Research Paper

# Single cell sequencing reveals cell populations that predict primary resistance to imatinib in chronic myeloid leukemia

**Weilong Zhang[1,2,*], Beibei Yang[2,*], Linqian Weng[3], Jiangtao Li[3], Jiefei Bai[3], Ting Wang[3], Jingwen Wang[4], Jin Ye[4], Hongmei Jing[1], Yuchen Jiao[2], Xixi Chen[5,6], Hui Liu[3], Yi-Xin Zeng[2,7]**

[1]Department of Hematology, Lymphoma Research Center, Peking University Third Hospital, Beijing 100191, China
[2]State Key Lab of Molecular Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China
[3]Department of Hematology, Beijing Hospital, National Center of Gerontology, Beijing, China
[4]Department of Hematology, Beijing Tongren Hospital, Capital Medical University, Beijing, China
[5]Genetron Health (Beijing) Co. Ltd., Beijing 102206, China
[6]Children's Hospital of Chongqing Medical University, Chongqing 400014, China
[7]Department of Experimental Research, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, Guangdong Province, China
*Equal contribution

**Correspondence to:** Hui Liu, Xixi Chen, Yi-Xin Zeng; **email:** liuhui3218@bjhmoh.cn, xiqian.chen@genetronhealth.com, zengyx@sysucc.org.cn
**Keywords:** chronic myeloid leukemia, peripheral immune structure, single cell sequencing, TKI resistance, stem cells
**Received:** June 6, 2020 **Accepted:** September 20, 2020 **Published:** November 23, 2020



## ABSTRACT

**The treatment of chronic myeloid leukemia (CML), a disease caused by t(9;22)(q34;q11) reciprocal translocation, has advanced largely through the use of targeted tyrosine kinase inhibitors (TKIs). To identify molecular differences that might distinguish TKI responders from non-responders, we performed single cell RNA sequencing on cells (n = 41,723 cells) obtained from the peripheral blood of four CML patients at different stages of treatment to generate single cell expression profiles. Analysis of our single cell expression profiles in conjunction with those previously obtained from the bone marrow of additional CML patients and healthy donors (total = 69,263 cells) demonstrated that imatinib treatment significantly altered leukocyte population compositions in both responders and non-responders, and affected the expression profiles of multiple cell populations, including non-neoplastic cell types. Notably, in imatinib poor-responders, patient-specific pre-treatment unique stem/progenitor cells became enriched in peripheral blood compared to the responders. These results indicate that resistance to TKIs might be intrinsic in some CML patients rather than acquired, and that non-neoplastic immune cell types may also play vital roles in dispersing the responsiveness of patients to TKIs. Furthermore, these results demonstrated the potential utility of peripheral blood as a diagnostic tool in the TKI sensitivity of CML patients.**

## INTRODUCTION

Chronic myeloid leukemia (CML) is a hematopoietic stem cell disease caused by a single translocation event, t(9;22)(q34;q11), which generates the fusion protein BCR-ABL. CML responds to treatment with targeted tyrosine kinase inhibitors (TKI), such as imatinib and dasatinib, which bind to the active site of BCR-ABL. However, some CML patients fail to benefit from TKI therapy, and 3-5% progress into blast crisis after

treatment [1–3]. Underlying mechanisms for resistance to TKI treatment have been associated with genomic aberrations in addition to the BCR-ABL translocation, including mutations in the fusion protein and triosomy of chr8 or chr17p [4, 5].

Immune function in CML is impaired as in other malignancies. For example, Rossignol et al. [6] reported dysfunctional invariant natural killer T cells (iNKT) in untreated CML patients, and Chen et al. [7] also observed a decreased proportion of NK cells in the peripheral blood that did not recover after treatment with TKIs. In addition, immune suppressive cells, such as Tregs and myeloid-derived suppressor cells (MDSCs), were also reported to increase in high-risk populations [8]. Despite these intriguing results, the population structure of the peripheral blood has not been well characterized in CML patients. The extent to which the population structure of the peripheral blood differs among patients, whether these differences have an impact on the response to TKI treatment and how the TKI treatment alters the immune system have also not yet been fully elucidated. In fact, an immune-modulatory role for TKI treatment has been previously observed in studies examining the frequencies of different immune compartments in the peripheral blood [9]. Yet, the specific changes in the expression programs remain a mystery.

Recent advances in single cell analysis have challenged and revised the trajectory map of hematopoiesis [10–12]. Thus, to refine the immune cell population structure at a higher resolution in CML might lead to further insight into alternative therapies. Giustacchini et al. and Warfvinge et al. [13, 14] analyzed single-cell transcriptomes of CML cancer stem cells in bone marrow and identified subpopulations and gene markers of TKI therapy-resistance. However, there are some limitations of these studies. First, these studies examined only a subgroup of CD34+ cells rather than whole CML cell populations. Second, only a small number of cells were studied (~ 2,000 cells), which limited cell classification. Finally, cells were only characterized from the bone marrow, and not the peripheral blood, which can be used as a noninvasive biopsy for monitoring of CML. Here, we performed single cell RNA-sequencing on a total of 41,723 cells from peripheral blood, covering different TKI treatment stages of four CML patients, to generate expression profiles for individual cells. We furthermore analyzed this data in association with expression profiles of 17,540 bone marrow cells from both CML patients and healthy donors to generate a comprehensive landscape for CML. Using the combined datasets, we investigated the immune cell structure in parallel with CML progression, and characterized the molecular/cell signature of the immune response to TKI treatment. Our study provides insight into the pathogenic mechanisms involved in CML beyond BCR-ABL translocation and new therapeutic strategies to complement TKI.

## RESULTS

### Comparison of the expression profiles of primitive stem/progenitor populations in peripheral blood from CML patients with healthy bone marrow components

To investigate the comprehensive immune cell composition in the peripheral blood of CML patients, we performed single cell RNA sequencing (scRNA-seq) on PBMCs collected from four representative patients with CML at multiple time points before (BT) and after treatment with imatinib. Despite treatment with imatinib, one patient (P04) rapidly progressed to the blast phase within three months, and had acquired a mutation in the kinase domain region of BCR-ABL (p.M448V). This patient received dasatinib for the second round of treatment. PBMC samples from P04 in the blast phase as well as post-blast treatment (blast cells (BC)-BT and BC-AT, respectively) were also included. We also obtained PBMCs from a single healthy donor (N). These single-cell suspensions were subjected to scRNA-seq using barcodes and UMIs for individual cells and unique transcript counting. After filtering out low-quality cells, we obtained a total of 41,723 cells from peripheral blood with at least 200 detected genes (Supplementary Table 2). The cell numbers obtained from different stages including all individuals were the following: normal, 5,082; BT, 5,562; AT, 24,815; BC-BT, 4,515; and BC-AT, 1,749.

Unsupervised clustering of the expression data obtained from all cells revealed a total of 11 clusters (Figure 1A). Through examination of known cell type specific markers, eight clusters were classified as functionally common immune cell types, including CD4+ T cells, CD8+ T cells, NK cells, B cells, CD14+ monocytes, CD16+ monocytes, and megakaryocytes (Supplementary Figure 1A). Sample-of-origin of cells in these immune cell compartments were well mixed (Figure 1B, Supplementary Figure 1B), and the fractions of each compartment in the healthy blood sample were consistent with the previously reported composition of PBMCs, thus confirming the clustering methods and results (Supplementary Figure 1C).

The most remarkable result was the identification of four clusters with apparent features of primitive cells, including Clu-CD34, Clu-MPO, Clu-MME and a subset of erythrocytes ($GATA1^{high}$) (Figure 1C). These four clusters exhibited a significant enrichment in the patients

with poor prognosis, such as Clu-CD34 and Clu-MPO in P03 at diagnosis, early-stage erythrocytes (GATA1$^{high}$) in P03 and P04 at diagnosis, and Clu-MME in P04 at the blast-crisis stage (Figure 1B, Supplementary Figure 1D. To further understand the biological status of these primitive cells, we used an expression dataset representing healthy Lin- bone marrow cells (n=17,540) as a comprehensive reference (referred to as BM-reference) [15]. Visualization using uniform manifold approximation and projection (UMAP) effectively recapitulated the intermediate clusters from our analysis during the continuous development process (Figure 1D, BM-1 to BM-11). We then mapped the primitive cells from peripheral blood onto the BM-reference to understand the hierarchy of these cells (Figure 1E). Clu-CD34 correlated with a collection of early stem cells with heterogenous differentiation destinies. The blended lineage potentials in this cluster were also confirmed using the lineage-specific signatures defined in a separate study (Supplementary Figure 1E). The primitive cell cluster Clu-MPO contained cells from the earliest myeloid progenitors to neutrophil-defined or monocyte-defined progenitors, while the Clu-MME cluster represented early B cell progenitors.

Since BCR-ABL fusion is usually considered to be associated with CD34+CD38- stem cells, we focused on the primitive cell cluster Clu-CD34 and compared the expression profile of this cluster with early HSCs and their immediate progenies from the reference dataset (BM-1 and BM-7) to identify differentially expressed genes (DEG). Up-regulated DEGs included *LGALS1*, *MALAT1*, *TGFB1*, *MZB1*, while down-regulated DEGS included *KIT*, *PTEN*, and *CXCR4* (Figure 1F). GSEA analysis revealed that inflammation signatures (interferon signaling, TNF signaling) were significantly up-regulated in Clu-CD34 which was consistent with an enhanced inflammatory response in these patients (Figure 1G). These same inflammation signatures were also associated with Clu-MPO (Supplementary Figure 1F).

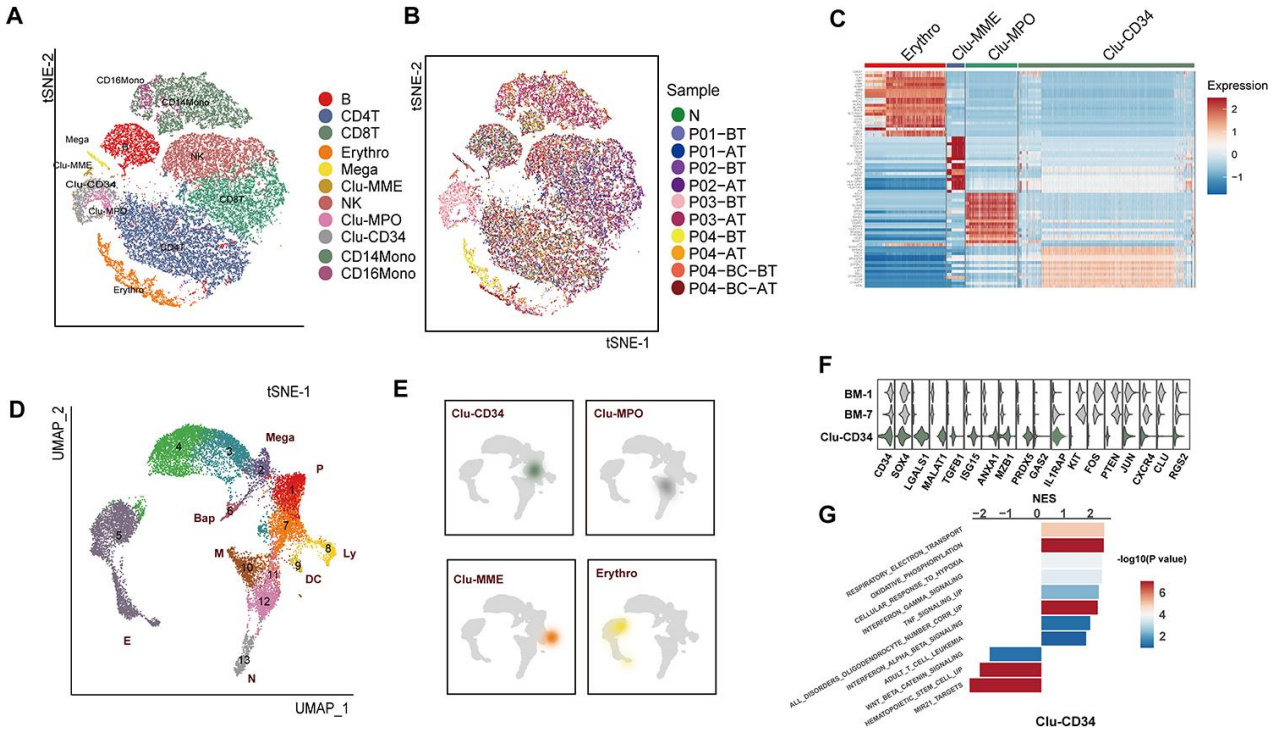



**Figure 1. Comparison of the primitive stem/progenitor populations in peripheral blood with the healthy BM components.** (**A**) TSNE plot for the resultant 11 clusters identified based on single cell RNA sequencing. The total number of cells is 41,723. Clusters are highlighted by different colors, and the number of cells in each cluster is listed in Supplementary Table 2. (**B**) TSNE plot of all cells. Colors indicate sample origin. (**C**) Heatmap showing the expression profiles of the four primitive clusters identified in peripheral blood. Expression of the scaled value of the top 30 significant marker genes in each cluster are shown. (**D**) UMAP plot displaying the resultant clusters identified in healthy Lin- bone marrow datasets [15]. The total number of cells is 15,253. Colors indicate clusters, and lineage destinations are labeled. P, early progenitor cells; Meg, megakaryocytes; (**E**) erythroid cells; BaP, basophil progenitors; N, neutrophils; M, monocytes; DC, dendritic cells; Ly, lymphoid cells. (**E**) Kernel density plot showing the projection result of the four primitive clusters identified in peripheral blood onto the BM reference map. (**F**) Violin plot showing the expression distribution of selected genes in Clu-CD34 in comparison with BM-1 and BM-7. (**G**) Bar plot displaying the GSEA result on the ordered expression profiles in Clu-CD34. X-axis indicates the normalized enrichment score (NES) and colors indicate the -log10(*P* value).

Clu-CD34 and Clu-MPO were mainly composed of cells from the non-responder P03 at diagnosis. Most of the altered expression signatures identified in these two clusters were consistent with TKI non-responding signatures identified in previous single cell studies performed on bone marrow samples from CML patients. For example, Giustacchini et al. [13]. observed enrichment of signatures related to inflammation, TGF-beta and TNF-alpha in BCR-ABL- stem cells at diagnosis from poor relative to good responders. In addition, a subgroup of BCR-ABL+ stem cells with selective persistence during TKI treatment was found exhibiting increased expression of *TGFB1*, *GAS2*, *CTNNB1*, and *HIF1A* but reduced expression of *CXCR4*, which was consistent with our observations. Finally, TNF-alpha and TGF-beta signaling became progressively enriched in this subgroup during the course of treatment. Warfvinge et al. [14]. also found that the most TKI-insensitive cells were negative for cKIT. Results from all the datasets suggested that despite the extensive heterogeneity that was present among CML patients, some of the features were shared by TKI-insensitive cells in the bone marrow, and these features could also be detected in the peripheral blood of some patients.

**Integration analysis of bone marrow and peripheral blood primitive cells in CML patients**

We next combined the primitive clusters in peripheral blood and the stem cells in bone marrow (CML-reference) [13] to illustrate the variation from BM to peripheral blood in CML patients. The CML-reference contained 2,287 CD34+CD38- cells from 34 CML patients and 6 healthy controls. Integration of Clu-CD34 and Clu-MPO with the CML-reference revealed both well-mapped and unmapped cells (Figure 2A). Using specific cell type markers, we identified that the well-mapped cells included three groups of cells: proliferating, megakaryocyte/erythrocyte progenitors (MEP) and erythroid progenitors (clustered together with K562 cells) (Figure 2B). Interestingly, we also found that a small subset of the cells in Clu-CD34 grouped together with the cells at the blast crisis stage within the CML-reference (Figure 2B). This observation may provide evidence for the existence of CML stem cells at diagnosis which already have potential for seeding blast crisis. However, two groups of cells clustered separately from the CML-reference, including a subset of Clu-CD34 cells and most Clu-MPO cells. The unmapped subset in Clu-CD34 uniquely expressed several leukemia-related genes (Figure 2C), one of which was *CD99*. This gene is known for playing key roles in promoting the mobilization of the hematopoietic cells [16–18] and has been recognized as a robust marker and promising therapeutic target in several tumor types [19]. Through inspection of known ligand-receptor pairs, we found that Clu-CD34 could possibly affect other immune cell types through elevated interaction between CD99 and PILRA (Figure 2D). Finally, *HOXA9*, which is frequently overexpressed in acute myeloid leukemia and may promote leukemia through epigenetic landscape remodeling [20, 21], was up-regulated.

Erythropoiesis is one of the major hallmarks of CML. In the peripheral blood dataset, we did observe a large number of erythroid cells with extremely high expression of *HBA1/2* and *HBB*, although most of these cells were filtered out due to the low number of detected genes. However, among all the erythroid cell types, the most primitive erythroid progenitors ($GATA1^{high}$) originated overwhelmingly in the non-responder P03 and the fast-BC-transforming patient P04 at diagnosis (Ery-TMCC2+ from P03-BT and Ery-CA2+ from P04-BT, Figure 3A, 3B). Compared with the $GATA1^{low}$ erythroid cells, these $GATA1^{high}$ cells had higher expression of *ABCG2*, which is an ATP-binding cassette (ABC) transporters and has been implicated as a potential mechanism of primary resistance to TKI (Figure 3C) [22]. Interestingly, cells in the Ery-CA2+ cluster (originated from P04) were predominantly synchronized in the G2/M phase (Figure 3D). In previous studies, cells just entering terminal differentiation have been identified in committed erythroid progenitor initiates, in part due to their synchronization in S phase [23, 24]. This cluster therefore possibly represents erythroid-terminal differentiation initiating populations.

In the bone marrow dataset, cells from the BC phase were separated from those from the chronic phase. In the peripheral blood dataset, we also found that Clu-MME, composed of 184 cells, was significantly enriched for blast phase cells from P04 (Supplementary Figure 1D, Figure 2A). This group of cells expressed common acute lymphocytic leukemia antigen *MME*, and the pro-B cell specific gene *VPREB1*, indicating that a minor subpopulation of P04-BC-BT was a lymphocytic cell type. We then integrated the BC cells from bone marrow and peripheral blood, and observed a well-mixed cell population [13] (Figure 3E). This result indicated that cells at the BC phase had unique expression signatures, and probably exhibited less inter-patient heterogeneity. The exclusively up-regulated genes in BC cells included *SOCS2* and *S100A16*, which are pivotal in promoting progression of leukemia as well as other types of cancer (Figure 3F, 3G) [25–27]. In addition, Clu-MME from peripheral blood also exhibited signatures consistent with a *RUNX1-RUNXT1* translocation, indicating that selection for an additional genomic alteration had occurred in P04, even though it

was represented by only a small fraction of cells (Supplementary Figure 2B). This detection of this translocation in this small cell population also emphasizes the high sensitivity of single cell transcriptome data in the identification of subclonal genetic alterations.

### Validation of the pre-treatment predictive signatures

Clu-CD34, Clu-MPO, Ery-TMCC2, and Ery-CA2 were present in the peripheral blood of patients with a poor-prognosis at diagnosis. These cells showed limited developmental trajectories towards three directions, erythroid, myeloid and BC (Figure 4A). Since they showed unique expression signatures compared with either normal stem cells or peripheral blood cells, we assumed that these primitive clusters could possibly predict the prognosis of CML patients as early as at primary diagnosis. Therefore, we analyzed a transcriptome dataset of 59-patients with known imatinib-treatment response [28], and found that the expression signatures of Clu-CD34 and Clu-MPO had

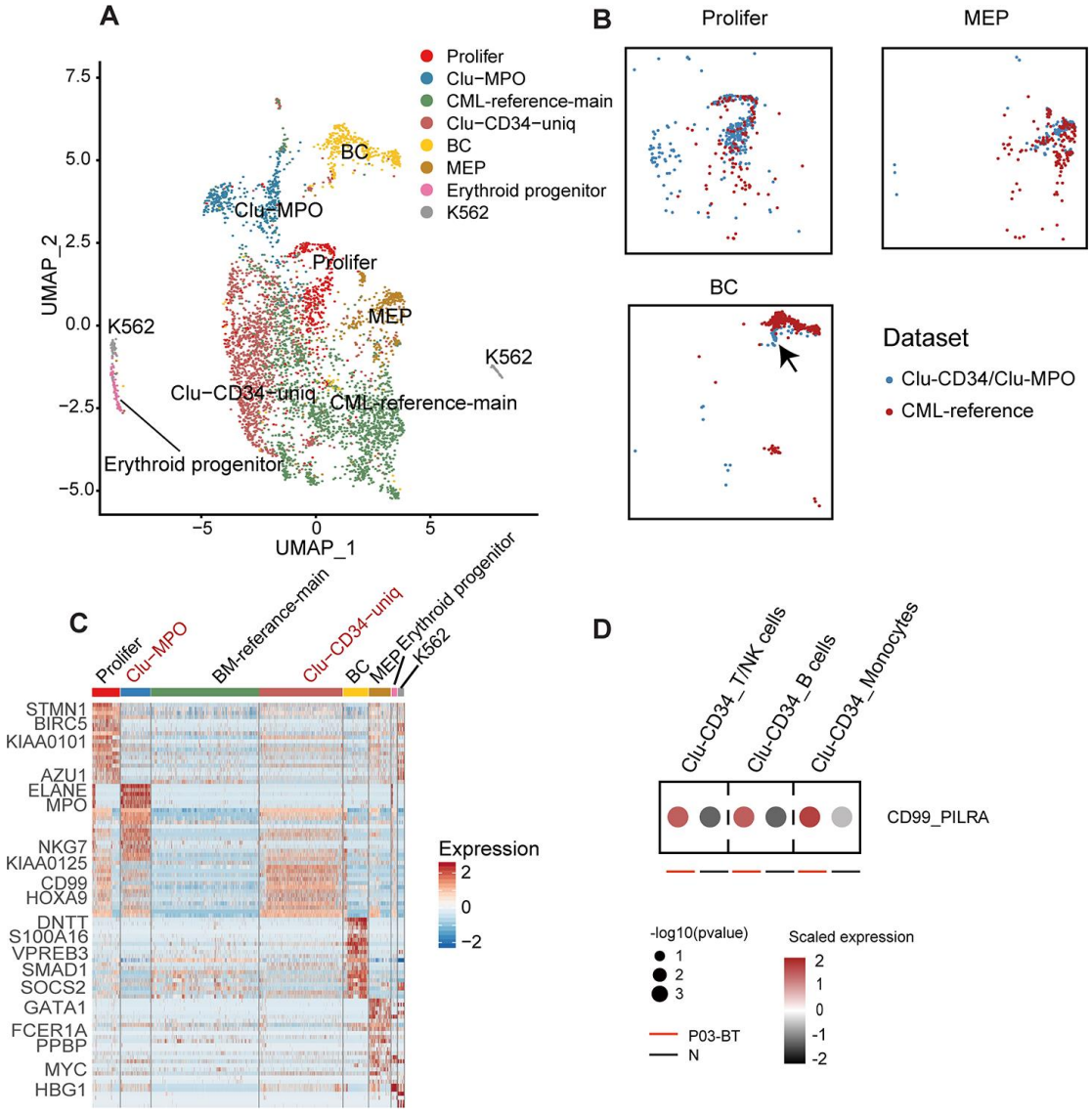



**Figure 2. Comparison of Clu-CD34 and Clu-MPO with bone marrow CML datasets.** (**A**) UMAP plot showing the clustering result of the integrated datasets comprising Clu-CD34, Clu-MPO and the BM-reference [13]. The total number of cells is 4,603 (Clu-CD34: 1,789; Clu-MPO: 527; BM-reference: 2,287). (**B**) UMAP plots of cells. Colors indicate dataset origin. Arrow in the lowest panel indicates the mixture of cells from the Clu-CD34 (n = 61) clustered together with BC cells from the BM-reference. (**C**) Heatmap showing expression of marker genes across different subtypes from the integrated dataset. Two specific subtypes from our dataset (absent in the BM-reference) are marked as red at top. (**D**) Dot plots comparing the interaction of the CD99-PILRA ligand-receptor pair between P03-BT and N. The p value was calculated using a permutation test.

significantly higher correlations with the imatinib non-responders compared with the imatinib responders, (Supplementary Figure 2C; Clu-CD34, P=0.0024, Clu-MPO, P=0.0068; unpaired t test). We generated ROC curves to demonstrate the prediction performance of Clu-CD34 and Clu-MPO gene expression signatures on the dataset. The curves were able to predict resistance to imatinib before initiation of the therapy (Supplementary Figure 2D; Clu-CD34, AUC = 0.74, Clu-MPO, AUC = 0.69).

We analyzed three other independent datasets derived from cohorts of patients (total = 217) treated with imatinib or dasatinib, or in different phases of CML, to achieve a significant scientific conclusion [29–31]. For the CML patients treated with imatinib (n = 96, GSE130404), Clu-CD34 and Clu-MPO gene expression signatures were able to predict early resistance to imatinib (Clu-CD34, AUC = 0.80, Clu-MPO, AUC = 0.82; Figure 4B–4D). When we integrated these two clusters together, the AUC reached 0.86 (95%CI 0.76-0.95; Figure 4E). However, in patients treated with dasatinib (n = 14, GSE33224), Clu-CD34 and Clu-MPO gene expression signatures were not able to predict dasatinib resistance (P-value > 0.05; Supplementary Figure 2E, 2F). These two clusters therefore are of more value in predicting response to imatinib than dasatinib. In the last cohort of CML patients in different phases of the

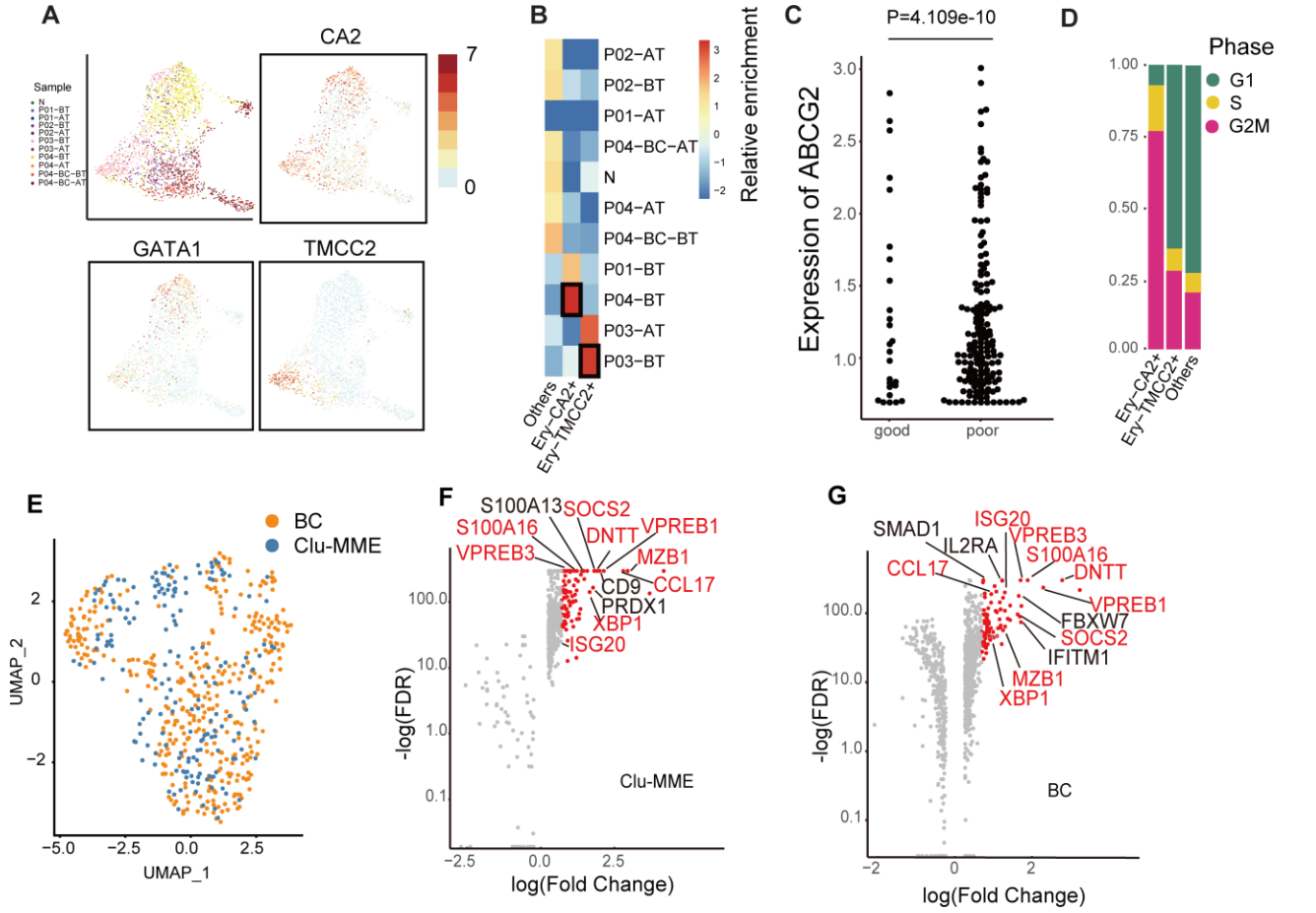



**Figure 3. Comparison of erythrocytes and Clu-MME with bone marrow CML datasets.** (**A**) UMAP plot presenting the re-clustering result of erythrocytes. Cells are highlighted with colors indicating sample origin. The expression of *GATA1*, *CA2* and *TMCC2* are indicated. The *CA2*- and *TMCC2*-expressing erythrocytes are separately enriched in P04-BT and P03-BT. (**B**) Heatmap showing enrichment of samples in each erythroid cluster. Enrichment score was calculated using the Fisher's exact test and indicated by log10(Odd Ratio). (**C**) Beeswarm plots showing the expression of *ABCG2* between good responders (P01 and P02) and poor responders (P03 and P04). P= 4.109e-10, unpaired t test. (**D**) Bar plots showing the fraction of cells from different cell cycle phases across different erythrocyte subtypes. (**E**) UMAP plot showing the integration result of the BC cluster (from the integrated dataset shown in Fig. 2a) and Clu-MME (the numbers of cells in the BC cluster and Clu-MME are 370 and 183, respectively). (**F**) Scatter plot showing the highly-expressed marker genes in Clu-MME (left) and the BC cluster (right). Significant markers (FDR<0.05, fold change > 2) are shown as red dots. The name of shared marker genes of these two clusters are indicated in red. (**G**) Heatmap comparing the expression profiles of CD16+ monocytes across different samples. The selected marker genes are indicated.

disease (n = 107, GSE4170), Clu-CD34 and Clu-MPO gene expression signatures were highly associated with blast crisis phase (P < 0.05; Supplementary Figure 2G, 2H).

### Modulation of the immune structure in response to imatinib treatment

In addition to the primitive cell populations, we investigated changes in the common functional non-neoplastic immune compartments during the treatment course. Fractions of these immune compartments differed significantly among different time points or different response status to imatinib. Within untreated samples (BT), the frequencies of lymphocytes and monocytes were significantly reduced relative to the healthy donor, especially in patients with adverse prognosis (P < 2e-16, Fisher's exact test), which were dominated by stem/progenitor cells (P03) or erythroid cells (P04). Imatinib treatment had an obvious positive

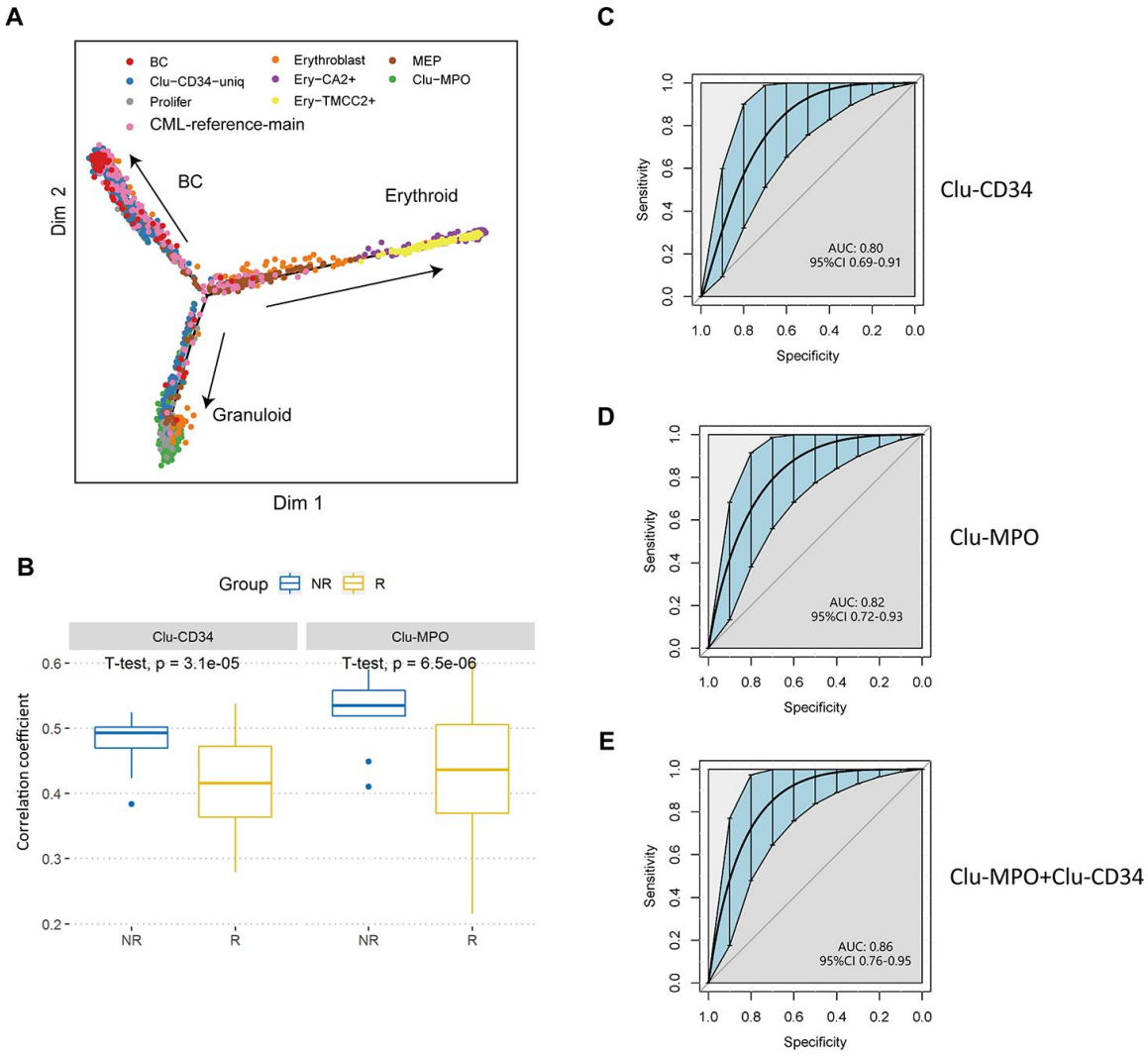



**Figure 4. Validation of prognostic cell populations in 59 CML patients treated with imatinib.** (**A**) Pseudotime trajectory of the primitive cells in peripheral blood. Cells are colored based on their identities in the integrated datasets with CML-reference or the erythroid clusters. (**B**) Box plots comparing the correlation coefficients of Clu-CD34 (left) and Clu-MPO (right) between 13 imatinib nonresponders (NR) and 83 responders (R) from GSE130404. The correlation coefficients were calculated using Pearson correlation of gene expression signatures of these two clusters with the gene expression profile in each CML patient treated with imatinib (see Methods). (**C**, **D**) ROC curves illustrating the classification performance of Clu-CD34 (**C**) and Clu-MPO (**D**) gene expression signatures of 13 imatinib nonresponders and 83 responders from GSE130404. The blue shade denotes the 95% confidence interval of the sensitivity at a given specificity point. AUC, area under the ROC curve are indicated. (**E**) We integrated Clu-CD34 and Clu-MPO clusters together to predict imatinib resistance. ROC curves illustrating the classification performance of gene expression signatures of 13 imatinib nonresponders and 83 responders from GSE130404. The blue shade denotes the 95% confidence interval of the sensitivity at a given specificity point. AUC, area under the ROC curve are indicated.

effect in the restoration of the functional immune structures (Figure 5A). However, the monocytes in the poor responders clustered independently from those in healthy donors (Figure 5B, 5C). GSVA analysis revealed aberrant activation of multiple pathways in P03-AT, including the TNF-α signaling, MAPK, and IL12-STAT4 pathways (Supplementary Figure 3A). Moreover, monocytes from P04-AT showed significantly higher expression levels of several genes, including AP-1 members *FOS* and *JUNB* (Supplementary Figure 3B).

We also observed that CD16+ monocytes were significantly enriched with cells originating from P04-BC-BT. They accounted for 49.7% (429/863) of the total number of monocytes within the sample, which was much higher than in the healthy control (101/690, 14.6%; P < 2.0e-16, Fisher's exact test). Their expression profile was also distinct from CD14+ monocytes (Supplementary Figure 3C, 3D). Several genes, including *IFI27* and *VAMP5*, were highly up-regulated in P04-BC-BT compared to the CD16+ monocytes from the healthy donor (Figure 5D). The increased frequency of macrophages may indicate that there is a pro-inflammatory stimulation in the peripheral blood during the blast crisis phase. In addition, further dissection of CD16+ monocytes revealed a higher proportion of cells expressing *C1QA*, which has been implicated in suppression of the cytotoxicity of CD8+ T cells (Figure 5E) [32, 33]. The Clu-MME in P04-BC-BT also showed evidence for significantly elevated interaction between *CD52* and *SIGLEC10*, which has been shown to suppress T cell function (Figure 5F) [34]. *LAG3* was also increased in P04-BC-BT T cells (Supplementary Figure 3E), revealing suppressed T cell function through the presence of both T cell inhibitory signals and suppressive macrophages during blast crisis.

In addition to the possible interaction between macrophages and T cells, we noticed that the NK cell inhibitory receptor and licensing mediator *KIR2DL2* was only restored in the patients with favorable prognosis after treatment (Figure 5G). Previous studies have suggested that *KIR2DL2* plays a protective role in CML as well as in solid tumor types, representing a good prognostic factor [35, 36]. Expression of NK cell inhibitory receptors are highly heterogeneous among human populations; thus, it was unclear whether this discrepancy in *KIR2DL2* was due to differences in tumorigenesis or in genetic backgrounds among individuals. However, we did not find discrepancies for other inhibitory receptors, including *KIR2DL3*, which is considered to bind *HLA-C1* like *KIR2DL2*, but with lower affinity (Supplementary Figure 3F) [37]. Previous studies have also suggested that KIR polymorphism was associated with clinical response [38]. Although this result requires further validation due to the low number of cells detected with expression of *KIRD2L2*, it might indicate that this gene is crucial in sensitizing NK cells against leukemic cells.

## DISCUSSION

Ultra-high-throughput single cell RNA-seq enabled characterization of the expression profiles of leukemic cells and immune structures simultaneously. Although the number of cases was limited, we managed to identify features of leukemic cells common among patients, and to excavate significant discriminations between poor and favorable prognosis, based principally on the cell population configurations. The presence of even small cell populations with unique expression profiles could be detected. Cells with similar transcriptomes existing in all patients represented "classic leukemic" cells in CML. Thus, molecular analysis at the single cell level has provided a comprehensive overview of many cell populations in peripheral blood of patients, which has implications for management of patients and their response to the current standard of care for CML.

Many of the previously reported genes playing important roles in CML progression or resistance were detected in our study, such as *c-FOS* and *DUSP1* in CD34+ cells [39]. However, preservation of these genes might vary in different subpopulations of terminally differentiated cells, as *c-FOS* was enriched in a subgroup of CD14+ monocytes (Supplementary Figure 3B) and *DUSP1* was elevated in CD16+ macrophages (Supplementary Figure 3G), which were significantly expanded in the BC phase. This observation emphasizes the necessity of understanding the precise role of different lineage priming processes and the exact function of these terminal immune compartments in the context of leukemia.

The presence of sample-specific subpopulations in untreated cells from patients with a poor prognosis indicated that the clinical resistance phenotype is more intrinsic than acquired. Nevertheless, the exact cell subpopulations indicative of clinical TKI resistance varied among patients, and they represented/covered different developmental stages along the trajectory of hematopoiesis. Using a recently revised hematopoiesis map at the resolution level of single cells, we refined the (accurate) hierarchy of the circulating progenitor cells previously determined as common myeloid progenitors (CMP) and megakaryocyte/erythrocyte progenitors (MEP) [40, 41]. The differentiation destiny of the leukemic HSCs in CML was basically confined to granulocytes and erythrocytes, but it may vary among patients. Notably, in the non-responder for imatinib, an

accumulation of cells in the population occurred at a higher level in the hierarchy of the differentiation of stem cells, suggesting that restoring the ability of cells to differentiate may further facilitate TKI treatment.

The mutations accompanying the development of cancer continue to be characterized for the purposes of understanding the biology of the disease and the design of therapies specific for malignant cells in individual patients [42–44]. While targeting mutations works for many cancers [45–47], the approach is of significant benefit to CML patients due to the application of TKI targeted at BCR-ABL kinase [48]. However, a significant proportion of CML patients still suffer from the resistance to TKI treatment. Thus, although the prediction of primary resistance of CML patients to imatinib is difficult, it is a clinically important goal to achieve [28].

The uniquely expressed gene sets in resistant-relevant cell populations discriminate them from the normal functional compartments, as well as classic leukemic cells, providing possible surrogates for their identification in clinical application. In our study, we identified two unique cell populations as primary imatinib resistance clusters, which can predict imatinib response before treatment using available CML expression profiles. Of more clinical importance is that these two clusters can be detected in peripheral blood,

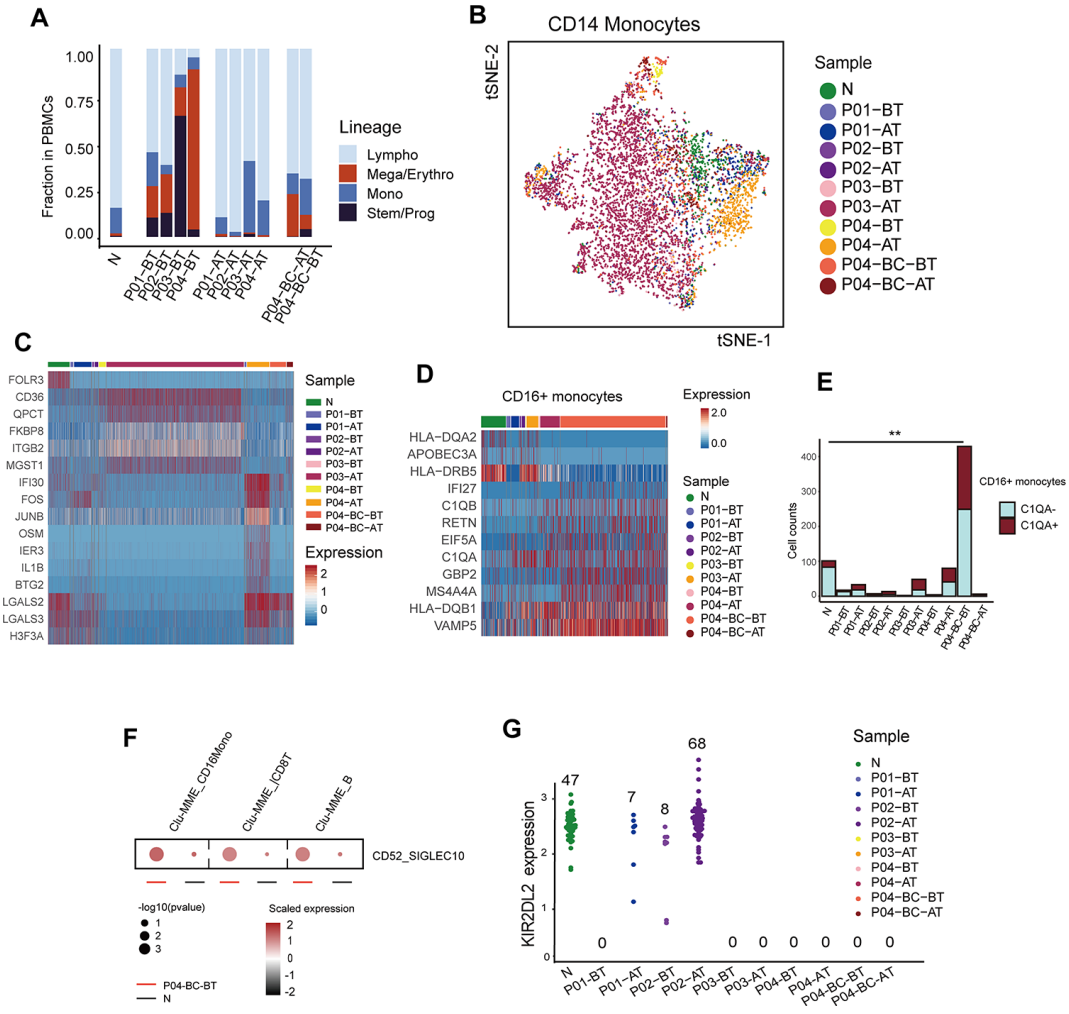



**Figure 5.** Modulation of the immune structure in response to imatinib treatment (**A**) Bar plots showing the fraction of different cell lineages in each sample. (**B**) TSNE plot of the re-clustering result of CD14+ monocytes. Cells are highlighted in colors indicating sample of origin. The number of cells in each cluster is listed in Supplementary Table 2. (**C**) Heatmap displaying the expression profiles of CD14+ monocytes across different samples. Top differentially expressed genes in each sample are indicated. (**D**) Heatmap comparing the expression profiles of CD16+ monocytes across different samples. The selected marker genes are indicated. (**E**) Bar plots comparing the detected *C1QA*-expressing monocytes across different samples. P value was calculated using a Fisher's exact test. (**F**) Dot plots comparing the CD52-SIGLECT10 ligand-receptor interactions between P04-BC-BT and N. The p value was calculated using a permutation test. (**G**) Beeswarm plots showing the expression of *KIR2DL2* among all cells across samples. The number of cells expressing *KIR2DL2* in each sample is indicated.

but not necessarily in bone marrow. Nevertheless, some of the possible prognostic populations were too small, such as the Clu-CD34 subset seeding blast crisis and Clu-MME, to be detected using traditional strategies. These results further highlight the importance and necessity of high-throughput methodologies, such as single cell RNA sequencing.

In addition, molecular distinctions between favorable and poor prognosis was found not only in leukemic stem/progenitor clones, but also in the terminal immune context which consist mainly of macrophage/monocytes and T cells. The observed molecular differences between leukemic monocytes emphasizes the fact that although the morphology might be normal, expression can differ in leukemic myeloid cells. Both macrophages and T cells are known to play vital roles in the suppression of malignancy, which underscores the value of the exploration of immunotherapy in combination with TKI treatment, especially for patients at high risk for resistance or relapse [49].

## MATERIALS AND METHODS

### Ethics statement

The protocols in this study were approved by the Institutional Review Board (IRB) of Beijing Hospital following the guidelines issued by the Ministry of Science and Technology of the People's Republic of China. Written informed consent was obtained from all individuals participating in the study in accordance with the Declaration of Helsinki. Human samples were collected and anonymously coded.

### Patients and samples

Our study included four newly diagnosed CML patients who had undergone treatment with imatinib between 2017 and 2018 at the Department of Hematology, Beijing Hospital and the Department of Hematology, Beijing Tongren Hospital. One healthy donor was also included in our study. All patients were diagnosed according to the 2008 World Health Organization (WHO) consensus criteria. Peripheral blood mononuclear cells (PBMCs) for single cell RNA sequencing were isolated from 10 mL of peripheral blood within 2 h of the draw using Human Lymphocyte Separation Medium (DAKEWE; Shenzhen, China) according to the manufacturer's instructions. The clinical information is summarized in Supplementary Table 1**.**

### 10X Genomics Single-Cell RNA Sequencing

PBMCs were suspended and loaded on the Chromium Single Cell Controller, and single-cell RNA-seq libraries were prepared following the manufacturer's instructions using the Single Cell 3' Library Gel Bead Kit V1 (10 X Genomics; San Francisco, CA, USA). The captured libraries were sequenced on the Illumina HiSeq genome analyzer with paired-end 150-base reads (Illumina; San Diego, CA, USA).

### Data Analysis

#### *Pre-processing of sequencing data*

The sequencing raw data was first de-multiplexed, aligned to the reference genome (hg19, UCSC), and quality-filtered, and barcodes and unique molecular identifiers (UMIs) were counted using CellRanger for each PBMC sample, according to the manufacturer's instructions. We applied Seurat v2.3.4 [50] and Seurat v3.5.2 [51] for merging of samples, filtering (cells with > 200 genes detected, genes with > 3 cells detected, and cells with > 5% UMIs derived from mitochondrial genes), down-sampling, data normalization and scaling.

#### *Integration of datasets*

Integration of datasets from different studies was generated with procedures implemented in Seurat [51]. In brief, features with high variances in each dataset were first selected individually, and those features identified as highly-variable in multiple datasets were combined. The top ranked 1,000 to 2,000 features were selected for downstream processing. The paired datasets were then placed in sharing low-dimension spaces using canonical correlation analysis (CCA). Mutual nearest neighbors (MNN) were then identified based on K-nearest neighbors (KNN), and these MNNs were then defined as "anchors" connecting each pair of datasets. The parameter *K.filters* in the command *FindIntegratedAnchors* was set to 50 when dealing with small sized datasets with < 500 cells. Datasets were then assembled using the identified anchors and processed as a single scRNA-seq object for subsequent analysis.

#### *Dimension reduction and unsupervised clustering of the data*

Principle component analysis (PCA) was performed using variable genes with *RunPCA* implemented in Seurat. The standard deviations of the first 40 principle components (PCs) were plotted to determine which PC would be used for further clustering and dimension reduction. Unsupervised clustering was performed using *FindClusters* in Seurat. Clustering results were visualized in 2-dimensional images by applying t-distributed Stochastic Neighbor Embedding (t-SNE) or Uniform Manifold Approximation and Projection (UMAP) [52]. Cluster-specific markers and differentially expressed genes among clusters or between any two given cell groups were identified using *FindAllMarkers* or *FindMarkers* in Seurat, and cell types were determined

using the known lineage-specific markers listed in Supplementary Figure 1A. The scores of different features were calculated based on the average expression levels of the feature-related gene sets (Supplementary Table 3) using *AddModuleScore* in Seurat. Cell cycle status was imputed with *CellCycleScoring* in Seurat, using implemented *cc.genes* for scoring and phase mapping (Supplementary Table 3).

***Projection of single cells onto reference maps***

When projecting the query dataset onto the reference dataset, we first selected 2,000 of the most variable genes among the intersecting genes between the reference and query datasets. Using these 2000 features, we calculated the PCs in the reference dataset, and the top 30 PCs were projected onto the query dataset. For each cell from the query dataset in the common lower dimensional space, the nearest five cells from the reference dataset were considered as the anchors of the query cell. The projection result is represented by a kernel density plot of the anchors on the reference map.

***Pathway and gene set analysis***

Standard GSEA [53] or GSVA [54] was applied for the identification of significantly enriched gene sets between two clusters or given cell groups. The required input files (including *.gct and *.cls) were extracted from the expression matrix (the integrated transformed matrix was used for integrated datasets), and the gene set files were downloaded from The Molecular Signature Database (MSigDB) [55]. Hallmark, curated and GO gene sets (H, C2 and C5, respectively) from MSigDB were used for analysis.

***Cell interaction evaluation***

The possible interaction between cell populations was evaluated using CellphoneDB [56], based on the curated known ligand-receptor pairs implemented in the package. In brief, for every two given clusters, the genes encoding certain receptors or ligands which were expressed in more than 30% of cells in a specific cluster were chosen for downstream analysis, and the significance of a ligand-receptor pair between two clusters was calculated through the permutation test by randomly assigning the cluster labels of each cell 1,000 times. An empirical P value was determined by the rank of the actual average expression of a given ligand and receptor pair in two clusters among the 1,000 permutated results.

***Pseudo-time trajectory construction***

Pseudo-time trajectory was constructed using Monocle v2.6.4 R package [57, 58]. The top 50 differentially expressed genes in each cluster identified using *FindAllMarkers* in Seurat were combined and used for dimension reduction (DDRTRee) and cell ordering.

***Statistical analysis and data availability***

Statistical analyses were performed in R 3.4.3 (foundation for statistical computing, or functions implemented in Seurat). Detailed descriptions are specified in the text.

**Validation in independent datasets of 276 CML patients**

Gene expression profiles of CML patients (n = 59) who had received imatinib treatment were obtained from the GEO database (GSE14671) [28]. Forty-one were imatinib nonresponders (NR) while 18 were responders (R). Samples of each CML were collected before the initiation of imatinib therapy. Gene expression profiles were obtained with the Affymetrix Human Genome U133 Plus 2.0 Array. The robust multiarray averaging method (RMA) was used to analyze the RNA expression microarray of each CML sample. The expression levels of each gene were log2 transformed. The correlation coefficients between the gene expression profiles of the 59 CML patients and the Clu-CD34 (or Clu-MPO) gene expression signatures (Supplementary Table 4) were calculated using the Pearson correlation. The correlation coefficients of the 41 imatinib nonresponders and the 18 responders were further used to build ROC curves with the pROC package (R 3.4.3).

Similarly, 3 other independent datasets with a total of 217 CML patients were analyzed to achieve significant scientific results. The clinical characteristics of the cohorts are the following: 1) imatinib therapy (GSE130404, 96 CML patients) [29]: 13 imatinib nonresponders and 83 responders; 2) dasatinib therapy (GSE33224, 14 CML patients) [30]: 8 dasatinib nonresponders and 6 responders; and 3) different phases of CML (GSE4170, 107 CML patients) [31]: 17 accelerated phase (AP), 33 (blast crisis (BC) and 57 chronic phase (CP) CML patients. The first two cohorts were used to build receiver operating curves (ROC).

## Abbreviations

CML: chronic myeloid leukemia; TKI: tyrosine kinase inhibitor; PBMC: peripheral blood mononuclear cell; CCA: canonical correlation analysis; PCA: principle component analysis; KNN: K-nearest neighbors; MNN: mutual nearest neighbors; tSNE: t-distributed stochastic neighbor embedding; UMAP: uniform manifold approximation and projection; GSEA: gene set enrichment analysis; GSVA: gene set variation analysis; UMI: unique molecular index; HSC: hematopoietic stem cells; CMP: common myeloid progenitors; GMP: granuloid progenitors; MEP: megakaryocyte/erythrocyte progenitors; BM: bone marrow; BC: blast crisis; BT: before treatment; AT: after treatment.

## AUTHOR CONTRIBUTIONS

Yuchen Jiao, Hui Liu, Yi-Xin Zeng and Weilong Zhang planned and designed the research; Weilong Zhang, Beibei Yang performed the benchwork for sequencing; Weilong Zhang and Xixi Chen performed the data processing and analysis; Linqian Weng, Jiangtao Li, Jiefei Bai, Ting Wang, Jingwen Wang, and Jin Ye collected blood samples and prepared clinical information; Weilong Zhang and Xixi Chen wrote the manuscript and Yuchen Jiao, Hongmei Jing, Hui Liu, Yi-Xin Zeng revised the manuscript.

## CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

## FUNDING

This work was supported by the National Key R&D Program of China (2018YFC1312100 to Yuchen Jiao), the National Key Basic Research Program of China (973 program 2015CB553902 to Yuchen Jiao), the Beijing Municipal Science and Technology commission (grants Z171100001017200 and Z181100001718162 to Hui Liu), the National Natural Science Foundation Fund (81472559 to Yuchen Jiao), Peking Union Medical College Fundamental Research Funds (JK2013A21, JK2014B10 to Yuchen Jiao), and the CAMS Innovation Fund for Medical Sciences (CIFMS; 2016-I2M-1-001 to Yuchen Jiao).

## REFERENCES

1. Pfirrmann M, Baccarani M, Saussele S, Guilhot J, Cervantes F, Ossenkoppele G, Hoffmann VS, Castagnetti F, Hasford J, Hehlmann R, Simonsson B. Prognosis of long-term survival considering disease-specific death in patients with chronic myeloid leukemia. Leukemia. 2016; 30:48–56. https://doi.org/10.1038/leu.2015.261 PMID:26416462

2. Chen Z, Shao C, Wang W, Zuo Z, Mou X, Hu SJ, DiGiuseppe JA, Zu Y, Medeiros LJ, Hu S. Cytogenetic landscape and impact in blast phase of chronic myeloid leukemia in the era of tyrosine kinase inhibitor therapy. Leukemia. 2017; 31:585–92. https://doi.org/10.1038/leu.2016.231 PMID:27560111

3. Marin D, Kaeda J, Szydlo R, Saunders S, Fleming A, Howard J, Andreasson C, Bua M, Olavarria E, Rahemtulla A, Dazzi F, Kanfer E, Goldman JM, Apperley JF. Monitoring patients in complete cytogenetic remission after treatment of CML in chronic phase with imatinib: patterns of residual leukaemia and prognostic factors for cytogenetic relapse. Leukemia. 2005; 19:507–12. https://doi.org/10.1038/sj.leu.2403664 PMID:15703781

4. Gorre ME, Mohammed M, Ellwood K, Hsu N, Paquette R, Rao PN, Sawyers CL. Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. Science. 2001; 293:876–80. https://doi.org/10.1126/science.1062538 PMID:11423618

5. Johansson B, Fioretos T, Mitelman F. Cytogenetic and molecular genetic evolution of chronic myeloid leukemia. Acta Haematol. 2002; 107:76–94. https://doi.org/10.1159/000046636 PMID:11919388

6. Rossignol A, Levescot A, Jacomet F, Robin A, Basbous S, Giraud C, Roy L, Guilhot F, Turhan AG, Barra A, Herbelin A, Gombert JM. Evidence for BCR-ABL-dependent dysfunctions of iNKT cells from chronic myeloid leukemia patients. Eur J Immunol. 2012; 42:1870–75. https://doi.org/10.1002/eji.201142043 PMID:22585600

7. Chen CI, Koschmieder S, Kerstiens L, Schemionek M, Altvater B, Pscherer S, Gerss J, Maecker HT, Berdel WE, Juergens H, Lee PP, Rossig C. NK cells are dysfunctional in human chronic myelogenous leukemia before and on imatinib treatment and in BCR-ABL-positive mice. Leukemia. 2012; 26:465–74. https://doi.org/10.1038/leu.2011.239 PMID:21904381

8. Christiansson L, Söderlund S, Svensson E, Mustjoki S, Bengtsson M, Simonsson B, Olsson-Strömberg U, Loskog AS. Increased level of myeloid-derived suppressor cells, programmed death receptor ligand 1/programmed death receptor 1, and soluble CD25 in sokal high risk chronic myeloid leukemia. PLoS One. 2013; 8:e55818. https://doi.org/10.1371/journal.pone.0055818 PMID:23383287

9. Seggewiss R, Price DA, Purbhoo MA. Immunomodulatory effects of imatinib and second-generation tyrosine kinase inhibitors on T cells and dendritic cells: an update. Cytotherapy. 2008; 10:633–41. https://doi.org/10.1080/14653240802317639 PMID:18836918

10. Giladi A, Paul F, Herzog Y, Lubling Y, Weiner A, Yofe I, Jaitin D, Cabezas-Wallscheid N, Dress R, Ginhoux F, Trumpp A, Tanay A, Amit I. Single-cell characterization of haematopoietic progenitors and their trajectories in homeostasis and perturbed haematopoiesis. Nat Cell Biol. 2018; 20:836–46. https://doi.org/10.1038/s41556-018-0121-4 PMID:29915358

11. Paul F, Arkin Y, Giladi A, Jaitin DA, Kenigsberg E, Keren-Shaul H, Winter D, Lara-Astiaso D, Gury M, Weiner A, David E, Cohen N, Lauridsen FK, et al. Transcriptional heterogeneity and lineage commitment in myeloid progenitors. Cell. 2015; 163:1663–77. https://doi.org/10.1016/j.cell.2015.11.013 PMID:26627738

12. Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, Hirche C, Lutz C, Buss EC, Nowak D, Boch T, Hofmann WK, Ho AD, et al. Human haematopoietic stem cell lineage commitment is a continuous process. Nat Cell Biol. 2017; 19:271–81. https://doi.org/10.1038/ncb3493 PMID:28319093

13. Giustacchini A, Thongjuea S, Barkas N, Woll PS, Povinelli BJ, Booth CA, Sopp P, Norfo R, Rodriguez-Meira A, Ashley N, Jamieson L, Vyas P, Anderson K, et al. Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. Nat Med. 2017; 23:692–702. https://doi.org/10.1038/nm.4336 PMID:28504724

14. Warfvinge R, Geironson L, Sommarin MN, Lang S, Karlsson C, Roschupkina T, Stenke L, Stentoft J, Olsson-Strömberg U, Hjorth-Hansen H, Mustjoki S, Soneji S, Richter J, Karlsson G. Single-cell molecular analysis defines therapy response and immunophenotype of stem cell subpopulations in CML. Blood. 2017; 129:2384–94. https://doi.org/10.1182/blood-2016-07-728873 PMID:28122740

15. Pellin D, Loperfido M, Baricordi C, Wolock SL, Montepeloso A, Weinberg OK, Biffi A, Klein AM, Biasco L. A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. Nat Commun. 2019; 10:2395. https://doi.org/10.1038/s41467-019-10291-0 PMID:31160568

16. Schenkel AR, Mamdouh Z, Chen X, Liebman RM, Muller WA. CD99 plays a major role in the migration of monocytes through endothelial junctions. Nat Immunol. 2002; 3:143–50. https://doi.org/10.1038/ni749 PMID:11812991

17. Chung SS, Eng WS, Hu W, Khalaj M, Garrett-Bakelman FE, Tavakkoli M, Levine RL, Carroll M, Klimek VM, Melnick AM, Park CY. CD99 is a therapeutic target on disease stem cells in myeloid Malignancies. Sci Transl Med. 2017; 9:eaaj2025. https://doi.org/10.1126/scitranslmed.aaj2025 PMID:28123069

18. Imbert AM, Belaaloui G, Bardin F, Tonnelle C, Lopez M, Chabannon C. CD99 expressed on human mobilized peripheral blood CD34+ cells is involved in transendothelial migration. Blood. 2006; 108:2578–86. https://doi.org/10.1182/blood-2005-12-010827 PMID:16825498

19. Pasello M, Manara MC, Scotlandi K. CD99 at the crossroads of physiology and pathology. J Cell Commun Signal. 2018; 12:55–68. https://doi.org/10.1007/s12079-017-0445-z PMID:29305692

20. Collins CT, Hess JL. Role of HOXA9 in leukemia: dysregulation, cofactors and essential targets. Oncogene. 2016; 35:1090–98. https://doi.org/10.1038/onc.2015.174 PMID:26028034

21. Sun Y, Zhou B, Mao F, Xu J, Miao H, Zou Z, Phuc Khoa LT, Jang Y, Cai S, Witkin M, Koche R, Ge K, Dressler GR, et al. HOXA9 reprograms the enhancer landscape to promote leukemogenesis. Cancer Cell. 2018; 34:643–58.e5. https://doi.org/10.1016/j.ccell.2018.08.018 PMID:30270123

22. Dohse M, Scharenberg C, Shukla S, Robey RW, Volkmann T, Deeken JF, Brendel C, Ambudkar SV, Neubauer A, Bates SE. Comparison of ATP-binding cassette transporter interactions with the tyrosine kinase inhibitors imatinib, nilotinib, and dasatinib. Drug Metab Dispos. 2010; 38:1371–80. https://doi.org/10.1124/dmd.109.031302 PMID:20423956

23. Tusi BK, Wolock SL, Weinreb C, Hwang Y, Hidalgo D, Zilionis R, Waisman A, Huh JR, Klein AM, Socolovsky M. Population snapshots predict early haematopoietic and erythroid hierarchies. Nature. 2018; 555:54–60. https://doi.org/10.1038/nature25741 PMID:29466336

24. Pop R, Shearstone JR, Shen Q, Liu Y, Hallstrom K, Koulnis M, Gribnau J, Socolovsky M. A key commitment step in erythropoiesis is synchronized with the cell cycle clock through mutual inhibition between PU.1 and s-phase progression. PLoS Biol. 2010; 8:e1000484. https://doi.org/10.1371/journal.pbio.1000484 PMID:20877475

25. Vitali C, Bassani C, Chiodoni C, Fellini E, Guarnotta C, Miotti S, Sangaletti S, Fuligni F, De Cecco L, Piccaluga PP, Colombo MP, Tripodo C. SOCS2 controls proliferation and stemness of hematopoietic cells under stress conditions and its deregulation marks unfavorable acute leukemias. Cancer Res. 2015; 75:2387–99. https://doi.org/10.1158/0008-5472.CAN-14-3625 PMID:25858143

26. Marenholz I, Heizmann CW. S100A16, a ubiquitously expressed EF-hand protein which is up-regulated in tumors. Biochem Biophys Res Commun. 2004; 313:237–44.

https://doi.org/10.1016/j.bbrc.2003.11.115
PMID:14684152

27. Brenner AK, Bruserud Ø. S100 proteins in acute myeloid leukemia. Neoplasia. 2018; 20:1175–86. https://doi.org/10.1016/j.neo.2018.09.007 PMID:30366122

28. McWeeney SK, Pemberton LC, Loriaux MM, Vartanian K, Willis SG, Yochum G, Wilmot B, Turpaz Y, Pillai R, Druker BJ, Snead JL, MacPartlin M, O'Brien SG, et al. A gene expression signature of CD34+ cells to predict major cytogenetic response in chronic-phase chronic myeloid leukemia patients treated with imatinib. Blood. 2010; 115:315–25. https://doi.org/10.1182/blood-2009-03-210732 PMID:19837975

29. Kok CH, Yeung DT, Lu L, Watkins DB, Leclercq TM, Dang P, Saunders VA, Reynolds J, White DL, Hughes TP. Gene expression signature that predicts early molecular response failure in chronic-phase CML patients on frontline imatinib. Blood Adv. 2019; 3:1610–21. https://doi.org/10.1182/bloodadvances.2019000195 PMID:31126916

30. Silveira RA, Fachel AA, Moreira YB, De Souza CA, Costa FF, Verjovski-Almeida S, Pagnano KB. Protein-coding genes and long noncoding RNAs are differentially expressed in dasatinib-treated chronic myeloid leukemia patients with resistance to imatinib. Hematology. 2014; 19:31–41. https://doi.org/10.1179/1607845413Y.0000000094 PMID:23676950

31. Radich JP, Dai H, Mao M, Oehler V, Schelter J, Druker B, Sawyers C, Shah N, Stock W, Willman CL, Friend S, Linsley PS. Gene expression changes associated with progression and response in chronic myeloid leukemia. Proc Natl Acad Sci USA. 2006; 103:2794–99. https://doi.org/10.1073/pnas.0510423103 PMID:16477019

32. Clarke EV, Tenner AJ. Complement modulation of T cell immune responses during homeostasis and disease. J Leukoc Biol. 2014; 96:745–56. https://doi.org/10.1189/jlb.3MR0214-109R PMID:25210145

33. Ling GS, Crawford G, Buang N, Bartok I, Tian K, Thielens NM, Bally I, Harker JA, Ashton-Rickardt PG, Rutschmann S, Strid J, Botto M. C1q restrains autoimmunity and viral infection by regulating $CD8^+$ T cell metabolism. Science. 2018; 360:558–63. https://doi.org/10.1126/science.aao4555 PMID:29724957

34. Bandala-Sanchez E, Bediaga NG, Goddard-Borger ED, Ngui K, Naselli G, Stone NL, Neale AM, Pearce LA, Wardak A, Czabotar P, Haselhorst T, Maggioni A, Hartley-Tassell LA, et al. CD52 glycan binds the proinflammatory B box of HMGB1 to engage the siglec-10 receptor and suppress human T cell function. Proc Natl Acad Sci USA. 2018; 115:7783–88. https://doi.org/10.1073/pnas.1722056115 PMID:29997173

35. Kärre K, Ljunggren HG, Piontek G, Kiessling R. Selective rejection of H-2-deficient lymphoma variants suggests alternative immune defence strategy. Nature. 1986; 319:675–78. https://doi.org/10.1038/319675a0 PMID:3951539

36. Middleton D, Diler AS, Meenagh A, Sleator C, Gourraud PA. Killer immunoglobulin-like receptors (KIR2DL2 and/or KIR2DS2) in presence of their ligand (HLA-C1 group) protect against chronic myeloid leukaemia. Tissue Antigens. 2009; 73:553–60. https://doi.org/10.1111/j.1399-0039.2009.01235.x PMID:19493232

37. Moesta AK, Norman PJ, Yawata M, Yawata N, Gleimer M, Parham P. Synergistic polymorphism at two positions distal to the ligand-binding site makes KIR2DL2 a stronger receptor for HLA-C than KIR2DL3. J Immunol. 2008; 180:3969–79. https://doi.org/10.4049/jimmunol.180.6.3969 PMID:18322206

38. Ureshino H, Shindo T, Tanaka H, Kimura S. [Chronic myeloid leukemia and NK cell immunity]. Rinsho Ketsueki. 2017; 58:381–88. https://doi.org/10.11406/rinketsu.58.381 PMID:28484170

39. Kesarwani M, Kincaid Z, Gomaa A, Huber E, Rohrabaugh S, Siddiqui Z, Bouso MF, Latif T, Xu M, Komurov K, Mulloy JC, Cancelas JA, Grimes HL, Azam M. Targeting c-FOS and DUSP1 abrogates intrinsic resistance to tyrosine-kinase inhibitor therapy in BCR-ABL-induced leukemia. Nat Med. 2017; 23:472–82. https://doi.org/10.1038/nm.4310 PMID:28319094

40. Diaz-Blanco E, Bruns I, Neumann F, Fischer JC, Graef T, Rosskopf M, Brors B, Pechtel S, Bork S, Koch A, Baer A, Rohr UP, Kobbe G, et al. Molecular signature of CD34(+) hematopoietic stem and progenitor cells of patients with CML in chronic phase. Leukemia. 2007; 21:494–504. https://doi.org/10.1038/sj.leu.2404549 PMID:17252012

41. Bruns I, Czibere A, Fischer JC, Roels F, Cadeddu RP, Buest S, Bruennert D, Huenerlituerkoglu AN, Stoecklein NH, Singh R, Zerbini LF, Jäger M, Kobbe G, et al. The hematopoietic stem cell in chronic phase CML is characterized by a transcriptional profile resembling normal myeloid progenitor cells and reflecting loss of quiescence. Leukemia. 2009; 23:892–99. https://doi.org/10.1038/leu.2008.392 PMID:19158832

42. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, Boot A, Covington KR, Gordenin DA, Bergstrom EN, Islam SM, Lopez-Bigas N, Klimczak LJ, et al, and PCAWG Mutational Signatures Working Group, and PCAWG Consortium. The repertoire of mutational signatures in human cancer. Nature. 2020; 578:94–101. https://doi.org/10.1038/s41586-020-1943-3 PMID:32025018

43. Koag MC, Kou Y, Ouzon-Shubeita H, Lee S. Transition-state destabilization reveals how human DNA polymerase β proceeds across the chemically unstable lesion N7-methylguanine. Nucleic Acids Res. 2014; 42:8755–66. https://doi.org/10.1093/nar/gku554 PMID:24966350

44. Kou Y, Koag MC, Lee S. N7 methylation alters hydrogen-bonding patterns of guanine in duplex DNA. J Am Chem Soc. 2015; 137:14067–70. https://doi.org/10.1021/jacs.5b10172 PMID:26517568

45. Wen R, Umeano AC, Kou Y, Xu J, Farooqi AA. Nanoparticle systems for cancer vaccine. Nanomedicine (Lond). 2019; 14:627–48. https://doi.org/10.2217/nnm-2018-0147 PMID:30806568

46. Rawluk J, Waller CF. Gefitinib. Recent Results Cancer Res. 2018; 211:235–46. https://doi.org/10.1007/978-3-319-91442-8_16 PMID:30069771

47. Sunshine J, Taube JM. PD-1/PD-L1 inhibitors. Curr Opin Pharmacol. 2015; 23:32–38. https://doi.org/10.1016/j.coph.2015.05.011 PMID:26047524

48. Jabbour E, Kantarjian H. Chronic myeloid leukemia: 2018 update on diagnosis, therapy and monitoring. Am J Hematol. 2018; 93:442–59. https://doi.org/10.1002/ajh.25011 PMID:29411417

49. Noonan KA, Huff CA, Davis J, Lemas MV, Fiorino S, Bitzan J, Ferguson A, Emerling A, Luznik L, Matsui W, Powell J, Fuchs E, Rosner GL, et al. Adoptive transfer of activated marrow-infiltrating lymphocytes induces measurable antitumor immunity in the bone marrow in multiple myeloma. Sci Transl Med. 2015; 7:288ra78. https://doi.org/10.1126/scitranslmed.aaa7014 PMID:25995224

50. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. Nat Biotechnol. 2015; 33:495–502. https://doi.org/10.1038/nbt.3192 PMID:25867923

51. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, Hao Y, Stoeckius M, Smibert P, Satija R. Comprehensive integration of single-cell data. Cell. 2019; 177:1888–902.e21. https://doi.org/10.1016/j.cell.2019.05.031 PMID:31178118

52. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IW, Ng LG, Ginhoux F, Newell EW. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 2019; 37:38–44 https://doi.org/10.1038/nbt.4314 PMID:30531897

53. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005; 102:15545–50. https://doi.org/10.1073/pnas.0506580102 PMID:16199517

54. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013; 14:7. https://doi.org/10.1186/1471-2105-14-7 PMID:23323831

55. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. Cell Syst. 2015; 1:417–25. https://doi.org/10.1016/j.cels.2015.12.004 PMID:26771021

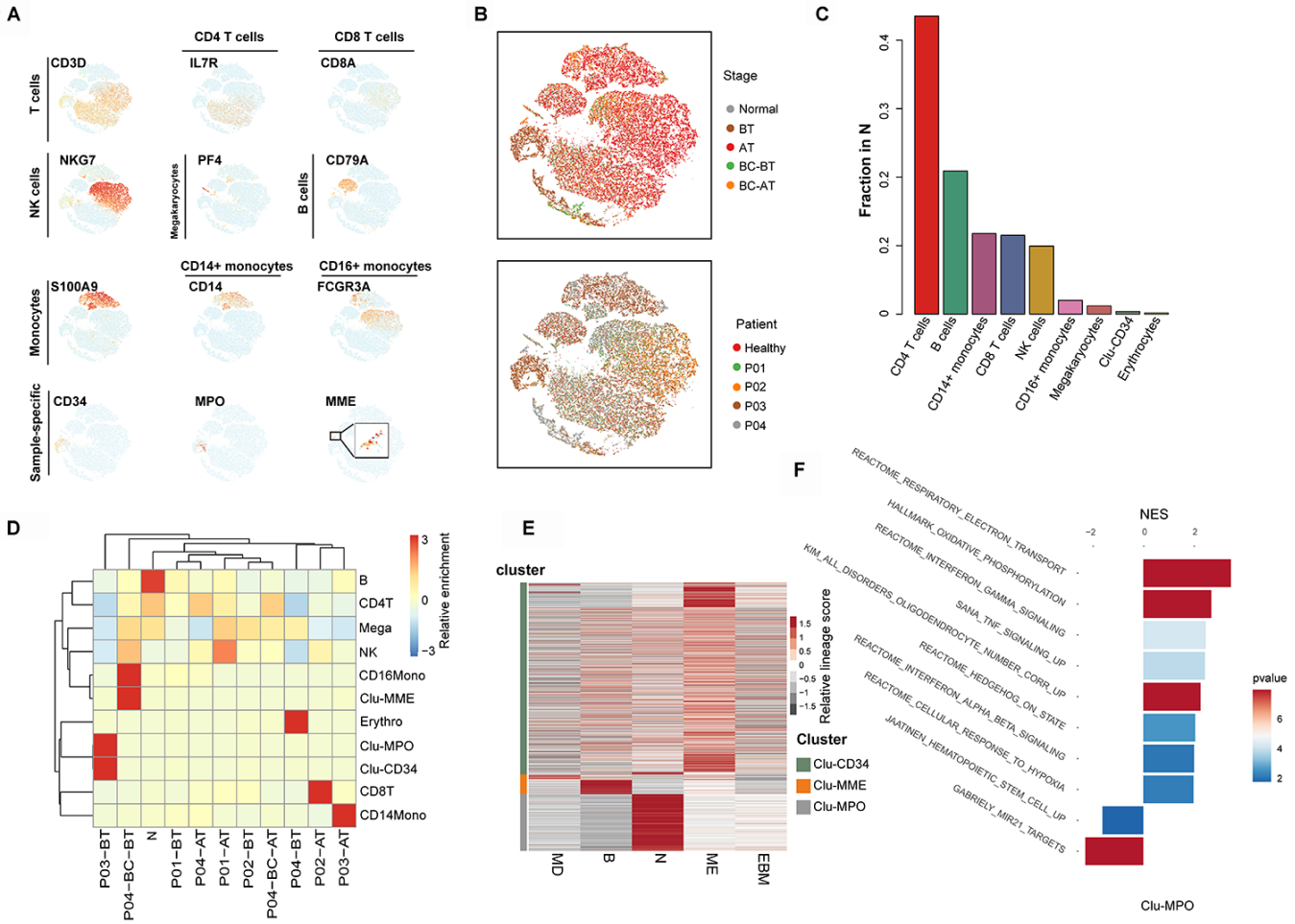
56. Vento-Tormo R, Efremova M, Botting RA, Turco MY, Vento-Tormo M, Meyer KB, Park JE, Stephenson E, Polański K, Goncalves A, Gardner L, Holmqvist S, Henriksson J, et al. Single-cell reconstruction of the early maternal-fetal interface in humans. Nature. 2018; 563:347–53. https://doi.org/10.1038/s41586-018-0698-6 PMID:30429548

57. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, Trapnell C. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods. 2017; 14:979–82. https://doi.org/10.1038/nmeth.4402 PMID:28825705

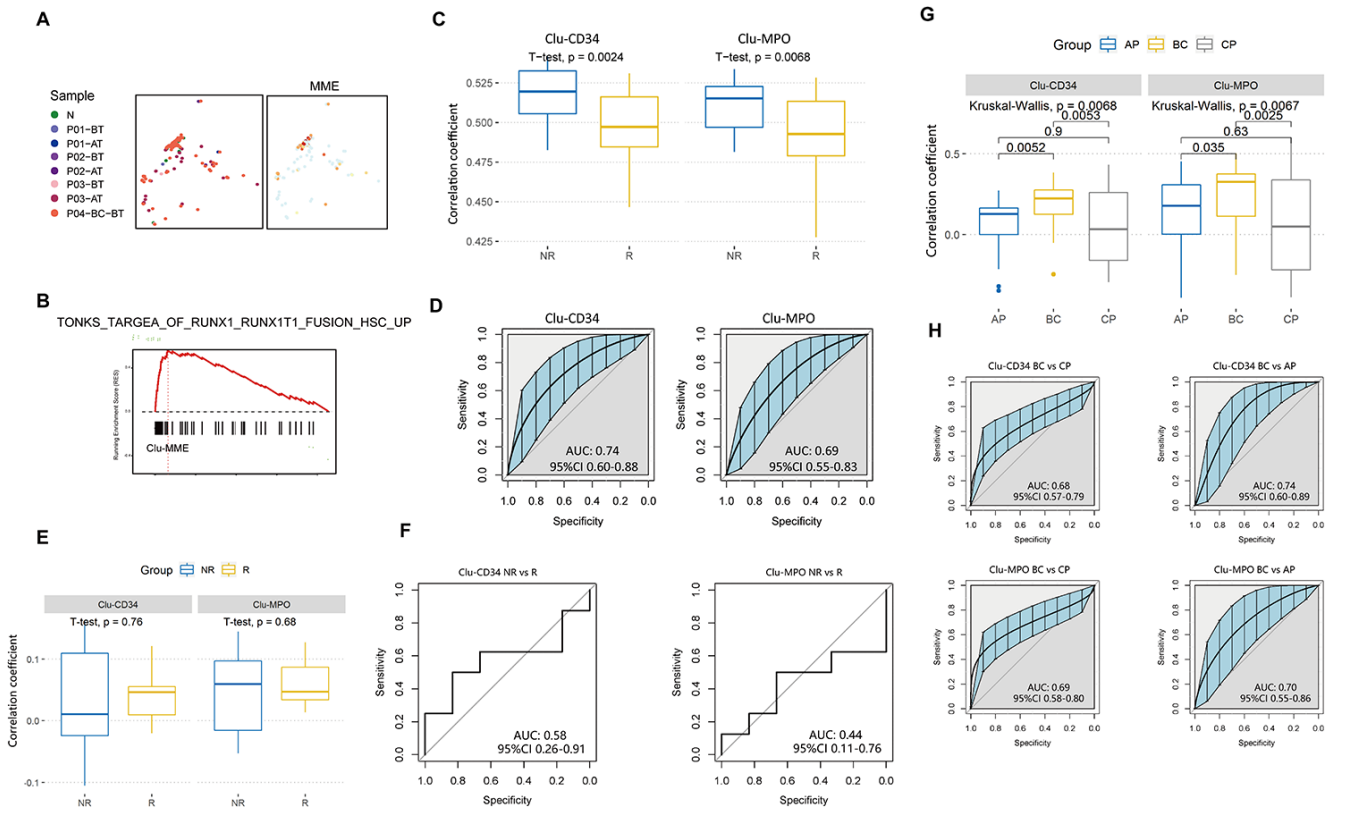
58. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 2014; 32:381–86. https://doi.org/10.1038/nbt.2859 PMID:24658644

# SUPPLEMENTARY MATERIALS

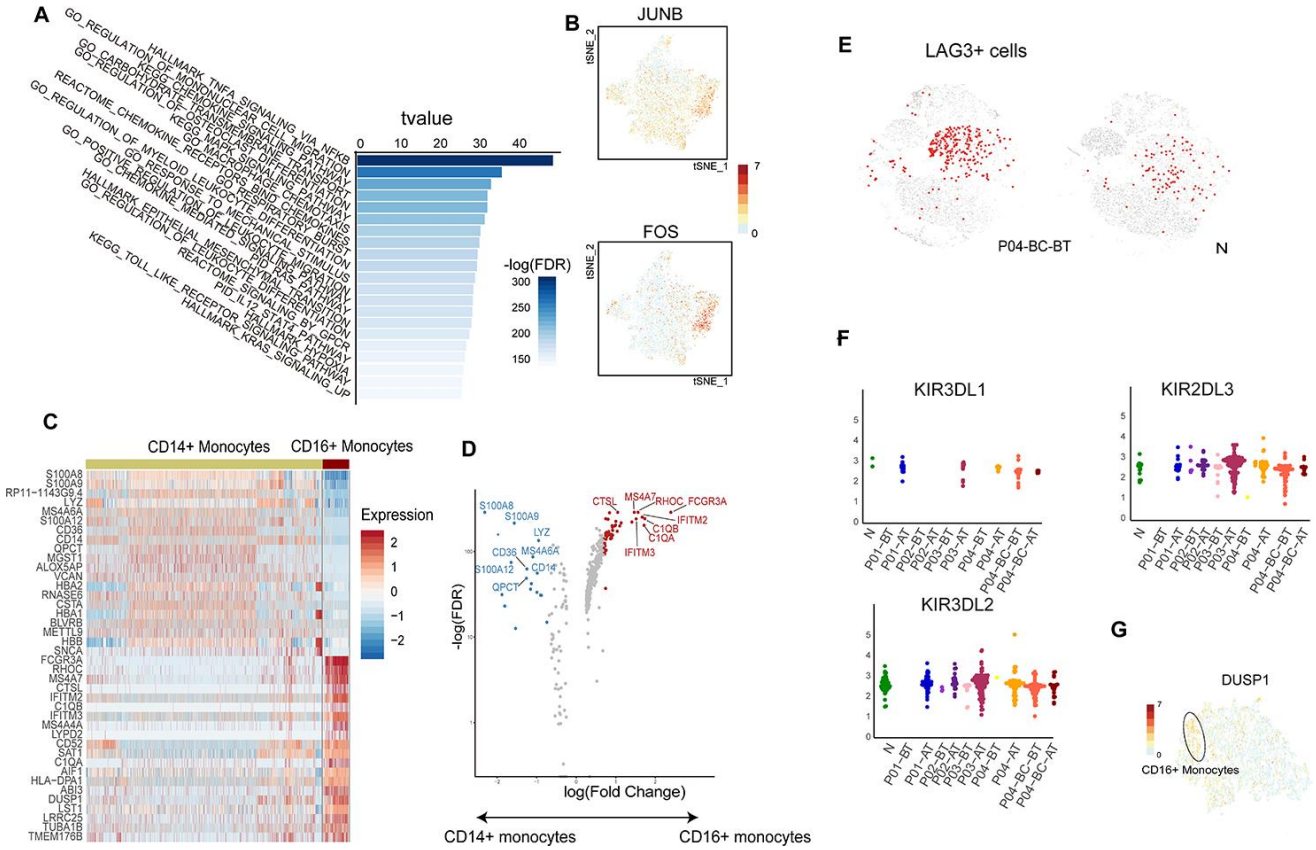
## Supplementary Figures




**Supplementary Figure 1.** (**A**) TSNE plots showing the cell type specific markers in the sequenced dataset. (**B**) TSNE plots of the sequenced dataset colored by sample types (top) and clinical stages (bottom). (**C**) Bar plots showing the fraction of different lineages from the healthy donor sample. The number of cells in each lineage is indicated. (**D**) Heatmap showing enrichment of samples in each cluster. Enrichment scores were calculated using the Fisher's exact test and indicated by log10(Odd Ratio). (**E**) Heatmap showing the lineage potential scores of each stem/progenitor cell towards the different development directions (adopted from Velten et al. [12] as shown in Supplementary Table 3) EBM: esophils/basophils/mast cell; N: neutrophils; ME: megakyrocytes/erythrocytes. B: B cells, MD: multiple direction. (**F**) Bar plots displaying the GSEA result on the ordered expression profile in Clu-MPO. X-axis indicates the normalized enrichment score (NES) and colors indicate the -log10(P value).




**Supplementary Figure 2.** (**A**) TSNE plots showing the sample origins (left) and expression of MME of cells in the Clu-MME cluster. (**B**) Gene set enrichment analysis revealing an enriched signature of RUNX1-RUNX11 fusion in Clu-MME. (**C**) Box plots comparing the correlation coefficients of Clu-CD34 (left) and Clu-MPO (right) between 41 imatinib nonresponders (NR) and 18 responders (R). The correlation coefficients were calculated using Pearson correlation of gene expression signatures of these two clusters with the gene expression profile for each CML patient treated with imatinib (see Methods). (**D**) ROC curves illustrating the classification performance of Clu-CD34 (left) and Clu-MPO (right) gene expression signatures of 41 imatinib nonresponders and 18 responders. The blue shade denotes the 95% confidence interval of the sensitivity at a given specificity point. AUC, area under the ROC curve are indicated. (**E**) Box plots comparing the correlation coefficients of Clu-CD34 (left) and Clu-MPO (right) between 8 dasatinib nonresponders (NR) and 6 responders (R) from GSE33224. The correlation coefficients were calculated using Pearson correlation of gene expression signatures of these two clusters with the gene expression profile in each CML patient treated with imatinib (see Methods). (**F**) ROC curves illustrating the classification performance of Clu-CD34 (left) and Clu-MPO (right) gene expression signatures of 8 dasatinib nonresponders (NR) and 6 responders (R) from GSE33224. The blue shade denotes the 95% confidence interval of the sensitivity at a given specificity point. AUC, area under the ROC curve are indicated. (**G**) Box plots comparing the correlation coefficients of Clu-CD34 (left) and Clu-MPO (right) for 17 AP (accelerated phase) CML patients, 33 BC (blast crisis) CML patients and 57 CP (chronic phase) CML patients from GSE4170. The correlation coefficients were calculated by Pearson correlation of gene expression signature of these two clusters with the gene expression profile in each CML patient treated with imatinib (see Methods). (**H**) ROC curves illustrating the classification performance of Clu-CD34 (up) and Clu-MPO (down) gene expression signatures for 17 AP (accelerated phase), 33 BC (blast crisis) and 57 CP (chronic phase) CML patients from GSE4170. The blue shade denotes the 95% confidence interval of the sensitivity at a given specificity point. AUC, area under the ROC curve are indicated.




**Supplementary Figure 3.** (**A**) Bar plots displaying the GSVA *t* values of top ranked gene sets. Bars are colored by the FDR *Q* value. (**B**) TSNE plots showing expression of JUNB and FOS in CD14+ monocytes. (**C**) Heatmap comparing the expression profiles for all CD14+ and CD16+ monocytes. (**D**) Scatter plot showing the highly-expressed markers in CD14+ monocytes and CD16+ monocytes, separately. Significant markers (FDR<0.05, fold change > 2) are colored as red in CD16+ monocytes and blue in CD14+ monocytes. (**E**) TSNE plots displaying LAG3 expression status of T cells/NK cells in P04-BC-AT in comparison with N. Cells positive for LAG3 were indicated as red. (**F**) Beeswarm plots showing the expression of other KIRs. The number of cells in each sample is indicated. (**G**) TSNE plot showing the enrichment of DUSP1 expression in CD16+ monocytes among all monocytes. Dashed line marks the cluster of CD16+ monocytes.

## Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1, 3, 4.

**Supplementary Table 1. Clinical information.**

**Supplementary Table 2. Number of cells in each cluster across samples**

| | N | P01-AT | P01-BT | P02-AT | P02-BT | P03-AT | P03-BT | P04-AT | P04-BC-AT | P04-BC-BT | P04-BT | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| B | 1045 | 526 | 54 | 197 | 44 | 872 | 99 | 237 | 103 | 377 | 24 | 3578 |
| CD4T | 2176 | 2100 | 171 | 1309 | 354 | 2191 | 120 | 1629 | 759 | 1535 | 4 | 12348 |
| CD8T | 576 | 1027 | 31 | 2674 | 243 | 1694 | 67 | 591 | 48 | 167 | 1 | 7119 |
| CD14Mono | 589 | 481 | 84 | 102 | 59 | 3776 | 189 | 619 | 186 | 434 | 49 | 6568 |
| CD16Mono | 101 | 33 | 17 | 13 | 7 | 48 | 2 | 80 | 6 | 429 | 4 | 740 |
| Erythro | 9 | 3 | 91 | 2 | 256 | 55 | 425 | 30 | 374 | 285 | 698 | 2228 |
| Mega | 61 | 75 | 3 | 17 | 17 | 19 | 5 | 5 | 20 | 58 | 9 | 289 |
| Clu-MME | 8 | 2 | 1 | 3 | 1 | 31 | 3 | 0 | 0 | 135 | 0 | 184 |
| NK | 497 | 1517 | 53 | 864 | 178 | 1236 | 105 | 606 | 248 | 1040 | 9 | 6353 |
| Clu-MPO | 0 | 0 | 5 | 0 | 28 | 29 | 451 | 0 | 0 | 5 | 9 | 527 |
| Clu-CD34 | 20 | 12 | 52 | 5 | 141 | 101 | 1375 | 4 | 5 | 50 | 24 | 1789 |
| **Total** | 5082 | 5776 | 562 | 5186 | 1328 | 10052 | 2841 | 3801 | 1749 | 4515 | 831 | **41723** |

**Supplementary Table 3. Gene modules introduced for cell cycle scoring or lineage priming scoring.**

**Supplementary Table 4. Expression signatures in Clu-CD34 and Clu-MPO.**