30% of the CD34$^+$ population (100% vs 37.5%, $p = .0023$). Moreover, we analyzed the characteristics of the 32 AML patients. We discovered that the percentage of bone marrow blast, white blood cell, hemoglobin, the proportion of CD52$^+$ cells in CD34$^+$ population, and CD74$^+$DAP12$^+$ cells in the CD34$^+$ population were statistically related to the response of AML patients to induction chemotherapy (Table 2).

In addition, we integrated the more abundant marker genes of clusters 0 and 8 (Figure 6D). We took advantage of RNA-seq data from 134 patients with AML in the TCGA database and 268 patients with AML in GSE165430 and calculated scores of the signature genes of cluster 0 and 8, respectively. The 134 patients from the TCGA database were divided into three groups based on the risk stratification (the favorable group [$n = 17$], the intermediate group [$n = 89$], and the poor group [$n = 26$]). We discovered a significantly lower gene signature score of cluster 0 or 8 in the poor group compared with the favorable group (cluster 0: $p = .0015$, cluster 8: $p = .0001$) and the intermediate group (cluster 0: $p = .019$, cluster 8: $p = .0002$; Figure 6E). Then, we compared the gene signature scores between patients who relapsed ($n = 164$) and patients who remained in CR for ≥3 years ($n = 104$) from GSE165430 (Table S4). We found the gene signature scores of cluster 0 or 8 in relapsed patients were lower than in CR patients (cluster 0: $p = .024$; cluster 8: $p = .009$; Figure 6F). These results suggested that the response of patients with AML to induction chemotherapy was related to the composition of CD34$^+$ cells, and that newly diagnosed patients with high proportions of clusters 0 and 8 were more likely to achieve CR and probably have good prognoses.

DISCUSSION

AML is a group of heterogeneous hematologic malignancies characterized by numerous cytogenetic and molecular alterations.[20] The "3 + 7 regimen" has long been considered the standard of care for AML.[21] Although novel agents such as venetoclax, FLT3 inhibitors, and IDH inhibitors have been approved for various indications in AML since 2017, the 3 + 7 regimen is still the first-line induction chemotherapy. However, the response to this standard induction chemotherapy has significant individual variability because of the heterogeneity in patients with AML.[22]

The heterogeneity of AML has been appreciated since the 1960s.[23] scRNA-seq provides a method to measure and compare the levels of gene expression at single-cell resolution,[24,25] and it has become possible to study the complexity of inter- and intraindividual in AML. Through scRNA-seq, Galen et al. identified six malignant AML cell types that projected along the HSC to myeloid differentiation axis.[26] Pei et al. found monocytic AML was more resistant to venetoclax + azacytidine and that the outgrowth of monocytic subpopulations was favored through a selective process at relapse by scRNA-seq analysis.[27] Riether et al. demonstrated that leukemia stem cells upregulated CD70 in response to hypomethylating agent treatment, resulting in increased CD70/CD27 signaling and citatuzumab could eliminate leukemia stem cells.[28] Duy et al. discovered that AML relapse was facilitated by a senescence-like resilience phenotype that occurred regardless of the stem cell status based on scRNA-seq data.[29] However, there was no research focusing on AML CD34$^+$

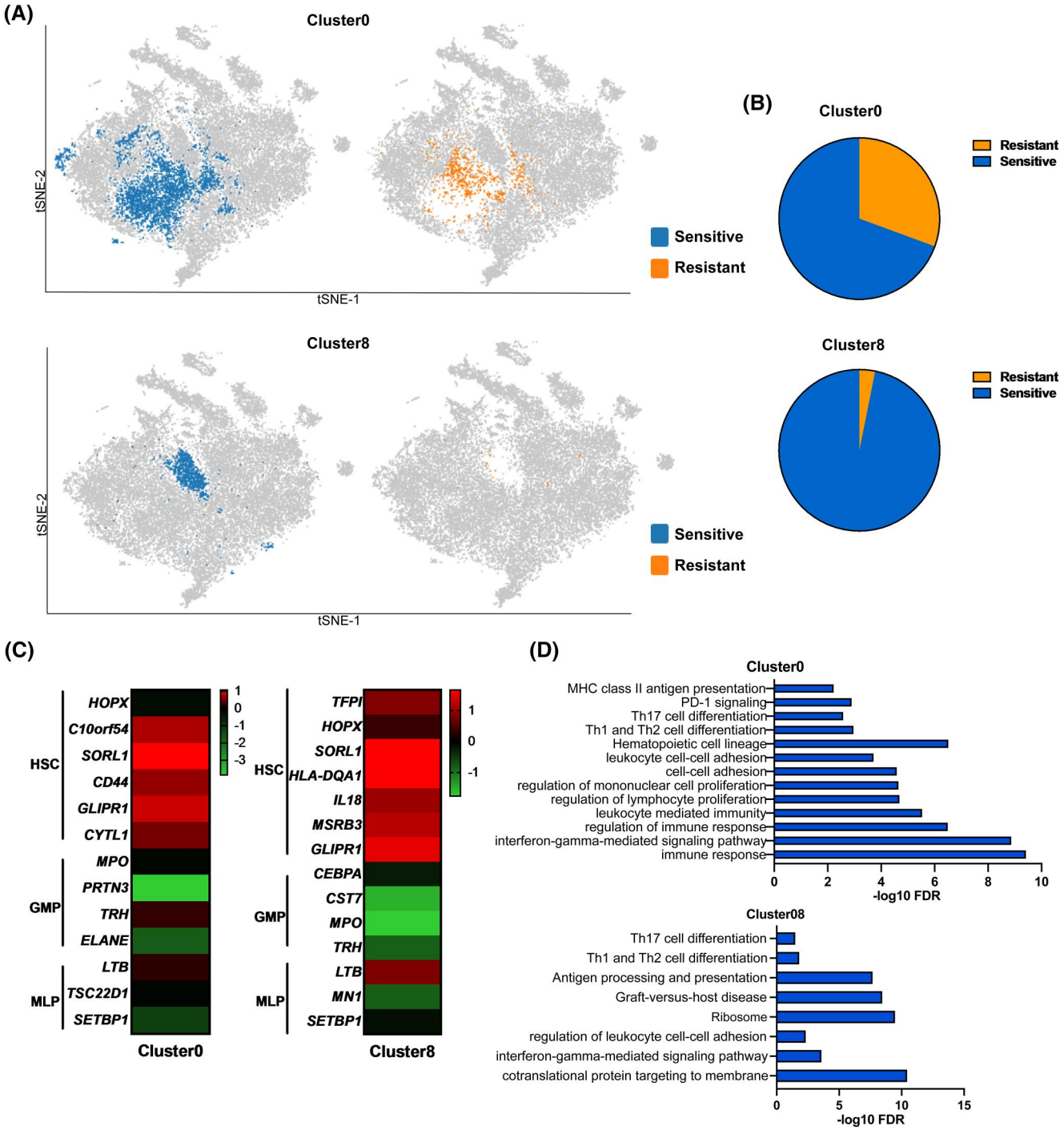

FIGURE 5 Subclusters of CD34$^+$ cells associated with good response to induction chemotherapy in AML. (A) t-SNE visualization of cluster 0 and cluster 8 between resistant and sensitive groups. (B) Pie chart showing cell percentage between resistant and sensitive groups in cluster 0 (top) and cluster 8 (down). (C) Gene expression heatmap of the cell-type–specific marker genes in cluster 0 and cluster 8. (D) Bar graph showing significantly KEGG pathways enrichment analysis of molecular signatures in cluster 0 and cluster 8. AML indicates acute myeloid leukemia; t-SNE, t-distributed stochastic neighbor embedding

hematopoietic stem/progenitor cells at the single-cell level and subpopulations that confer responses to cytarabine + anthracyclines.

AML hematopoietic stem/progenitor cells are thought to generate and perpetuate leukemic populations. To study the heterogeneity of AML hematopoietic stem/progenitor cells, we purified CD34$^+$ cells from bone marrow samples of newly diagnosed patients with AML and performed scRNA-seq by the 10× Genomics platform. We classified AML CD34$^+$ cells into six cell types (HSC, MLP, MEP, GMP, Pro-B cell, and ETP) according to their similarity to normal CD34$^+$ cells because the gene expression program of AML CD34$^+$ cells is not matched completely to the recognized characteristics of normal hematopoietic stem/progenitor cells. We found there were

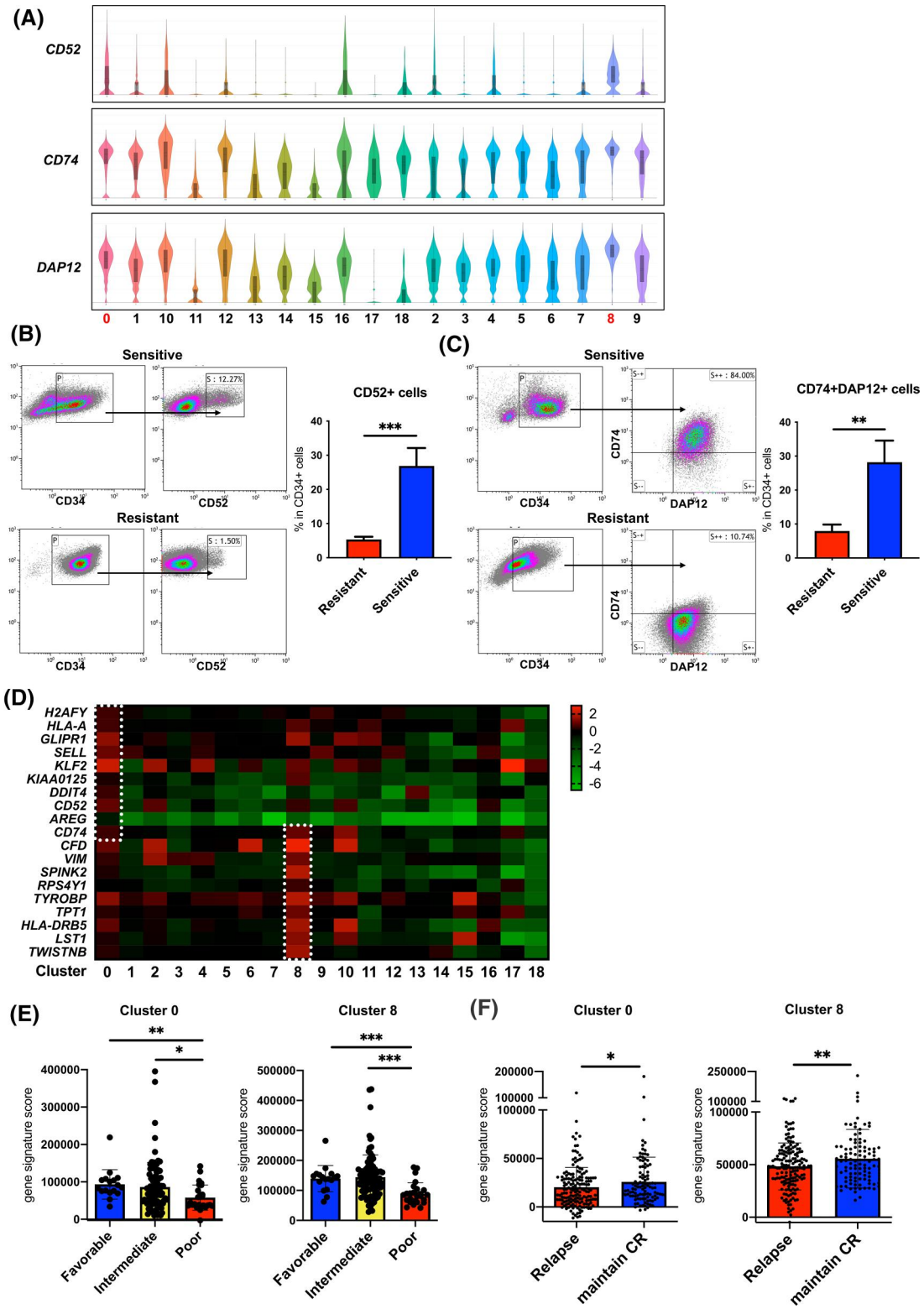

FIGURE 6 Clusters 0 and 8 confer outcomes of newly diagnosed patients with AML. (A) Violin plots showing relative expression of top markers in cluster 0 and cluster 8. (B, C) Bone marrow mononuclear cells were isolated from samples of 32 newly diagnosed patients with AML before treatment, 17 of whom achieved CR (sensitive) and 15 of whom showed NR (resistant) after induction chemotherapy. Flow cytometry was used to detect the percentage of $CD52^+$ (marker of cluster 0) (B) or $CD74^+DAP12^+$ (markers of cluster 8) (C) population in $CD34^+$ cells from patients with AML. Typical flow cytometric plots and histograms were shown. (D) Heatmap showing relative expression of top marker genes in clusters 0 and 8. (E) Histogram showing the statistical analysis results of gene signature scores of cluster 0 (left) and cluster 8 (right) among the favorable ($n = 17$), intermediate ($n = 89$), and poor ($n = 26$) groups of patients with AML from the TCGA database. (F) Histogram showing the statistical analysis results of gene signature scores of cluster 0 (left) and cluster 8 (right) between the relapsed ($n = 164$) and maintain CR ($n = 104$) group of patients with AML from GSE165430. $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$. AML indicates acute myeloid leukemia; CR, complete response; NR, no response

TABLE 2 Association of the cluster 0 and 8 with clinical characteristics of patients with AML

Characteristics	Overall cohort (n = 32)		
	Sensitive n = 17	Resistant n = 15	*p*
Gender (male/female)	13/4	6/9	.0702
Age: y, median (range)	57 (35-69)	53 (15-67)	.2439
Peripheral blast %, median (range)	60 (1-97)	76 (22-98)	.0697
Bone marrow blast %, median (range)	58.03 (18.18-95)	74.33 (27.07-93.09)	.0323
WBC count (10^9/L), median (range)	27.35 (1.89-130.81)	81.23 (1.89-382.48)	.0408
Hemoglobin (g/L), median (range)	97 (66-126)	81 (57-112)	.0052
Platelet count (10^9/L), median (range)	53 (10-207)	87 (2-692)	.2159
Cytogenetic risk category (*n*)			
Favorable	9	5	.2246
Intermediate	4	6	.2674
Adverse	4	4	.5787
Molecular abnormality			
FLT3-ITD mutation			.1583
Positive	4	7	
Negative	13	8	
NPM1 mutation			.4516
Positive	1	2	
Negative	16	13	
CEBPA mutation			.4213
Positive	5	3	
Negative	12	12	
$CD52^+$ cells in $CD34^+$ population			<.0001
>10%	15	0	
≤10%	2	15	
$CD74^+DAP12^+$ cells in $CD34^+$ population			.0023
>30%	8	0	
≤30%	9	15	

Abbreviations: AML, acute myeloid lymphoma; WBC, white blood cell.

obvious differences in the composition of $CD34^+$ cells between newly diagnosed patients with AML and healthy controls. We identified malignant-like clusters according to cell source, copy number variations, and gene set enrichment analysis. Granja et al. defined clusters to be healthy-like if a high percentage (>80% for scRNA-seq and >90% for scATAC [assay for transposase-accessible chromatin]) of the cells were from the normal data.[30] Jin et al. inferred large-scale copy number variations based on scRNA-seq data to distinguish malignant from nonmalignant cells.[31] These methods were also used in our study to identify malignant-like clusters.

To further demonstrate the reason for distinct responses of patients with AML to induction chemotherapy, we divided the six newly diagnosed patients with AML into two groups: the sensitive group and the resistant group. By comparing the heterogeneity of $CD34^+$ cells in the two groups, we discovered one cell population was related to no response of patients with AML to induction therapy. This cell population had a specific gene signature program, including *CRIP1*, *LGALS1*, and *S100As*. *CRIP1* was reported to dramatically recover the 5-fluorouracil–inhibited cancer cell proliferation *in vitro* and stimulate the tumor formation *in vivo*.[32] *LGALS1* is best known for its role in RAS signaling and is associated with shorter disease-free survival and increased blasts.[33] Moreover, *LGALS1* mediates immune evasion by preventing T-cell migration into the tumor.[34] Although *S100As* are significantly associated with poor prognosis in patients with low-grade glioma as reported,[35] the prognostic and oncologic values of the S100A family have not been systematically

investigated in most cancers including AML. Here, we combined these genes as a specific gene signature program to mark the cell population, which was different between AML-sensitive and AML-resistant groups, and TCGA data confirmed this gene signature program was associated with poor survival of AML.

Interestingly, the two patients (AML01 and AML03) in the resistant group both had MLL rearrangement, but they belonged to different cytogenetic risk stratifications (adverse and intermediate). To figure out whether the cluster (*CRIP1*^{high}*LGALS1*^{high}*S100As*^{high}) associated with poor responses to induction therapy had MLL rearrangement–related gene signatures, we compared the transcriptional expression of each cluster on the basis of MLL rearrangement–related genes by GSEA enrichment analysis.[27] The results showed that MEP-1 and MLP in the resistant group were associated with MLL rearrangement (Figure S6). However, pathway enrichment analysis showed that MEP-1 in the resistant group did not significantly enrich for AML, cell cycle, oxidative phosphorylation or uncontrolled transcription, compared with the sensitive group. MLP was excluded from resistant-related clusters during subclusters analysis (Figure 3A). Furthermore, no prominent enrichment of MLL-related genes was observed in the resistant-related cluster featured with *CRIP1*^{high}*LGALS1*^{high}*S100As*^{high}. Therefore, our results suggested that the signatures (*CRIP1*^{high}*LGALS1*^{high}*S100As*^{high}) may be relatively independent genetic characteristics for poor prognosis, which is not associated with MLL rearrangement.

In addition, two clusters in CD34$^+$ cells were found to be correlated with good response to induction chemotherapy in AML. The surface markers of the two clusters are CD34$^+$CD52$^+$ and CD34$^+$CD74$^+$DAP12$^+$, respectively. Multiparameter flow cytometry showed that newly diagnosed patients with AML with high proportions of these two clusters were more likely to achieve CR after the first induction chemotherapy. Interestingly, cells of these two clusters display characteristics of HSC and were enriched in immune response including Th17 cell differentiation and interferon-gamma–mediated signal pathway.

In summary, we leveraged high-throughput single-cell transcriptomics to parse heterogeneous CD34$^+$ cells in newly diagnosed patients with AML. Our results provide insight into the aberrant gene programs of AML CD34$^+$ cells comparing with the normal, reveal a striking difference between the "resistant group" and "sensitive group" and identify the specific cell populations correlated with chemoresistance and chemosensitivity. It is the first discovery for clusters of CD34$^+$ cells in newly diagnosed patients with AML that is related with the response of different individuals to induction chemotherapy. Our data and findings can predict the response of patients with AML to induction treatment at diagnosis phase and guide therapeutic strategies to target critical and specific cell components of chemoresistance.

AUTHOR CONTRIBUTIONS

Ruinan Jia: Experiments, data analysis, manuscript writing and bioinformatic analysis. **Min Ji:** Research design, experiments, data analysis, and manuscript writing. **Guosheng Li:** Experiments, data analysis, and manuscript writing. **Yuan Xia:** Bioinformatic analysis. **Shouhui Guo:** Bioinformatic analysis. **Peng Li:** Clinical sample and information collection. **Yanping Sun:** Clinical sample and information collection. **Fei Lu:** Clinical sample and information collection. **Jingru Zhang:** Research design. **Shaolei Zang:** Experiments, data analysis, and manuscript writing. **Shuxin Yan:** Scientific and technical support. **Jingjing Ye:** Research design. **Fuzhong Xue:** Scientific and technical support. **Daoxin Ma:** Scientific and technical support. **Tao Sun:** Research supervision and coordination. **Chunyan Ji:** Research supervision and coordination. All authors revised, read and approved the final manuscript.

ACKNOWLEDGMENTS

This work was supported by grants from the Distinguished Taishan Scholars in Climbing Plan (tspd20210321) and Young Taishan Scholars (tsqn201812132), the National Natural Science Foundation of China (81770159, 82070160, 81873425, 81700143), and the Shandong Provincial Natural Science Foundation (ZR2020MH118, ZR2020KH016, ZR2020MH119).

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

ORCID

Fuzhong Xue https://orcid.org/0000-0003-0378-7956
Chunyan Ji https://orcid.org/0000-0003-4438-1768

REFERENCES

1. Schwind S, Marcucci G, Maharry K, et al. BAALC and ERG expression levels are associated with outcome and distinct gene and microRNA expression profiles in older patients with de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *Blood*. 2010;116(25):5660-5669. https://doi.org/10.1182/blood-2010-06-290536
2. Attar EC, Johnson JL, Amrein PC, et al. Bortezomib added to daunorubicin and cytarabine during induction therapy and to intermediate-dose cytarabine for consolidation in patients with previously untreated acute myeloid leukemia age 60 to 75 years: CALGB (Alliance) study 10502. *J Clin Oncol*. 2013;31(7):923-929. https://doi.org/10.1200/jco.2012.45.2177
3. Morita K, Kantarjian HM, Wang F, et al. Clearance of somatic mutations at remission and the risk of relapse in acute myeloid leukemia. *J Clin Oncol*. 2018;36(18):1788-1797. https://doi.org/10.1200/jco.2017.77.6757
4. Tallman MS, Gilliland DG, Rowe JM. Drug therapy for acute myeloid leukemia. *Blood*. 2005;106(4):1154-1163. https://doi.org/10.1182/blood-2005-01-0178
5. Büchner T, Hiddemann W, Wörmann B, et al. Double induction strategy for acute myeloid leukemia: the effect of high-dose cytarabine with mitoxantrone instead of standard-dose cytarabine with daunorubicin and 6-thioguanine: a randomized trial by the German AML Cooperative Group. *Blood*. 1999;93:4116-4124.
6. Loken MR, Alonzo TA, Pardo L, et al. Residual disease detected by multidimensional flow cytometry signifies high relapse risk in patients with de novo acute myeloid leukemia: a report from Children's Oncology Group. *Blood*. 2012;120:1581-1588. https://doi.org/10.1182/blood-2012-02-408336
7. Pollyea DA, Jordan CT. Therapeutic targeting of acute myeloid leukemia stem cells. *Blood*. 2017;129(12):1627-1635. https://doi.org/10.1182/blood-2016-10-696039
8. Tremblay CS, Saw J, Chiu SK, et al. Restricted cell cycle is essential for clonal evolution and therapeutic resistance of pre-leukemic stem

cells. *Nat Commun*. 2018;9(1):3535. https://doi.org/10.1038/s41467-018-06021-7
9. Drissen R, Buza-Vidas N, Woll P, et al. Distinct myeloid progenitor-differentiation pathways identified through single-cell RNA sequencing. *Nat Immunol*. 2016;17(6):666-676. https://doi.org/10.1038/ni.3412
10. Paul F, Arkin Y, Giladi A, et al. Transcriptional heterogeneity and lineage commitment in myeloid progenitors. *Cell*. 2016;164(1-2):325. https://doi.org/10.1016/j.cell.2015.12.046
11. Giustacchini A, Thongjuea S, Barkas N, et al. Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. *Nat Med*. 2017;23(6):692-702. https://doi.org/10.1038/nm.4336
12. Savas P, Virassamy B, Ye C, et al. Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. *Nat Med*. 2018;24(7):986-993. https://doi.org/10.1038/s41591-018-0078-7
13. Mathys H, Davila-Velderrain J, Peng Z, et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature*. 2019;570(7761):332-337. https://doi.org/10.1038/s41586-019-1195-2
14. Witkowski MT, Dolgalev I, Evensen NA, et al. Extensive remodeling of the immune microenvironment in B cell acute lymphoblastic leukemia. *Cancer Cell*. 2020;37(6):867-882.e12. https://doi.org/10.1016/j.ccell.2020.04.015
15. Zheng GX, Terry JM, Belgrader P, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun*. 2017;8(1):14049. https://doi.org/10.1038/ncomms14049
16. Cheson BD, Bennett JM, Kopecky KJ, et al. Revised recommendations of the International Working Group for Diagnosis, Standardization of Response Criteria, Treatment Outcomes, and Reporting Standards for Therapeutic Trials in Acute Myeloid Leukemia. *J Clin Oncol*. 2003;21(24):4642-4649. https://doi.org/10.1200/jco.2003.04.036
17. Hua P, Roy N, de la Fuente J, et al. Single-cell analysis of bone marrow-derived CD34+ cells from children with sickle cell disease and thalassemia. *Blood*. 2019;134(23):2111-2115. https://doi.org/10.1182/blood.2019002301
18. Laurenti E, Doulatov S, Zandi S, et al. The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. *Nat Immunol*. 2013;14(7):756-763. https://doi.org/10.1038/ni.2615
19. Yagi T, Morimoto A, Eguchi M, et al. Identification of a gene expression signature associated with pediatric AML prognosis. *Blood*. 2003;102(5):1849-1856. https://doi.org/10.1182/blood-2003-02-0578
20. Zhang H, Savage S, Schultz AR, et al. Clinical resistance to crenolanib in acute myeloid leukemia due to diverse molecular mechanisms. *Nat Commun*. 2019;10(1):244. https://doi.org/10.1038/s41467-018-08263-x
21. Kantarjian H, Kadia T, DiNardo C, et al. Acute myeloid leukemia: current progress and future directions. *Blood Cancer J*. 2021;11(2):41. https://doi.org/10.1038/s41408-021-00425-3
22. Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic classification and prognosis in acute myeloid leukemia. *N Engl J Med*. 2016;374(23):2209-2221. https://doi.org/10.1056/nejmoa1516192
23. Levine JH, Simonds EF, Bendall SC, et al. Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell*. 2015;162(1):184-197. https://doi.org/10.1016/j.cell.2015.05.047
24. Eberwine J, Sul JY, Bartfai T, Kim J. The promise of single-cell sequencing. *Nat Methods*. 2014;11(1):25-27. https://doi.org/10.1038/nmeth.2769
25. Stegle O, Teichmann SA, Marioni JC. Computational and analytical challenges in single-cell transcriptomics. *Nat Rev Genet*. 2015;16(3):133-145. https://doi.org/10.1038/nrg3833
26. van Galen P, Hovestadt V, Wadsworth Ii MH, et al. Single-cell RNA-seq reveals AML hierarchies relevant to disease progression and immunity. *Cell*. 2019;176(6):1265-1281.e24. https://doi.org/10.1016/j.cell.2019.01.031
27. Pei S, Pollyea DA, Gustafson A, et al. Monocytic subclones confer resistance to venetoclax-based therapy in patients with acute myeloid leukemia. *Cancer Discov*. 2020;10(4):536-551. https://doi.org/10.1158/2159-8290.cd-19-0710
28. Riether C, Pabst T, Hopner S, et al. Targeting CD70 with cusatuzumab eliminates acute myeloid leukemia stem cells in patients treated with hypomethylating agents. *Nat Med*. 2020;26(9):1459-1467. https://doi.org/10.1038/s41591-020-0910-8
29. Duy C, Li M, Teater M, et al. Chemotherapy induces senescence-like resilient cells capable of initiating AML recurrence. *Cancer Discov* 2021;11(6):1542-1561. https://doi.org/10.1158/2159-8290.cd-20-1375
30. Granja JM, Klemm S, McGinnis LM, et al. Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat Biotechnol*. 2019;37(12):1458-1465. https://doi.org/10.1038/s41587-019-0332-7
31. Jin S, Li R, Chen MY, et al. Single-cell transcriptomic analysis defines the interplay between tumor cells, viral infection, and the microenvironment in nasopharyngeal carcinoma. *Cell Res*. 2020;30(11):950-965. https://doi.org/10.1038/s41422-020-00402-8
32. Zhang L, Zhou R, Zhang W, et al. Cysteine-rich intestinal protein 1 suppresses apoptosis and chemosensitivity to 5-fluorouracil in colorectal cancer through ubiquitin-mediated Fas degradation. *J Exp Clin Cancer Res*. 2019;38:120. https://doi.org/10.1186/s13046-019-1117-z
33. Ruvolo PP, Ma H, Ruvolo VR, Zhang X, Post SM, Andreeff M. LGALS1 acts as a pro-survival molecule in AML. *Biochim Biophys Acta Mol Cell Res*. 2020;1867(10):118785. https://doi.org/10.1016/j.bbamcr.2020.118785
34. Nambiar DK, Aguilera T, Cao H, et al. Galectin-1-driven T cell exclusion in the tumor endothelium promotes immunotherapy resistance. *J Clin Invest*. 2019;129(12):5553-5567. https://doi.org/10.1172/jci129025
35. Zhang Y, Yang X, Zhu XL, et al. S100A gene family: immune-related prognostic biomarkers and therapeutic targets for low-grade glioma. *Aging (Albany NY)*. 2021;13(11):15459-15478. https://doi.org/10.18632/aging.203103

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Jia R, Ji M, Li G, et al. Subclones of bone marrow CD34$^+$ cells in acute myeloid leukemia at diagnosis confer responses of patients to induction chemotherapy. *Cancer*. 2022;128(22):3929-3942. https://doi.org/10.1002/cncr.34481