

Subclones of Bone Marrow CD34+ Cells in Acute Myeloid Leukemia at Diagnosis Confer Responses of Patients to Induction Chemotherapy

Ruinan Jia

Qilu hospital of Shandong university

Min Ji

Qilu hospital of Shandong university

Guosheng Li

Qilu hospital of Shandong university

Yuan Xia

Qilu hospital of Shandong university

Shouhui Guo

Shandong university

Peng Li

Qilu hospital of Shandong university

Yanping Sun

Qilu hospital of Shandong university

Fei Lu

Qilu hospital of Shandong university

Jingru Zhang

Qilu hospital of Shandong university

Shaolei Zang

Qilu hospital of Shandong university

Shuxin Yan

Qilu hospital of Shandong university

Jingjing Ye

Qilu hospital of Shandong university

Fuzhong Xue

Shandong university

Daoxin Ma

Qilu hospital of Shandong university

Tao Sun

Qilu hospital of Shandong university

Chunyan Ji (✉ jjichunyan@sdu.edu.cn)

Research

Keywords: Acute myeloid leukemia, Single-cell RNA-sequencing, CD34+ cell, Heterogeneity, Induction chemotherapy

Posted Date: November 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1047797/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Abstract

Background

Acute myeloid leukemia (AML) is a fatal hematopoietic malignancy which has a prognosis that varies with genetic heterogeneity of its hematopoietic stem/progenitor cells (HSCPs). Intensive induction chemotherapy with cytarabine and anthracycline has been the standard of 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 10× Genomics, a high-throughput single cell RNA sequencing (scRNA-seq) platform, we analyzed the cellular heterogeneity of bone marrow CD34⁺ cells from newly diagnosed AML patients who were then divided into sensitive and resistant groups according to their responses to induction chemotherapy with cytarabine and anthracycline. ScRNA-seq data for healthy controls were obtained from the GEO database. We verified our findings by the TCGA database, GEO datasets and the multiparameter flow cytometry.

Results

From the integrative analysis of 60,402 cells, we established a landscape for single-cell CD34⁺ cells in AML patients and identified the hematopoietic stem/progenitor cell types based on the lineage signature genes. Interestingly, through recognizing the malignant-like clusters and comparing “resistant group” with “sensitive group”, we found a cell population with specific gene signatures (*CRIP1^{high}LGALS1^{high}S100As^{high}*) showing features of granulocyte-monocyte progenitors (GMP) was associated with poor prognosis of AML. And two cell populations of AML CD34⁺ cells marked by CD52⁺ or CD74⁺DAP12⁺ were related to good response of patients to induction therapy, showing characteristics of hematopoietic stem cell (HSC).

Conclusion

Our study indicates the heterogeneity of AML CD34⁺ cells confers for outcomes of AML and provides possible biomarkers to predict the response of AML patients to induction chemotherapy.

Background

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 an 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 AML are none remission to chemotherapy, and at least 50% of those who achieve remission will relapse [3–5]. The different outcomes of AML patients are due to the fact that AML is a heterogeneous and molecularly complex disease with variable hematologic phenotypes [6].

Normal hematopoietic stem cells (HSCs) give rise to mature blood cell types of the myeloid, lymphoid, and erythroid/megakaryocyte lineages. 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 inter-patient, 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⁺ T_{RM} 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 non-classical monocytes subcluster in bone marrow microenvironment supporting B-ALL progression and treatment evasion [14]. In our research, scRNA-seq showed the potential to improve treatment efficacy prediction by taking into account differences among subpopulations of CD34⁺ hematopoietic stem/progenitor cells in newly diagnosed AML patients.

Here, we adapted 10× Genomics scRNA-seq technology [15] to acquire transcriptional data for thousands of single cells from bone marrow CD34⁺ cells of newly diagnosed AML patients. We showed a comprehensive and heterogeneous map of the CD34⁺ cells in AML patients 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 AML patients 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 AML patients.

Materials And Methods

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. In addition, patients for scRNA-seq met the following criteria: 18 ~ 65 years old, without severe infection, none APL, without MDS history, without cancer chemotherapy or radiotherapy record, WBC count $<100 \times 10^9/L$. 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 for AML. CR was defined as bone marrow blasts $\leq 5\%$, no circulating blasts, with complete blood count recovery (neutrophil count $\geq 1000/\mu L$ and platelet count $\geq 100,000/\mu L$). Partial remission (PR) required all the hematologic values of CR with decrease of $\geq 50\%$ in the percentage of bone marrow blasts to 5% ~ 25%. None remission (NR) was defined as not up to the criteria of PR and CR.

Single-cell preparation

Mononuclear cells from bone marrow aspirates were separated by Ficoll-Paque Plus (Pharmacia LKB Biotechnology) density gradient centrifugation. Then $CD34^+$ cells were MACS-purified from the cell suspensions according to the manufacturer's instructions (Miltenyi Biotech). The $CD34^+$ single cell suspension used for scRNA-seq was washed twice with PBS containing 0.04% BSA and the final cell concentration was adjusted to ~ 1000 cells/ μL . Cells used for flow cytometry were washed with PBS containing 0.02% BSA then stained for markers.

Single-cell RNA library construction and sequencing

We used the Single Cell 3' Reagent kit (V2) and the Chromium instrument to prepare individually barcode single cell RNA-seq libraries following the manufacturer's protocols (10x Genomics). For quality control and to quantify the library concentration, we used the BioAnalyzer (Agilent BioAnalyzer High Sensitivity Kit). Sequencing with dual indexing was conducted on an Illumina NovaSeq PE150 machine using the 150 cycles High Output kit. ScRNA-seq data for normal $CD34^+$ cells from four healthy donors were obtained from the GEO database (GSE133181). The Cell Ranger Single Cell Software Suite was used to perform sample demultiplexing, barcode processing, and single-cell 3' gene counting. Further analysis including the identification of highly variable genes, dimensionality reduction, standard unsupervised clustering algorithms, and the discovery of differentially expressed genes was performed using the Seurat R package and Loupe software.

Assigning sequenced single cells to hematopoietic lineages

The cell type specific signature genes in each hematopoietic stem or progenitor cell (HSPC) population were downloaded from http://www.jdstemcellresearch.ca/index_files/ResearchData.htm. We build a reference profile using the top 250 specific signature genes. We assigned to each cluster a specific hematopoietic cell identity by comparing cluster-specific genes with lineage signature genes reference profile. Finally, cell clusters 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 (ProB) and earliest thymic progenitors (ETP).

Flow cytometry

Antibodies of FITC anti-human CD34, PC5.5 anti-human CD74, PE anti-human CD52 and Zombie Red Fixable Viability Kit were purchased from Biolegend (San Diego, USA) and DAP12 antibody was from Santa Cruz Biotechnology (Texas, USA). Cells were stained with five panels of 4 antibodies each for flow cytometry (Beckman Coulter) then analyzed by Kaluza analysis software.

Bulk expression analysis

Bulk RNA-seq expression levels from the TCGA database and GEO datasets were downloaded from the companion website of the original publication. We calculated expression scores for cluster specific signatures in bulk profiles using the following approach: For each gene in a given signature, we selected the 100 genes with the smallest difference in average expression level as a background gene set. The average expression of the background gene set was then subtracted from the respective signature gene, and the average of the resulting values of all signature genes was kept as the signature score. These scores were used to stratify patients into two groups, followed by survival analysis.

Survival analysis

The publicly available RNA sequencing data of 134 AML patients (without APLs) and complete clinical and survival data were downloaded from TCGA database. AML patients were divided into two groups according to the gene signature score of cluster specific genes program and the log-rank test was used to analyze the overall survival rate.

Statistical analysis

Differences between 2 groups were analyzed using an unpaired Student t-test. Data are presented as mean \pm standard deviation (SD). Fisher's exact test was used for comparing the CR rates of patients. $P < 0.05$ was considered statistically significant.

Results

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

In order to clarify cell diversity of CD34⁺ cells in AML, we performed scRNA-seq on the Chromium platform (10 \times Genomics) of primary CD34⁺ cells which were MACS-purified from the bone marrow of six newly-diagnosed AML patients before treatment. Normal CD34⁺ cells from four healthy donors were used as control whose scRNA-seq data were obtained from the GEO database. The patients' clinical information was shown in Table 1. We acquired high quality data of 60 402 CD34⁺ cells from AML patients and healthy controls and analyzed the data by the Seurat utilizing its anchoring-based integration method to account for technical and biological variance between individual samples, followed by uniform manifold approximation and projection dimensionality reduction for visualization (Figure 1A). Sample viability did not significantly correlate with the mean number of detected genes, the number of unique molecular identifiers (UMIs), and the fraction of mitochondrial RNA transcripts.

Then we 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 (ProB) and earliest thymic progenitors (ETP)), and seventeen clusters (HSC-1 & 2, MLP, MEP-1 ~ 3, GMP-1 ~ 7, ProB-1 & 2, ETP-1 & 2), based on comparing cluster-specific genes with reported lineage signature genes (Figures 1B-D, Supplemental Table 1) [16].

CD34⁺ cells of AML patients at diagnosis enriches in GMP population

We observed obvious differences in the composition of the stem/progenitor lineage between newly diagnosed AML patients and healthy controls. The results showed that the proportion of GMP population was significantly higher in AML patients than in healthy controls ($P = 0.000003$), while the proportion of HSC and pro-B populations were significantly lower ($P = 0.00002$, Figure 2A). The percentages of cluster GMP-1, 2, 3, 5 & 7 were higher in AML patients than in healthy controls ($P < 0.05$, Figure 2B).

We then defined the cluster to be normal-like if the high percentage (> 60%) of cells in the cluster was from healthy controls, and the remaining clusters were defined as malignant-like. As a result, the cluster GMP-1 ~ 7, MEP-1, MLP and ETP-1 were classified as malignant-like clusters, and the others were classified as normal-like clusters (Figure 2C). Obviously, CD34⁺ cells from AML patients were mainly composed of malignant-like clusters (Supplemental Figure 1). Copy number variations (CNVs) were also highly consistent with the result and malignant-like cells were grouped together with extensive CNVs across the whole genome (Supplemental Figure 2). 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 2D & E).

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

About 30% of newly diagnosed AML patients fail to reach complete remission (CR) after induction chemotherapy with cytarabine and anthracycline. We hypothesized that the distinct response of AML patients to chemotherapy was due to the heterogeneity of CD34⁺ cells. To address this, we divided the 6 AML patients into two groups according to the response of AML patients to the standard induction therapy with cytarabine and anthracycline. AML02, AML05, AML06 and AML07 achieved CR after induction therapy and were assigned to “sensitive” group, while AML01 and AML03 were assigned to “resistant” group since they were no response to induction therapy (Figure 3A & Table 1). The 6 AML patients were in different cytogenetic risk categories, especially AML01 and AML03 belonged to adverse and intermediate risk category respectively. We then analyzed the single cell data of CD34⁺ cells in the two groups (a total of 28 381 single-cell transcriptomes). We found that both groups of CD34⁺ cells were composed mainly of the 10 malignant-like clusters which belonged to 4 cell types (GMP, MEP, MLP and

ETP) (Figure 3A). The proportion of each cell cluster did not show significantly different between the sensitive and resistant groups (Figure 3B).

It seemed a modest increase in the proportion of malignant-like clusters was observed in the resistant group, when compared with the sensitive group (Figure 3C). However, there was no significant difference in the proportion of cells derived from the two groups in each malignant-like cluster (Figure 3D). Further, to search for the differences between AML sensitive and 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, 7 and MLP in AML resistant group obviously enriched for acute myeloid leukemia, cell cycle, oxidative phosphorylation and uncontrolled transcription, compared with the sensitive group (Figure 3E-G). In addition, these clusters in AML resistant group expressed higher levels of genes which were reported to be associated with poor prognosis [17] than in AML sensitive group (Supplemental Figure 3). The results indicated that although the composition of CD34⁺ cells were almost the same between AML sensitive and resistant groups, the gene expression profiles were quite different in the GMP-1, 2, 3, 5, 7 and MLP clusters, suggesting these clusters were probably composed of different subclusters which may confer to the distinct response of AML patients to chemotherapy.

Subclusters with specific gene signatures (*CRIP1*^{high} *LGALS1*^{high} *S100A6*^{high}) are associated with poor prognosis of AML

To figure out the specific subcluster of CD34⁺ cells which are associated with no response to treatment of AML, we subgrouped these six cell clusters (GMP-1, 2, 3, 5, 7 and MLP) and compared the proportion of subclusters in AML sensitive, AML resistant and control groups (Figure 4A).

In GMP-1 cluster, we identified 9 transcriptionally distinct cell subclusters and the proportion of subcluster GMP-1-0 was markedly higher in AML resistant group than in AML sensitive group and controls (Figure 4A). Specific gene expression data analysis indicated the highly expressing markers of subcluster GMP-1-0, such as *CRIP1*, *S100A10*, *S100A6*, *LGALS1*, which were all significantly related to poor prognosis in cancers (Figure 4B). Interestingly, in GMP-3 cluster, we found the proportion of subcluster GMP-3-0 was also higher in AML resistant group and its specific gene signatures were almost the same as the subcluster GMP-1-0 (Figure 4A & B). Further, 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 4C).

To further clarify the significance of these subclusters with specific gene signatures (*CRIP1*^{high}*LGALS1*^{high}*S100A6*^{high}) for AML, we took advantage of 134 AML patients from TCGA database and calculated the gene signature scores, then divided the patients into two groups based on the scores (the high score group (n = 67), the low score group (n = 67), Figure 4D). We noted a significantly inferior overall survival (OS) 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}*S100A6*^{high} plays a pernicious role in AML

survival (Figure 4E). Therefore, we speculated that GMP-1-0 and GMP-3-0 may be the specific subpopulations which were associated with poor responses to induction chemotherapy in newly diagnosed AML patients.

Another way of clustering also confirms the existence of the subcluster with specific gene signatures ($CRIP1^{high}$ $LGALS1^{high}$ $S100A6^{high}$)

To exclude the interference of healthy controls, we removed healthy controls and only reanalyzed the data for $CD34^{+}$ cells from AML patients, which contained a total of 28 381 single-cell transcriptomes from AML resistant ($n = 8\ 935$) and sensitive ($n = 19\ 446$) groups (Figure 5A). Sample viability did not significantly correlate with the mean number of detected genes, the number of UMIs nor the fraction of mitochondrial RNA transcripts. As shown in Figure 5B, $CD34^{+}$ cells from AML patients were partitioned into 19 clusters. After analysis of the composition in each group, we observed increases in the proportion of cluster 2, 3, 5, 6 in AML resistant group (Figure 5C).

In cluster 2, 3, 5 and 6, the high proportion (>60%) of cells was from AML resistant group (Figure 5D). Through matching the marker genes of cluster 2, 3, 5 and 6 with lineage signature genes, we found cluster 2, 5, 6 were similar to GMP and cluster 3 was similar to MEP (Figure 5E). The cell types of these clusters were defined as GMP-like and MEP-like. Figure 5F showed 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 performed gene expression enrichment analysis, revealing the upregulated genes were involved in cell adhesion, cell proliferation and S100 protein (Figure 5G). 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.

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 sensitive groups, we found that there was a significant increase in the proportion of cluster 0 and 8 in the sensitive group (Figure 6A & 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 6C). 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 cells differentiation and immune response (Figure 6D).

We also recognized the surface markers of cluster 0 (CD52) and 8 (CD74 and DAP12) (Figure 7A). To verify whether the distinct response of newly diagnosed AML patients to induction chemotherapy were

attributed to the differences of clusters in CD34⁺ cells, we used multiparameter flow cytometry to detect the HSC-like clusters in newly diagnosed AML patients. Bone marrow mononuclear cells were isolated from samples of 12 newly diagnosed AML patients before treatment. Six patients achieved CR after induction chemotherapy and were assigned to “sensitive” group, while the other six patients showed NR after induction chemotherapy and were assigned to “resistant” group (Supplemental Table 2). The result showed that the proportion of CD52⁺ cells and CD74⁺DAP12⁺ cells in CD34⁺ population at diagnosis was significantly higher in AML patients with CR (sensitive group) than those with NR (resistant group) (CD52⁺: $P = 0.0125$, CD74⁺DAP12⁺: $P = 0.0124$), which was consistent with the finding of scRNA-seq (Figure 7B & C). In AML patients with CD52⁺ cells >10% of CD34⁺ population, the CR rate after induction chemotherapy was significantly higher than that in patients with CD52⁺ cells ≤10% of CD34⁺ population (100% vs 0, $P = 0.0011$). The CR rate was also higher in patients with CD74⁺DAP12⁺ cells >30% than ≤30% of CD34⁺ population (100% vs 25%, $P = 0.03$). Moreover, we analyzed the characteristics of the 12 AML patients. We discovered that age, hemoglobin, the proportion of CD52⁺ cells in CD34⁺ population and CD74⁺DAP12⁺ cells in CD34⁺ population were statistically related to the response of newly diagnosed AML patients to induction chemotherapy (Table 2). In addition, we integrated the more abundant marker genes of cluster 0 and 8 (Figure 7D). We took advantage of 134 AML patients from TCGA database and 268 AML patients from GSE165430, and calculated the gene signature scores respectively. The 134 patients from 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: $P = 0.0015$, Cluster 8: $P = 0.0001$) and the intermediate group (Cluster 0: $P = 0.019$, Cluster 8: $P = 0.0002$, Figure 7E). 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. We found the gene signature scores of cluster 0 or 8 in relapsed patients were lower than those in maintain CR patients (Cluster 0: $P = 0.024$, Cluster 8: $P = 0.009$, Figure 7F). These results suggested that the response of AML patients 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 prognosis.

Discussion

Acute myeloid leukemia (AML) is a group of heterogeneous hematologic malignancies characterized by numerous cytogenetic and molecular alterations [18]. The discovery of the activity of cytarabine (Ara-C) and anthracyclines in AML and combining them in the 1970’s into what is known as the “3 + 7 regimen”, has long been considered the standard of care [19]. Although novel agents such as venetoclax, FLT3 inhibitors, IDH inhibitors, have been approved for various indications in AML since 2017, “3 + 7 regimen” is still the first line induction chemotherapy. However, the response to this standard induction chemotherapy has significant individual variability due to the heterogeneity in AML patients [20].

The heterogeneity of AML has been appreciated since the 1960s [21]. Flow cytometry is widely used for exploring cell heterogeneity in leukemia, however, it is limited to the surface markers. Bulk population sequencing can be used into cell genome and transcriptome, but cannot exhibit the information of individual cells. ScRNA-seq provides a method to measure and compare the levels of gene expression at single cell resolution [22, 23]. While high-throughput scRNA-seq technologies matured such that one can sequence and analyze thousands of cells per tumor, it has become possible to study the complexity of inter- and intra-individual in AML. Through scRNA-seq, Galen *et al*/ identified six malignant AML cell types that project along the HSC to myeloid differentiation axis, and found the dysregulated transcriptional programs with co-expression of stemness- and myeloid priming genes in AML [24]. Jiang *et al*/ reported patients with t(8;21) AML with a higher proportion of CD34⁺ CD117^{dim} cells had significantly worse clinical outcomes than those with a lower CD34⁺ CD117^{dim} proportion [25]. Pei *et al*/ found monocytic AML was more resistant to venetoclax + azacytidine and the outgrowth of monocytic subpopulations was favored through a selective process at relapse by scRNA-seq analysis [26]. Riether *et al*/ demonstrated that leukemia stem cells (LSCs) upregulated CD70 in response to hypomethylating agents (HMAs) treatment resulting in increased CD70/CD27 signaling and cusatuzumab could eliminate LSCs [27]. Duy *et al*/ discovered that AML relapse was facilitated by a senescence-like resilience phenotype that occurred regardless of their stem cell status based on scRNA-seq data [28]. However, there was no research focusing on AML CD34⁺ hematopoietic stem/progenitor cells at the single cell level and subpopulations which 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 6 newly diagnosed AML patients and performed scRNA-seq by 10× Genomics platform. We classified AML CD34⁺ cells into six cell types (HSC, MLP, MEP, GMP, ProB and ETP) according to their similarity to normal CD34⁺ cells, since 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 obvious differences in the composition of CD34⁺ cells between newly diagnosed AML patients and healthy controls. The proportion of GMP population was significantly higher in AML patients than in healthy controls, while the proportion of HSC population was significantly lower. Then, 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) of the cells were from the normal data [29]. Jin *et al*/ inferred large-scale copy number variations based on scRNA-seq data to distinguish malignant from non-malignant cells [30]. These methods were also used in our study to identify malignant-like clusters.

To further demonstrate the reason for distinct responses of AML patients to induction chemotherapy, we divided the six newly diagnosed AML patients into two groups. Patients who achieved CR after induction chemotherapy were assigned to the sensitive group, while patients who were NR were assigned to the resistant group. Although the two patients in the resistant group had KMT2A rearrangement, they belonged to different cytogenetic risk stratifications (adverse and intermediate). Through analyzing the

expression of KMT2A-rearrangement related genes among all the clusters, we found there was no obvious difference among all the clusters or outsized value of certain cluster (data not shown). By comparing the heterogeneity of CD34⁺ cells in the two groups, we discovered one cell population related to chemoresistance. This cell population had specific gene signature program, including *CRIP1*, *LGALS1* and *S100As*. *CRIP1* was reported to dramatically recover the 5-Fluorouracil (5-FU) inhibited cancer cell proliferation *in vitro* and stimulate the tumor formation *in vivo* [31]. *LGALS1* is best known for its role in RAS signaling, and is associated with shorter disease-free survival and increased blasts [32]. Moreover, *LGALS1* mediates immune evasion by preventing T cell migration into the tumor [33]. Although S100As are significantly associated with poor prognosis in low-grade glioma patients as reported [34], the prognostic and oncologic values of 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 related to chemoresistance in AML, and TCGA data confirmed this gene signature program were associated with poor survival of AML.

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 AML patients 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 are 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 AML patients. Our results provide insight into the aberrant gene programs of AML CD34⁺ cells comparing with the normal, reveal a striking difference between “resistant group” and “sensitive group”, and identify the specific cell populations correlated with chemoresistance and chemosensitivity.

Conclusion

It is the first discovery for clusters of CD34⁺ cells in newly diagnosed AML patients which is related with the response of different individuals to induction chemotherapy. Our data and findings can predict the response of AML patients to induction treatment at diagnosis phase and guide therapeutic strategies to target critical and specific cell components of chemoresistance.

Abbreviations

AML: Acute myeloid leukemia; HSCP: hematopoietic stem/progenitor cell; scRNA-seq: single cell RNA sequencing; CR: complete remission; NR: None remission; PR: Partial remission; HSC: hematopoietic stem cells; MLP: multilymphoid progenitor; MEP: megakaryocyte-erythroid progenitor; GMP: granulocyte-monocyte progenitor; ProB: pro-B cell; ETP: earliest thymic progenitors; SD: standard deviation; UMI:

unique molecular identifier; CNV: Copy number variation; LSC: leukemia stem cell; Ara-C: cytarabine; HMA: hypomethylating agent; 5-FU: 5-Fluorouracil

Declarations

Ethics approval and consent to participate

Informed consent was obtained in accordance with the Declaration of Helsinki. All laboratory experiments with primary samples procedures were reviewed and approved by the Medical Ethics Committee of Qilu Hospital of Shandong University.

Consent for publication

All authors have agreed to the publication of this manuscript.

Availability of data and materials

All data available within the article and supplementary files, or available from the author upon request.

Competing interests

The authors declare that they have no competing interests.

Funding

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), the Shandong Provincial Natural Science Foundation (ZR2020MH118, ZR2020KH016, ZR2020MH119).

Author contributions:

MJ, JZ, JY designed the research; RJ, MJ, GL, SZ performed the experiments, analyzed data, and wrote the manuscript; PL, YS, FL collected clinical samples and information; RJ, YX, SG did bioinformatic analysis; DM, FX, SY provided scientific and technical support; TS, CJ supervised and coordinated all aspects of the research. All authors revised, read and approved the final manuscript.

Acknowledgments

No applicable

Authors' information

¹ Department of Hematology, Qilu Hospital, Cheeloo College of Medicine, Shandong University, Jinan 250012, People's Republic of China; ² Shandong Key Laboratory of Immunohematology, Qilu Hospital, Shandong University, Jinan 250012, People's Republic of China; ³ Department of Biostatistics, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan 250002, People's Republic of China; ⁴ National Institute of Health Data Science of China, Shandong University, Jinan 250002, People's Republic of China

References

1. Schwind, S., G. Marcucci, K. Maharry, M.D. Radmacher, K. Mrózek, K.B. Holland, 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): 5660-9.
2. Attar, E.C., J.L. Johnson, P.C. Amrein, G. Lozanski, M. Wadleigh, D.J. DeAngelo, 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): 923-9.
3. Morita, K., H.M. Kantarjian, F. Wang, Y. Yan, C. Bueso-Ramos, K. Sasaki, 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.
4. Tallman, M.S., D.G. Gilliland, and J.M. Rowe, Drug therapy for acute myeloid leukemia. *Blood* 2005; 106(4): 1154-63.
5. Büchner, T., W. Hiddemann, B. Wörmann, H. Löffler, W. Gassmann, T. Haferlach, 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(12): 4116-24.
6. Loken, M.R., T.A. Alonzo, L. Pardo, R.B. Gerbing, S.C. Raimondi, B.A. Hirsch, 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(8): 1581-8.
7. Pollyea, D.A. and C.T. Jordan, Therapeutic targeting of acute myeloid leukemia stem cells. *Blood* 2017; 129(12): 1627-1635.
8. Tremblay, C.S., J. Saw, S.K. Chiu, N.C. Wong, K. Tsyganov, S. Ghotb, et al., Restricted cell cycle is essential for clonal evolution and therapeutic resistance of pre-leukemic stem cells. *Nat Commun* 2018; 9(1): 3535.

9. Drissen, R., N. Buza-Vidas, P. Woll, S. Thongjuea, A. Gambardella, A. Giustacchini, et al., Distinct myeloid progenitor-differentiation pathways identified through single-cell RNA sequencing. *Nat Immunol* 2016; 17(6): 666-676.
10. Paul, F., Y. Arkin, A. Giladi, D.A. Jaitin, E. Kenigsberg, H. Keren-Shaul, et al., Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell* 2016; 164(1-2): 325.
11. Giustacchini, A., S. Thongjuea, N. Barkas, P.S. Woll, B.J. Povinelli, C.A.G. Booth, et al., Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. *Nat Med* 2017; 23(6): 692-702.
12. Savas, P., B. Virassamy, C. Ye, A. Salim, C.P. Mintoff, F. Caramia, 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.
13. Mathys, H., J. Davila-Velderrain, Z. Peng, F. Gao, S. Mohammadi, J.Z. Young, et al., Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* 2019; 570(7761): 332-337.
14. Witkowski, M.T., I. Dolgalev, N.A. Evensen, C. Ma, T. Chambers, K.G. Roberts, et al., Extensive Remodeling of the Immune Microenvironment in B Cell Acute Lymphoblastic Leukemia. *Cancer Cell* 2020; 37(6): 867-882 e12.
15. Zheng, G.X., J.M. Terry, P. Belgrader, P. Ryvkin, Z.W. Bent, R. Wilson, et al., Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 2017; 8: 14049.
16. Laurenti, E., S. Doulatov, S. Zandi, I. Plumb, J. Chen, C. April, et al., The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. *Nat Immunol* 2013; 14(7): 756-63.
17. Yagi, T., A. Morimoto, M. Eguchi, S. Hibi, M. Sako, E. Ishii, et al., Identification of a gene expression signature associated with pediatric AML prognosis. *Blood* 2003; 102(5): 1849-56.
18. Zhang, H., S. Savage, A.R. Schultz, D. Bottomly, L. White, E. Segerdell, et al., Clinical resistance to crenolanib in acute myeloid leukemia due to diverse molecular mechanisms. *Nat Commun* 2019; 10(1): 244.
19. Kantarjian, H., T. Kadia, C. DiNardo, N. Daver, G. Borthakur, E. Jabbour, et al., Acute myeloid leukemia: current progress and future directions. *Blood Cancer J* 2021; 11(2): 41.
20. Papaemmanuil, E., M. Gerstung, L. Bullinger, V.I. Gaidzik, P. Paschka, N.D. Roberts, et al., Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med* 2016; 374(23): 2209-2221.
21. Levine, J.H., E.F. Simonds, S.C. Bendall, K.L. Davis, A.D. Amir el, M.D. Tadmor, et al., Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell* 2015;

162(1): 184-97.

22. Eberwine, J., J.Y. Sul, T. Bartfai, and J. Kim, The promise of single-cell sequencing. *Nat Methods* 2014; 11(1): 25-7.
23. Stegle, O., S.A. Teichmann, and J.C. Marioni, Computational and analytical challenges in single-cell transcriptomics. *Nat Rev Genet* 2015; 16(3): 133-45.
24. van Galen, P., V. Hovestadt, M.H. Wadsworth li, T.K. Hughes, G.K. Griffin, S. Battaglia, et al., Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. *Cell* 2019; 176(6): 1265-1281 e24.
25. Jiang, L., X.P. Li, Y.T. Dai, B. Chen, X.Q. Weng, S.M. Xiong, et al., Multidimensional study of the heterogeneity of leukemia cells in t(8;21) acute myelogenous leukemia identifies the subtype with poor outcome. *Proc Natl Acad Sci U S A* 2020; 117(33): 20117-20126.
26. Pei, S., D.A. Pollyea, A. Gustafson, B.M. Stevens, M. Minhajuddin, R. Fu, et al., Monocytic Subclones Confer Resistance to Venetoclax-Based Therapy in Patients with Acute Myeloid Leukemia. *Cancer Discov* 2020; 10(4): 536-551.
27. Riether, C., T. Pabst, S. Hopner, U. Bacher, M. Hinterbrandner, Y. Banz, 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.
28. Duy, C., M. Li, M. Teater, C. Meydan, F.E. Garrett-Bakelman, T.C. Lee, et al., Chemotherapy Induces Senescence-Like Resilient Cells Capable of Initiating AML Recurrence. *Cancer Discov* 2021; 11(6): 1542-1561.
29. Granja, J.M., S. Klemm, L.M. McGinnis, A.S. Kathiria, A. Mezger, M.R. Corces, et al., Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat Biotechnol* 2019; 37(12): 1458-1465.
30. Jin, S., R. Li, M.Y. Chen, C. Yu, L.Q. Tang, Y.M. Liu, 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.
31. Zhang, L., R. Zhou, W. Zhang, X. Yao, W. Li, L. Xu, 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(1): 120.
32. Ruvolo, P.P., H. Ma, V.R. Ruvolo, X. Zhang, S.M. Post, and M. Andreeff, LGALS1 acts as a pro-survival molecule in AML. *Biochim Biophys Acta Mol Cell Res* 2020; 1867(10): 118785.

33. Nambiar, D.K., T. Aguilera, H. Cao, S. Kwok, C. Kong, J. Bloomstein, et al., Galectin-1-driven T cell exclusion in the tumor endothelium promotes immunotherapy resistance. *J Clin Invest* 2019; 129(12): 5553-5567.

34. Zhang, Y., X. Yang, X.L. Zhu, H. Bai, Z.Z. Wang, J.J. Zhang, et al., S100A gene family: immune-related prognostic biomarkers and therapeutic targets for low-grade glioma. *Aging (Albany NY)* 2021; 13(11): 15459-15478.

Tables

Due to technical limitations, table 1 and 2 is only available as a download in the Supplemental Files section.

Figures

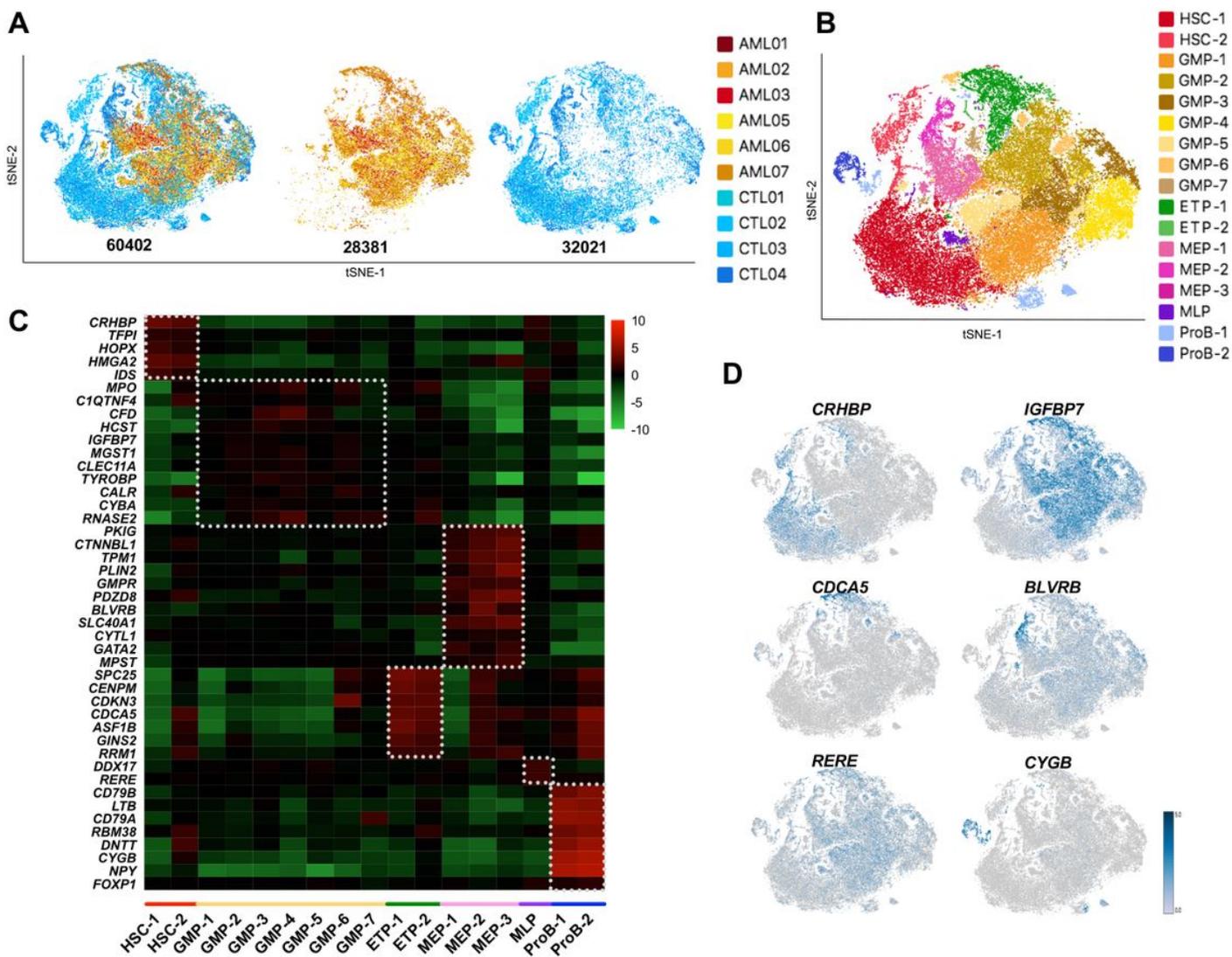


Figure 1

Identification of cell populations in CD34+ cells from newly diagnosed AML patients. (A) t-distributed Stochastic Neighbor Embedding (tSNE) visualization of 60,402 cells from 6 newly diagnosed AML patients and 4 healthy controls (CTL). (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 tSNE representation.

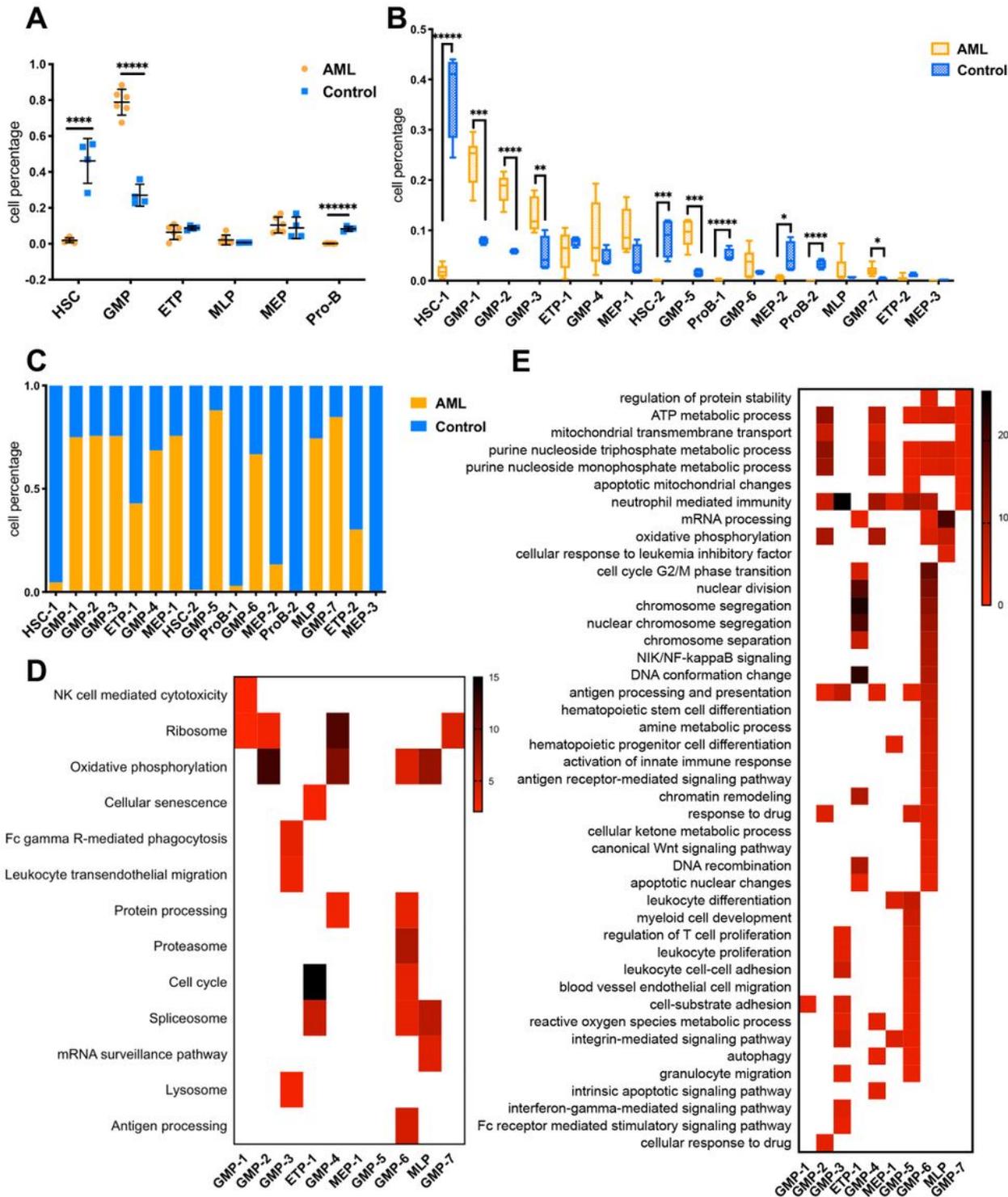


Figure 2

CD34+ cells of AML patients at diagnosis enriches in GMP population. (A) Boxplot showing fraction of HSC, GMP, MEP, MLP, ProB and ETP in AML patients and controls. (B) Boxplot showing fraction of each clusters in AML patients and controls. (C) Histogram showing percentage of each clusters between AML patients and controls after normalizing baseline to 100%. (D) Significantly KEGG pathways enrichment analysis of molecular signatures in malignant-like clusters. (E) Significantly GO pathways enrichment analysis of molecular signatures in malignant-like clusters. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$, ***** $P < 0.00001$, ***** $P < 0.000001$.

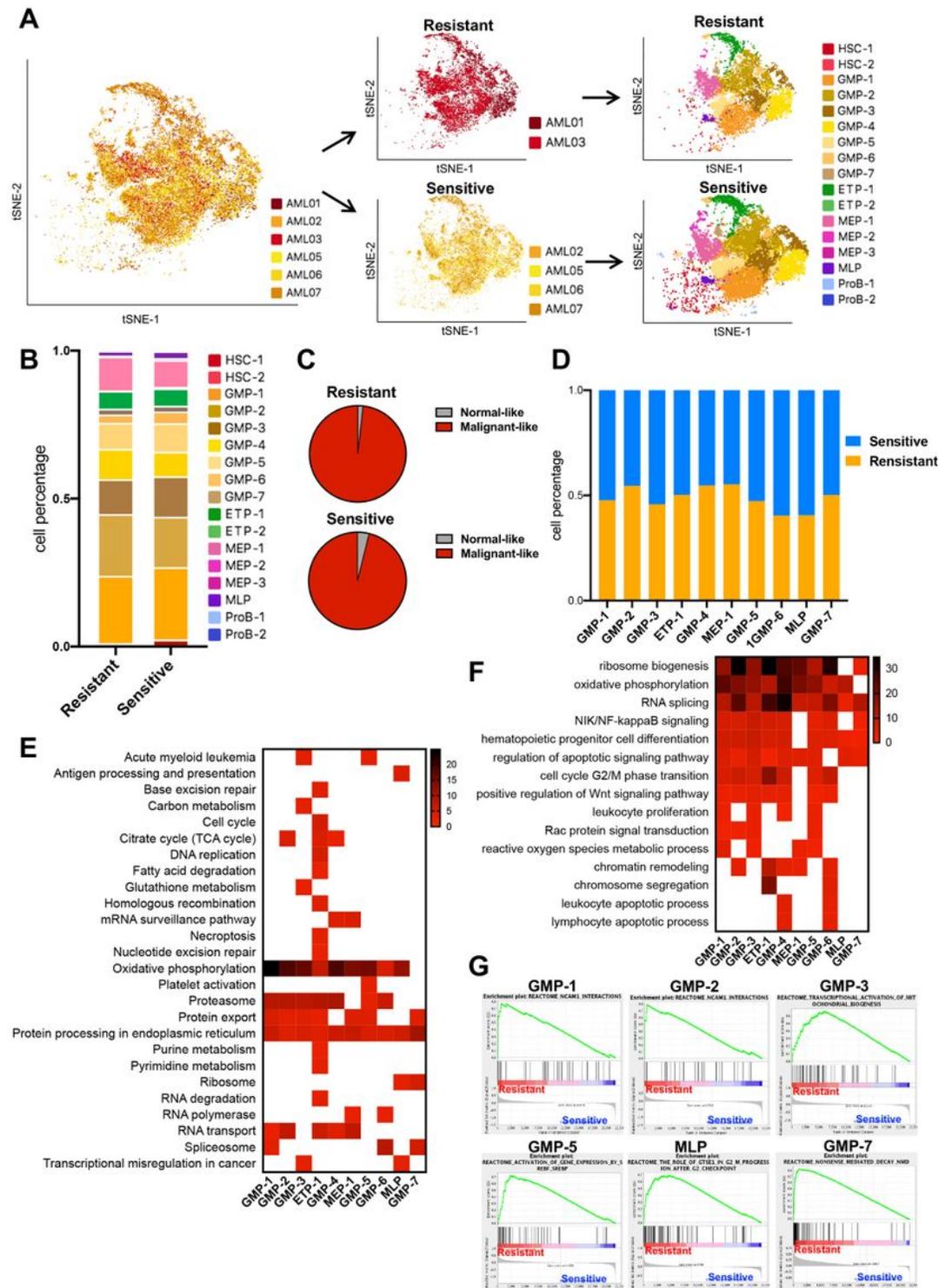


Figure 3

Heterogeneity of malignant-like clusters in CD34+ cells of AML. (A) tSNE visualization of CD34+ cells from 6 newly diagnosed AML patients (left), resistant and sensitive groups (medium), and marker-based cell type identification analysis including 6 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, 7 and MLP clusters.

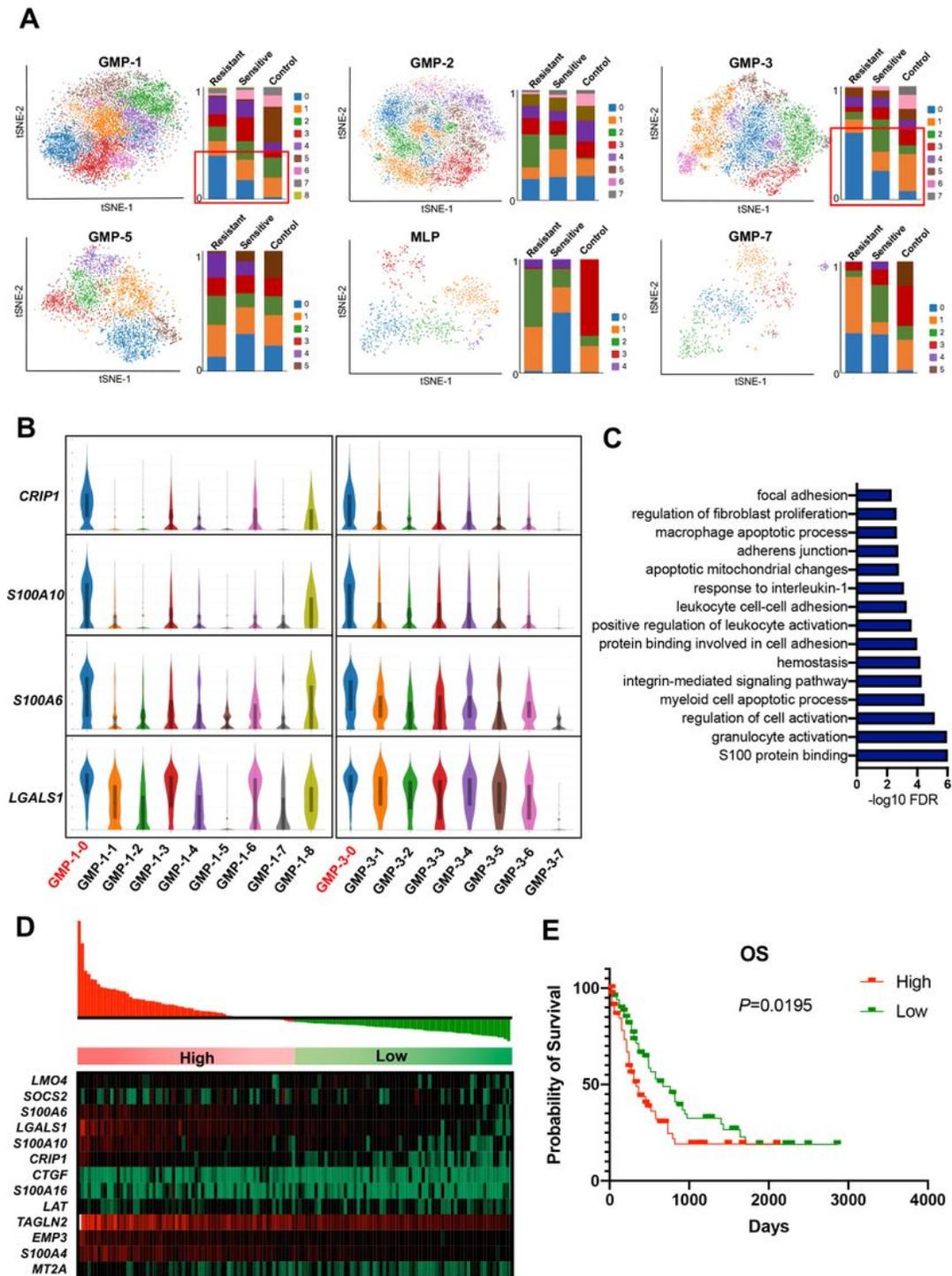


Figure 4

Subclusters of CD34+ cells associated with poor prognosis in AML. (A) tSNE visualization of the subgrouped results in 6 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) 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 AML patients with high and low scores of the signature genes. P value calculated by log-rank test.

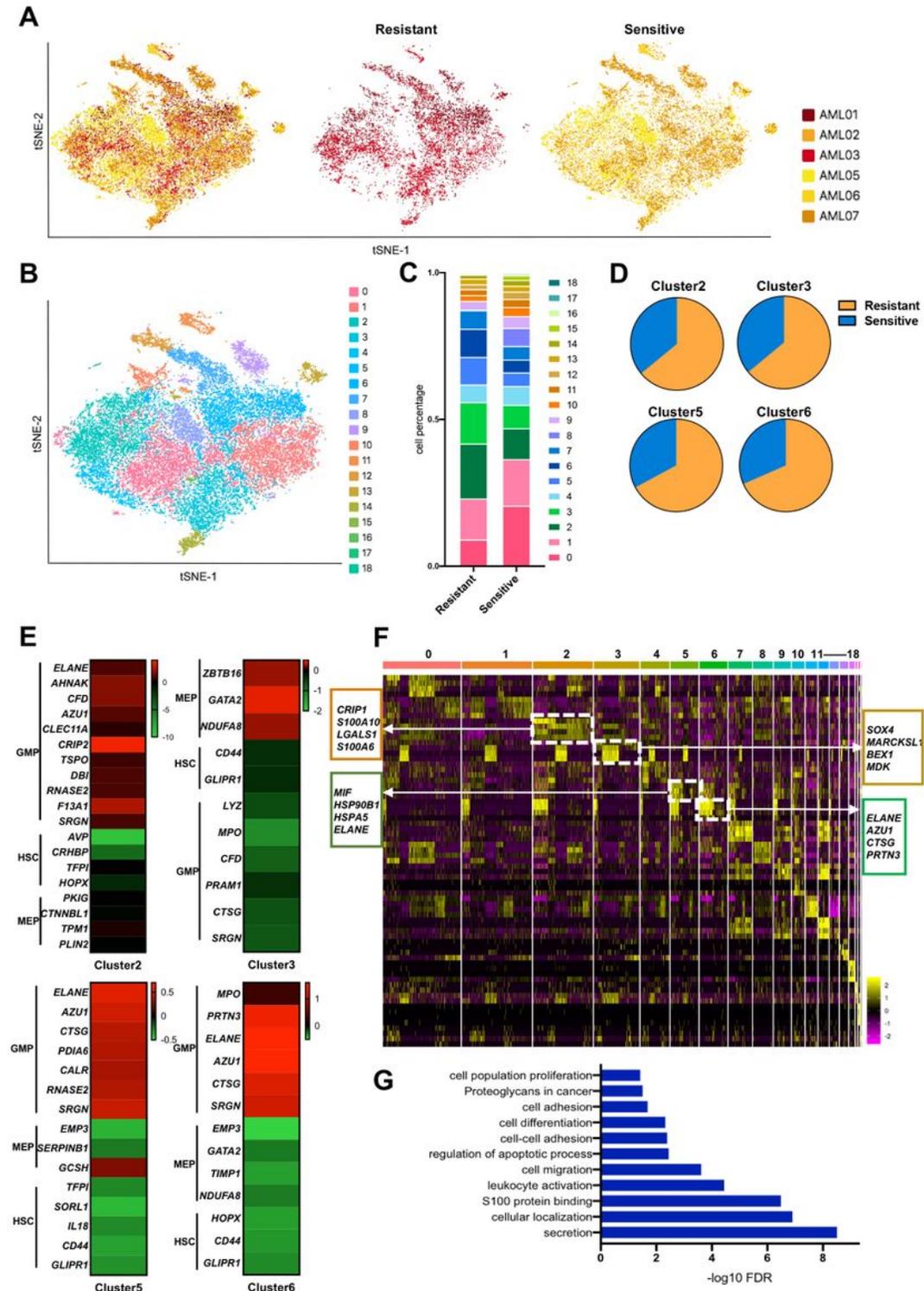


Figure 5

Regrouping of AML CD34+ cells and re-appearance of the CRIP1highLGALS1highS100Ashigh cluster. (A) tSNE visualization of 28,381 cells from 6 newly diagnosed AML patients. Total (left), resistant group

(medium) and sensitive group (right). (B) tSNE 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 cluster 2, 3, 5 and 6. (E) Gene expression heatmap of the cell-type-specific marker genes in cluster 2, 3, 5, 6. (F) Heatmap showing relative expression of top marker genes in 19 clusters and emphasized marker genes in cluster 2, 3, 5, 6. (G) Significantly KEGG pathways enrichment analysis of molecular signatures in cluster 2.

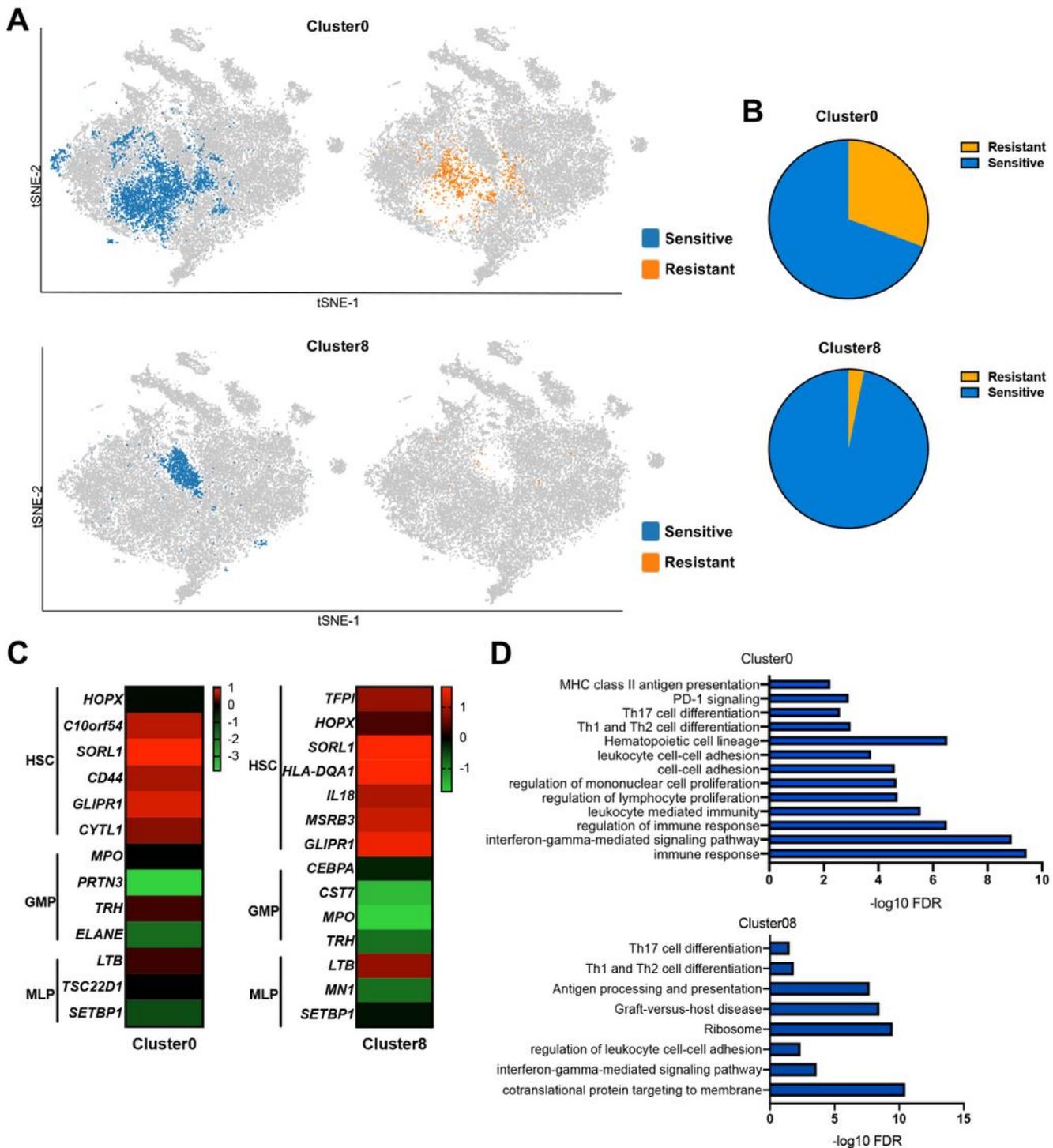


Figure 6

Subclusters of CD34+ cells associated with good response to induction chemotherapy in AML. (A) tSNE 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) Significantly KEGG pathways enrichment analysis of molecular signatures in cluster 0 and cluster 8.

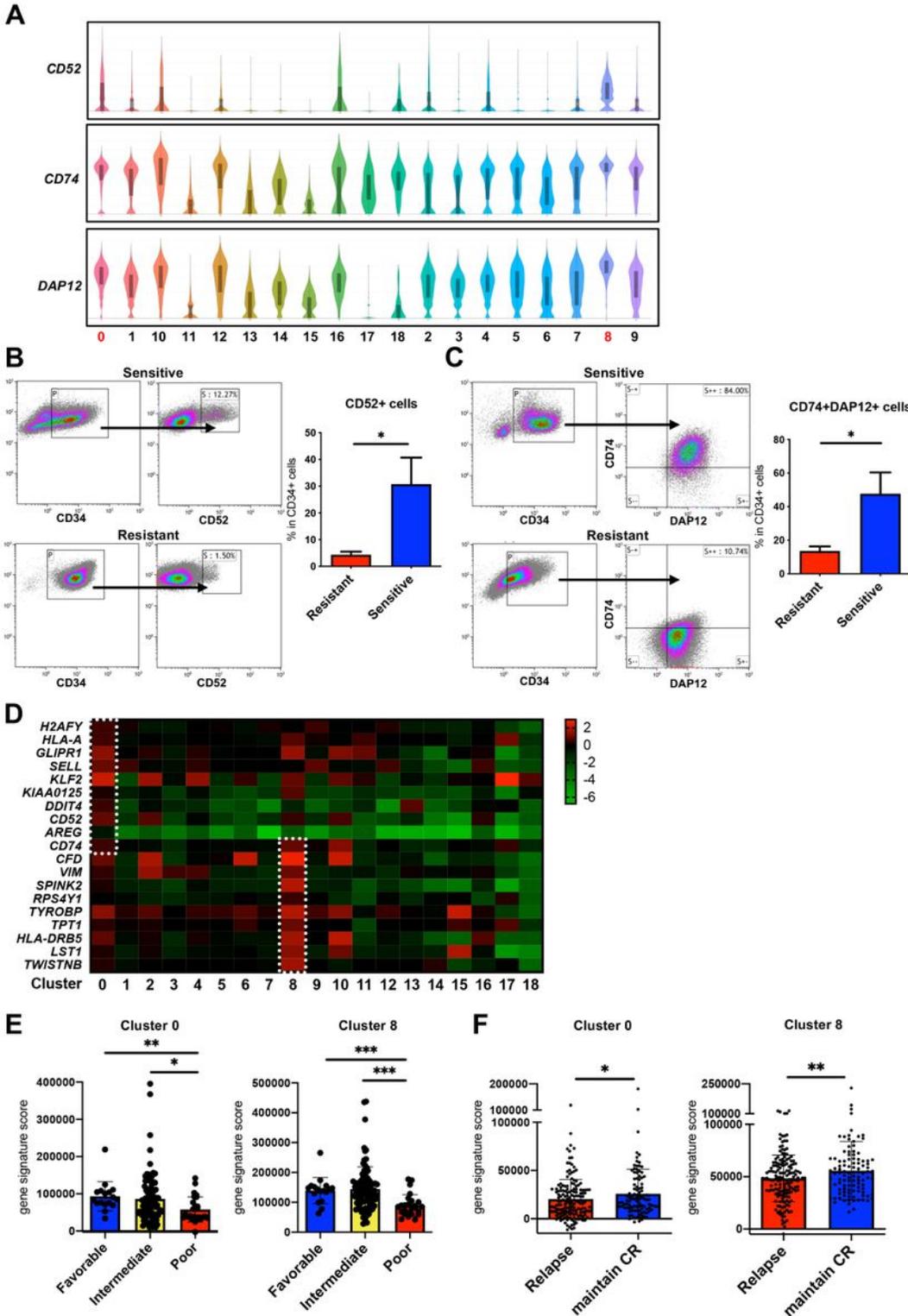


Figure 7

Cluster 0 and 8 confer outcomes of newly diagnosed AML patients. (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 12 newly diagnosed AML patients before treatment, 6 of whom achieved CR (Sensitive) and 6 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 AML patients. Typical flow cytometric plots and histograms were shown. (D) Heatmap showing relative expression of top marker genes in cluster 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 AML patients from 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 AML patients from GSE165430. * P < 0.05, ** P < 0.01, *** P < 0.001

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Table1.xlsx](#)
- [Supplementalfiles.doc](#)
- [Table2.xlsx](#)