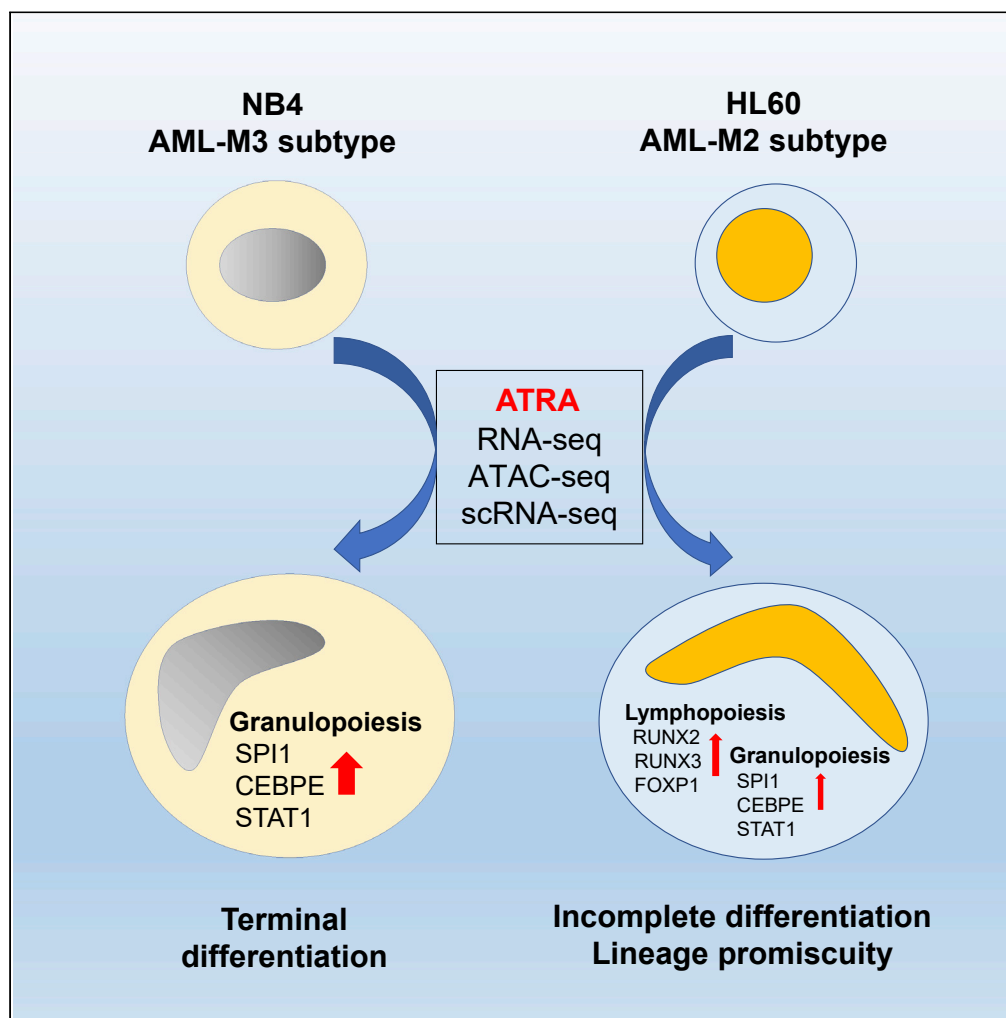


Article

Induced lineage promiscuity undermines the efficiency of all-trans-retinoid-acid-induced differentiation of acute myeloid leukemia



Yijia Tang, Xin Tian, Zihan Xu, ..., Zhu Chen, Saijuan Chen, Feng Liu

sjchen@stn.sh.cn (S.C.)
lf12034@rjh.com.cn (F.L.)

Highlights

Single-cell analysis of ATRA-induced differentiation in two AML cell types

In AML-M3/APL cells, ATRA induces a near normal trajectory of granulopoiesis

In AML-M2 cells, ATRA induces incomplete differentiation and lineage promiscuity

ATRA-induced lineage promiscuity involves transient cis-regulatory reprogramming

Tang et al., iScience 24, 102410
May 21, 2021 © 2021 The Author(s).
<https://doi.org/10.1016/j.isci.2021.102410>

Article

Induced lineage promiscuity undermines the efficiency of all-trans-retinoid-acid-induced differentiation of acute myeloid leukemia

Yijia Tang,^{1,3} Xin Tian,^{1,3} Zihan Xu,¹ Junke Cai,¹ Han Liu,¹ Nan Liu,² Zhu Chen,¹ Saijuan Chen,^{1,*} and Feng Liu^{1,4,*}

SUMMARY

All-trans retinoid acid (ATRA) can induce terminal differentiation of acute promyelocytic leukemia (APL), also known as the M3 subtype of acute myeloid leukemia (AML). However, non-APL types of AML respond poorly to ATRA-induced differentiation, and the mechanism underlying cell-type-specific resistance against ATRA remains unclear. Here, we use single-cell transcriptome analysis to compare the differentiation trajectories of two AML cell types during ATRA treatment. We show that in NB4 (APL/AML-M3) cells, ATRA activates canonical myeloid lineage factors—including SPI1, CEBPE, and STAT1—to direct near-normal differentiation toward mature granulocytes. By contrast, in HL60 (AML-M2) cells, ATRA-induced differentiation is incomplete and promiscuous, which is characterized by coinduction of both myelopoiesis and lymphopoiesis gene expression programs, as well as transient activation of cis-regulatory elements associated with myeloid differentiation. Our study suggests that the differentiation inducing capacity of ATRA in certain subtypes of AML may be compromised by therapy-induced lineage promiscuity.

INTRODUCTION

The hematopoietic system is generated by a hierarchy of stem/progenitor cells in the bone marrow, which proliferate and differentiate along distinct cell lineages, leading to terminal differentiation of a multitude of specialized cell types of the myeloid compartment (e.g. granulocytes, monocytes, macrophages, erythrocytes, mast cells, and dendritic cells) and the lymphoid compartment (e.g. B cells, T cells, and natural killer cells) (Laurenti and Gottgens, 2018). It has been suggested that leukemia often originates from aberrant differentiation blockage of hematopoietic progenitor cells, and unblocking such blockage may provide an efficient means of cure (Tenen, 2003). Indeed, this strategy, known as differentiation therapy, has been successfully applied in the treatment of acute promyelocytic leukemia (APL), also known as acute myeloid leukemia (AML)-M3 subtype, using all-trans retinoid acid (ATRA) in combination with arsenic compounds (Chen and Chen, 2017; Huang et al., 1988; Petrie et al., 2009; Wang and Chen, 2008). Earlier studies have suggested that ATRA has the potential to induce differentiation in non-APL types of AML as well as other cancers such as embryonic carcinoma, breast cancer, and neuroblastoma (de The, 2018; Petrie et al., 2009; Thiele et al., 1985). However, non-APL cancers generally respond poorly to ATRA-based differentiation therapy and the mechanisms underlying such cancer-specific resistance remain obscure.

ATRA belongs to the retinoid family of signaling molecules (Tang and Gudas, 2011). Activation of the retinoid signaling pathway is mediated by converting a transcriptional complex, composed of heterodimers of nuclear retinoid (RAR α , β and γ) and rexinoid (RXR α , β and γ) receptors, from transcription repressor to transcription activator (Altucci et al., 2007). In APL, t(15; 17) chromosomal translocation produces the PML-RARA fusion protein, which prevents the RA signaling pathway from being activated at physiological levels of retinoids (de The, 2018; Kakizuka et al., 1991; Tang and Gudas, 2011). A large body of evidence has shown that the transregulatory suppressor function of PML-RARA can be relieved by pharmacological levels of ATRA, leading to restoration of a normal granulocyte differentiation program and loss of clonal expansion capacity of APL-initiating cells (de The, 2018; Jiao et al., 2013) (Tan et al., 2020). In non-APL types of AML, a lack of therapeutic effect of ATRA is generally thought to be the lack of PML-RARA mutation, although, in principle, terminal differentiation programs may still be activated by ATRA via the intrinsic retinoid signaling pathway. However, the differentiation-inducing capacity of ATRA in most cancer cell types

¹Shanghai Institute of Hematology, State Key Laboratory of Medical Genomics, National Research Center for Translational Medicine at Shanghai, Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

²Interdisciplinary Research Center on Biology and Chemistry, Shanghai Institute of Organic Chemistry, Shanghai 200032, China

³These authors contributed equally

⁴Lead contact

*Correspondence: sjchen@stn.sh.cn (S.C.), lf12034@rjh.com.cn (F.L.)

<https://doi.org/10.1016/j.isci.2021.102410>



has not been well characterized, in part, owing to the difficulty in relating the developmental trajectory of tumor cells to those of their closely related cell types.

Here, we attempted to address the mechanisms of AML subtype-specific responses to ATRA using two well-established promyelocyte cell lines. The first cell line, NB4, was derived from a patient with APL with t(15; 17) chromosomal translocation. It expresses the PML-RARA fusion protein and can be induced by ATRA to terminally differentiate into granulocyte-like cells (Lanotte et al., 1991). The other line, HL60 (HL-60), was initially reported to derive from a patient diagnosed with APL (Collins et al., 1978) but was later re-defined as an AML-M2-like case owing to the absence of t(15; 17) translocation and its molecular profile featuring amplification of the Myc oncogene (Dalton et al., 1988). *In vitro*, HL60 can also be induced by ATRA to differentiate to granulocyte-like cells (Ramirez et al., 2017). However, a study using cDNA microarray experiments found few ATRA-regulated genes shared between NB4 and HL60 cells (Zheng et al., 2005), suggesting that very different gene expression programs were induced by ATRA in each cell type, although how such differences reflect ATRA-induced differentiation programs has not been explored.

In this study, we have applied single-cell RNA-seq analysis to delineate the trajectories of ATRA-induced differentiation in NB4 and HL60 cells, reasoning that a comparison of cell-type-specific differentiation processes may provide insights into the variations of ATRA's therapeutic efficacy between APL and non-APL types of AML. Our analysis indicates that in NB4, ATRA induces a near-normal process of terminal granulopoiesis. But, in HL60, ATRA-induced differentiation is incomplete and promiscuous, showing coactivation of granulopoiesis and lymphopoiesis gene expression programs. Through single-cell gene expression and gene network analyses, we also uncovered the complete sets of transcription factors underlying ATRA-induced differentiation. Furthermore, by profiling ATRA-regulated cis-regulatory elements during induced differentiation, we found that incomplete differentiation in HL60 cells is accompanied by weak and transient reprogramming of a group of cis-regulatory elements associated with myelopoiesis. We discuss the implication of these findings in understanding the efficiency of ATRA-based differentiation therapy in different subtypes of AML.

RESULTS

Overview of ATRA-regulated gene expression programs in NB4 and HL60 leukemia cell lines

To systematically compare ATRA-regulated gene expression programs between HL60 and NB4 cells, we first performed a side-by-side mRNA-seq experiment using each cell line cultured with 1 μ M of ATRA for 1/3/6 days (Figure 1A). As compared with untreated controls, 5461 and 4642 differentially expressed genes (DEGs) were identified in HL60 and NB4 cells after ATRA treatment (Wald test, padj <0.05), respectively. Most DEGs changed in a progressive manner, with 2438 genes shared between cell lines (1298 were upregulated; 1140 were downregulated) (Figures 1B and 1C, see also Tables S1 and S2). Gene ontology (GO) analysis indicated that the shared, upregulated DEGs were significantly associated with myeloid differentiation, neutrophil activation, and interferon response; the shared, downregulated DEGs were significantly associated with processes such as the cell cycle and metabolism (Figure 1D). Furthermore, 1832 activated DEGs private to NB4 showed enrichment of genes related to myeloid development and maturation (e.g. interferon response and autophagy) during the course of ATRA treatment, whereas 1149 activated DEGs private to HL60 only showed enrichment in myeloid activation genes at day 1 after treatment (Figure 1D). These studies indicated that even though NB4 and HL60 cells can both be induced to differentiate toward mature hematopoietic cell types, there were significant differences in the gene expression programs induced in each cell type, with ATRA-induced NB4 cell expressing more genes associated with the myeloid lineage.

To assess how these transcriptome profiles correlated with differentiation, we used the CIBERSORT algorithm to compute the relative similarity of each transcriptome to those a panel of major immune cell types (Newman et al., 2015). The transcriptome signature of granulocytes was most strongly and consistently increased in ATRA-treated NB4 cells (p value <0.05, one-way ANOVA), but its change in ATRA-treated HL60 cells was not as significant. By contrast, the signature of NK cells showed notable, albeit weak, increases in ATRA-treated HL60 cells (p value <0.05, one-way ANOVA), but such change was less obvious in ATRA-treated NB4 cells (Figure 1E). In addition, we retrieved genes associated with myeloid and lymphoid differentiation from the GO database and found that most of them showed ATRA-dependent upregulation in a cell-type-specific manner. In particular, a group of myeloid-differentiation genes (e.g. CEBPE, BATF2, MEF2C) were activated to high levels in NB4 but not in HL60. In addition, a group of lymphocyte-differentiation TFs showed coordinated increase in ATRA-induced HL60 cells (e.g. NFATC2, RUNX2, RUNX3), in consistence with weak increase of lymphocyte-related transcriptome signatures (Figure 1F; see also Figures S1 and S2, Table S3).

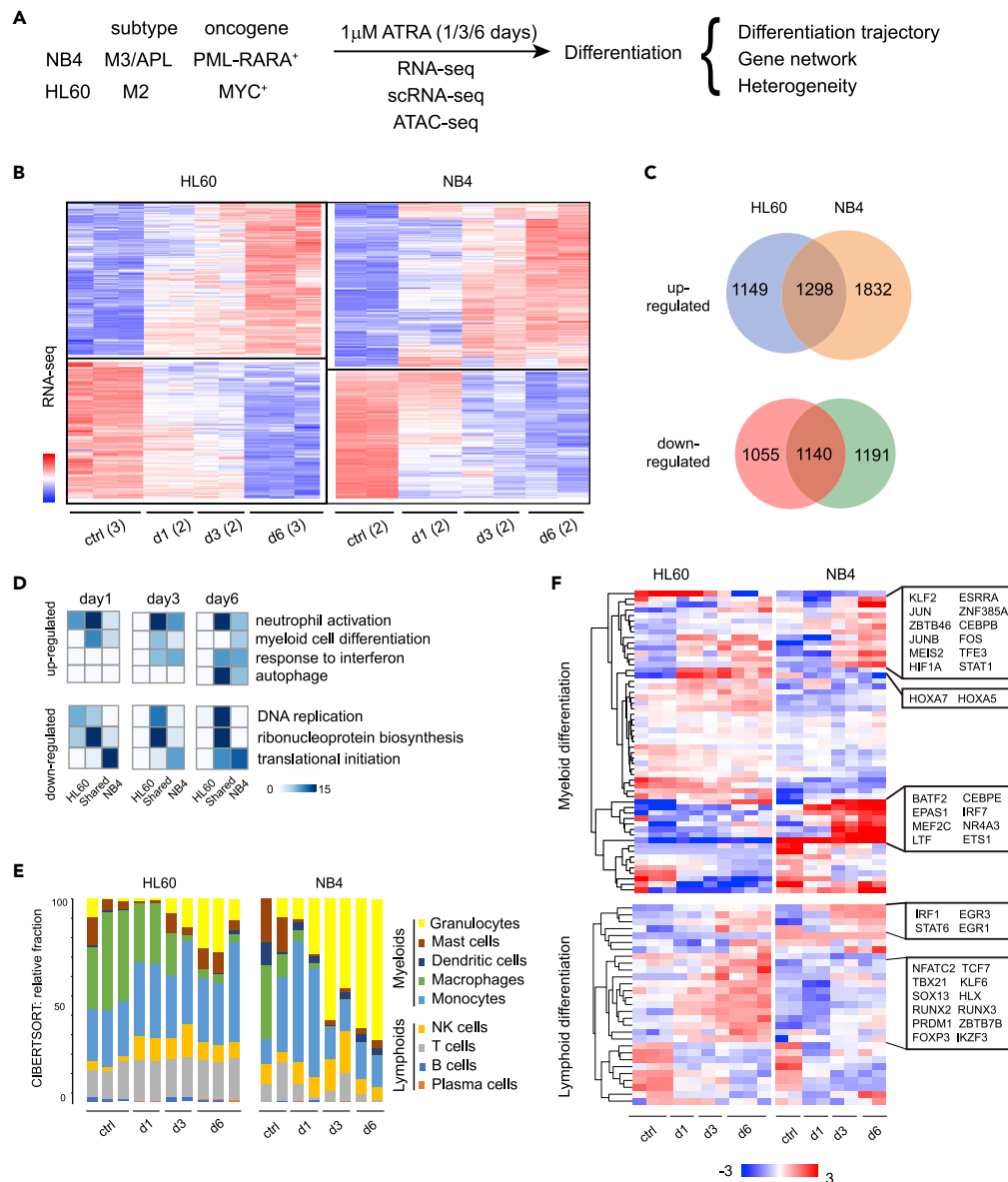


Figure 1. Overview of cell type-specific responses of leukemia cells to ATRA

(A) Experimental scheme. NB4 and HL60 cells were incubated with 1 μ M of ATRA in culture and collected at different time points for bulk and single cell RNA-seq analysis.

(B) Heat map of DEGs in the RNA-seq data using HL60 and NB4 cells treated with 1 μ M of ATRA for 1/3/6 days. Ctrl: control/untreated; day 1/3/6: day 1/3/6 after treatment. Numbers of biological repeats were shown in brackets.

(C) Venn diagrams of shared and cell type-specific DEGs between cell lines.

(D) Heat map of enrichment scores (-Log₁₀p value) of select GO terms related to ATRA-regulated biological processes.

(E) CIBERSORT analysis of the RNA-seq data. The relative fraction of the transcriptome signatures of major immune cell types is shown.

(F) Heatmap of TFs associated with myeloid and lymphoid differentiation in the RNA-seq data. Select TF clusters genes are indicated on the right.

Together, these results not only confirmed that ATRA induced highly cell-type-specific gene expression programs in NB4 and HL60 cells but also revealed a certain degree of heterogeneity of the gene regulatory networks activated in both cell types. Nevertheless, in these bulk RNA-seq experiments, a pool of cells were used as input, and the results only provided averaged gene expression profiles, which were insufficient to resolve the fate of individual cells or potential heterogeneity in their differentiation trajectories

(Kolodziejczyk et al., 2015). Therefore, in subsequent experiments, we turned to Drop-seq (Macosko et al., 2015), a high-throughput droplet-based scRNA-seq platform, to investigate the trajectories of ATRA-induced differentiation in each cell type at single-cell levels.

Single-cell analysis of ATRA-induced NB4 cells

We started our single-cell analysis with ATRA-induced NB4 cells, with the aim of using this PML-RARA⁺ APL cell type as reference for studying ATRA-induced differentiation trajectory. To do so, we collected NB4 cells cultured with 1 μ M of ATRA treatment for 1/3/6 days (untreated cells were used as controls) and performed Drop-seq experiments in replicates for each time point. A strict quality control pipeline was used to remove doublets and low-quality cells (see [transparent methods](#) for details). In total, this data set consisted of 6,135 single-cell transcriptomes (1,673 control/untreated cells; 2,068 ATRA day 1 cells; 1,261 ATRA day 3 cells; 1,133 ATRA day 6 cells). An unsupervised clustering analysis using this data set divided the cells into 7 clusters; of which, five (c1–c5) were close to each other, whereas the other two (c6 and c7) were separated (Figure 2A). Based on gene expression profiles and the duration of exposure to ATRA, the first six clusters (c1–c6) could be placed in a spectrum of differentiation toward granulocytes (Figures 2B–2D; see also Figure S3, Table S4). On one side of the spectrum were cells of c1, most of which were untreated, control cells, and marked by cell cycle genes (e.g. MKI67, a G1/S phase marker; TOP2A, an M phase marker), suggesting that they were proliferating naive NB4 cells (Figure 2C). On the other side of the spectrum were cells of c6, which resembled differentiated granulocyte-like cells because they were mainly composed of cells treated with ATRA for 6 days and marked by a large number of granulocyte genes (Figure 2B). For the intermediate clusters, c2 and c3 were marked by cell cycle genes and PRAM1, a known direct target of retinoid acid signaling (Moog-Lutz et al., 2001), suggesting that they had been exposed to ATRA but were yet to exit the cell cycle. Two other intermediate clusters, c4 and c5, were marked by many granulocyte differentiation-related genes, including three TFs—SPI1, CEBPE, and STAT1—that were known for directing granulocyte differentiation (Lekstrom-Himes et al., 1999; Nerlov and Graf, 1998; Wu et al., 2020). The levels of these genes showed a gradual increase from c4 to c5 and c6 (Figure 2B), which was indicative of a gradual process of ATRA-induced differentiation.

On the other hand, the cells in another separate cluster (c7) also expressed granulocyte-marker-like ENO3, which might suggest them as a distinct granulocyte-like cell type (Figures 2A–2C). However, c7 cells were mainly from untreated (control) experiment. In addition, they still expressed high levels of MKI67 but not granulopoiesis genes such as STAT1, CEBPE, and SPI1. Together, these features suggested that c7 represented a group of naive NB4 cells that spontaneously expressed a small set of granulocyte genes.

Next, we applied pseudotime analysis (Cacchiarelli et al., 2018) to infer ATRA-induced differentiation trajectory of NB4 cells. The resulted trajectory contained seven leaves (cellular states) connected by three branch points (Figure 2E). Projecting the sample collection time to this trajectory revealed that control cells resided in three states (s1–s3), with the majority in s1. Cells from day 1 were mainly in the same states as the control sample, yet the proportion of cells in s3 was greatly expanded (Figures 2F and 2G). In day 3 and day 6 samples, the fraction of cells in s3 progressively decreased, with a concomitant increase in s6 and then s7 (Figures 2F and 2G). These observations indicated s1 as the starting/root state and s7, the induced differentiated state. Furthermore, they suggested that the observed differentiation process was nonlinear, with multiple branches/states (i.e. s2, s3, and s6) transiently activated during the course of ATRA treatment.

Single-cell analysis of ATRA-induced HL60 cells

We then applied the same scRNA-seq analysis pipeline on ATRA-induced HL60 cells. This data set included 10,041 cells in total (1,075 control cells; 4,007 ATRA day 1 cells; 1,495 ATRA day 3 cells; 3,464 ATRA day 6 cells). Gene expression profile-based clustering divided them into six clusters that could also be placed in a spectrum of differentiation (Figures 3A–3D; see also Figure S4, Table S4). Cells of c1 were mainly composed of naive HL60 cells from the control sample and marked by the expression of cell cycle genes such as MKI67 and TOP2A. Three intermediate clusters (c2–c4) were mostly cells treated with ATRA for 1 and/or 3 days. The gene expression profiles of c2 cells were similar to those of c1, representing ATRA-treated cells (PRAM1⁺) that were yet to start differentiation; c3 and c4 were marked by increased expression of myeloid differentiation genes such as CD36, FCER1G, and S100A8. The last two clusters (c5 and c6) were largely composed of cells treated with ATRA for 6 days and expressed higher levels of differentiation-related genes (Figure 3C). Of note, three granulocyte differentiation TFs—CEBPE, SPI1, STAT1—marking

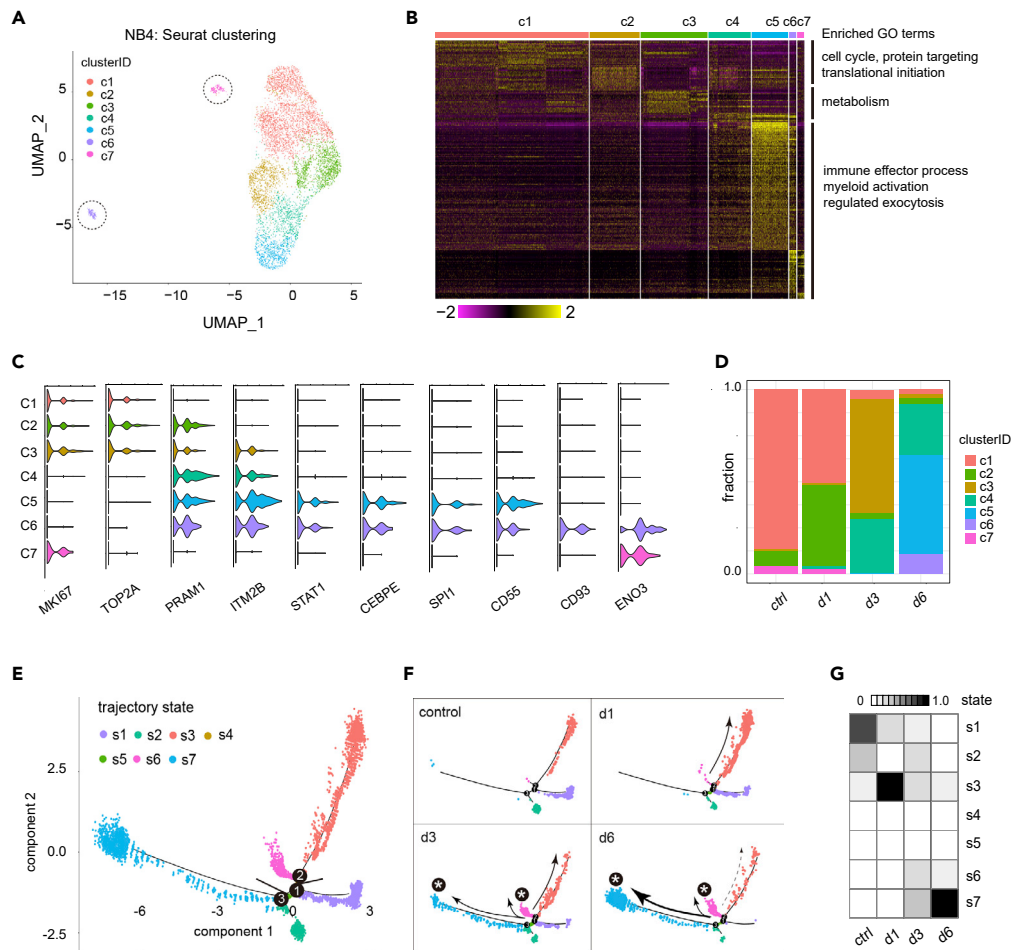


Figure 2. Single-cell analysis of ATRA-induced differentiation of NB4

(A) Seurat clustering scRNA-seq experiments using 6,135 NB4 cells with or without ATRA treatment for up to 6 days (1,673 control/untreated cells, 2,068 ATRA day 1 cells, 1,261 ATRA day 3 cells, and 1,133 ATRA day 6 cells). The majority of cells were closely positioned in U-MAP. Two separate clusters are circled.
 (B) Heat map of marker genes in different clusters. Enriched GO terms for different groups of genes are shown at right.
 (C) Violin plots of select genes showing cluster specificity.
 (D) Fraction of cells in different clusters per experiment.
 (E) Pseudotime trajectory constructed by monocle revealed seven states and three branch points.
 (F) Distribution of cells derived from different experiments in the trajectory tree.
 (G) The fraction of cells in different states per experiment.

ATRA-induced NB4 cells were not detectable in any group of ATRA-induced HL60 cell types; instead, FOXP1 was found to be expressed at high levels in differentiated cell types (i.e. those in c4–c6) (Figure 3C).

Pseudotime-based differentiation trajectory of ATRA-treated HL60 cells was composed of five cellular states connected by two branch points (Figure 3E). Untreated HL60 cells resided in three states (s1–s3), with the majority in s1 (Figures 3F and 3G). The fraction of cells in state s2 was expanded in the day 1 but then gradually decreased in day 3 and day 6. At the same time, there was an increase of cell fractions in states s3, s4, and s5. By day 6, most cells were in s5 and a minority in s3 (Figures 3F and 3G). These observations thus indicated s2 and s4 as intermediate states and s3–s5 as differentiated states.

Cell-specific differentiation trajectories induced by ATRA

We now return to comparing the differentiation trajectories of ATRA-induced NB4 and HL60 cells using the aforementioned single-cell data sets. Because multiple intermediates states appeared during induced

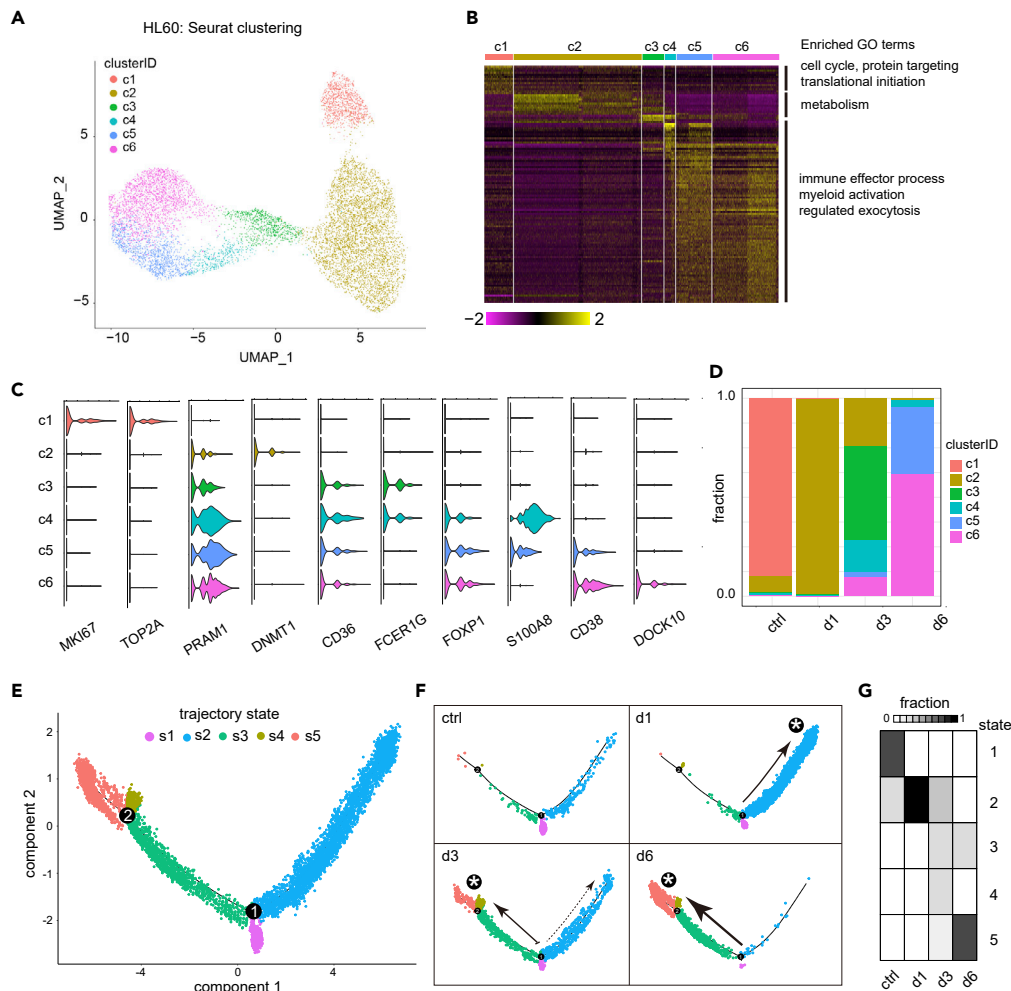


Figure 3. Single-cell analysis of ATRA-induced differentiation of HL60

(A) Seurat clustering scRNA-seq experiments using 14,389 cells HL60 cells with or without ATRA treatment for up to 6 days (2,118 control cells, 6,639 ATRA day 1 cells, 1,708 ATRA day 3 cells, and 3,924 ATRA day 6 cells). Expression based classification of cell clusters are labeled with colors.

(B) Heat map of marker genes in different clusters. Enriched GO terms for different groups of genes are shown at right.

(C) Violin plot of select genes showing cluster specificity.

(D) Fraction of cells in different clusters per experiment.

(E) Pseudotime trajectory constructed by monocle revealed five states and two branch points.

(F) Distribution of cells derived from different experiments in the trajectory tree.

(G) Heat map of the fraction of cells in different states per experiment.

differentiation of both NB4 and HL60, we performed branch-dependent gene expression analysis (Cacchiarelli et al., 2018) to determine if the same or different sets of genes were involved in cell fate decisions between them (Figures 4A–4G; see also Table S5).

In the NB4 trajectory, 3071, 128, and 966 branch-dependent genes were identified at branch points 1, 2, and 3, respectively. At each point, one set of genes was significantly associated with myeloid development, whereas another set, activated toward an alternative state, was associated with metabolic processes such as ribosome biogenesis and RNA processing. Furthermore, at branch point 1, which is close to the root state (i.e. s1), 31 myeloid differentiation genes were found (Figure 4B). Moving along the trajectory to branch point 2 toward the intermediate states s3 and s6, only 8 myeloid differentiation genes (8 of 31, 25.8%) were retrained (Figure 4C). By contrast, moving along the trajectory to branch point 3 toward the differentiated state (s7), most myeloid differentiation genes (21 of 31, 71.0%) were still present, and other

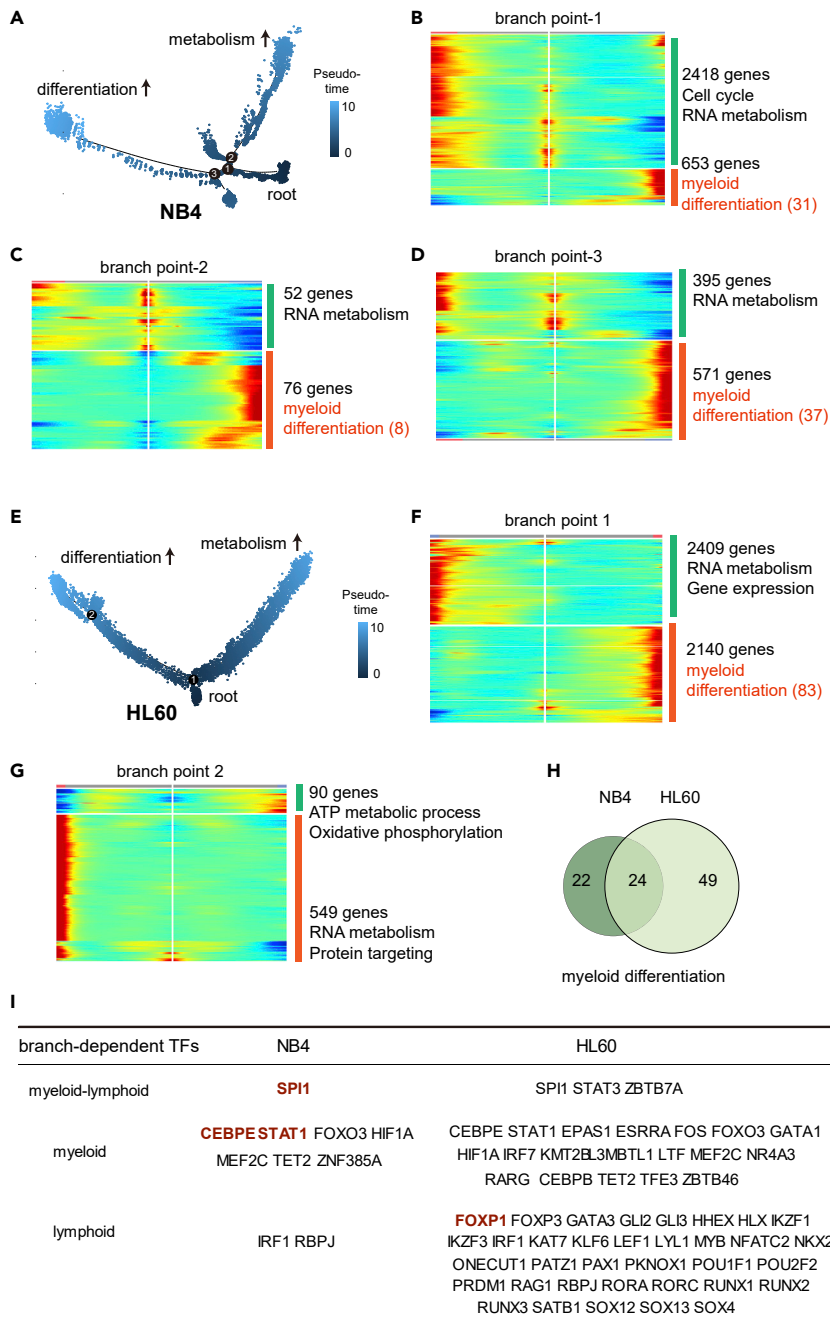


Figure 4. Comparison of ATRA-regulated gene sets during the differentiation of NB4 and HL60

(A) ATRA-induced NB4 differentiation trajectory as in Figure 2A. Cells were colored by pseudotime values. The root state and major branches with distinct gene expression profile changes are indicated.

(B–D) Heat map of genes differentially expressed at each branchpoint of the NB4 trajectory tree. Representative GO-enriched biological process terms are shown on the right. The number of genes in each category was indicated. Genes related to myeloid differentiation or activation are colored in red. The numbers of genes in each group are shown.

(E) ATRA-induced HL60 differentiation trajectory as in Figure 3A. Cells were colored by pseudotime values. The root state and two major branches with distinct gene expression profile changes are indicated.

(F and G) Heat map of genes differentially expressed at each branchpoint of the HL60 trajectory tree. Representative GO-enriched biological process terms are shown on the right. The number of genes in each category was indicated. Genes related to myeloid differentiation or activation are colored in red. The numbers of genes in each group are shown.

(H) Venn diagram of the number of myeloid differentiation genes in ATRA-induced NB4 and HL60 cells.

(I) Table of branch-dependent TFs in ATRA-induced NB4 and HL60 cells.

myeloid-related genes were induced (Figure 4D). These observations indicated that a differentiation program was continuously activated by ATRA along the path toward differentiated states but was lost when the cells entered the intermediate states.

In the HL60 trajectory, 4549 and 639 branch-dependent genes were identified at branch points 1 and 2, respectively. At branch point 1, one set of branch-dependent genes was associated with myeloid development (83 myeloid differentiation genes and 164 myeloid activation genes) and the other set was associated with metabolic activities including ribosome biogenesis and RNA processing. At branch point 2, however, none of the branch-dependent gene sets were significantly associated with myeloid development; instead, both were enriched with genes involved in RNA metabolism and mitochondria activities (Figures 4E–4G). Thus, while a myeloid differentiation program is initiated as the cells left the root state and progressed toward some differentiating states at the first branch point, this program appeared to be discontinued at the second branch point.

Although many branch-dependent genes in both NB4 and HL60 were found to be associated with hematopoietic cell differentiation, only a few of them were shared (Figure 4H). Notably, the branch-dependent TFs in NB4 were predominantly associated with granulopoiesis, but only about one-half of the branch-dependent TFs in HL60 were associated with granulopoiesis, whereas the other half were associated with lymphopoiesis (Figure 4I). These results confirmed the coexpression of myelopoiesis and lymphopoiesis TFs, including RUNX2 and RUNX3, observed in the bulk RNA-seq analysis shown previously (compare Figures 4I and 1F).

Cell-type-specific expression of lineage factors during ATRA-induced differentiation

The single-cell differentiation trajectories of both NB4 and HL60 cells contained multiple differentiating/differentiated states and intermediate metabolically active states. We reasoned that those aligned along the differentiation axis might be used for a more focused analysis of ATRA-induced differentiation program. Therefore, we selected cells from each data set three states that constituted the shortest path from control cells to ATRA-induced differentiating/differentiated cell types and reordered them by pseudotime values (Figure 5A). The resulted trajectory of each cell type was a three-leaf structure containing a single branch point in which most cells were aligned along the major axis, and their pseudotime-time values were largely in accordance with the duration of ATRA treatment (Figure 5A).

Next, we retrieved all TF genes showing pseudotime-dependent expression patterns in each subtrajectory (p value <0.001, Figures 5B and 5C). Of them, 31 and 27 TFs were activated and repressed in NB4, respectively; 79 and 5 TFs were activated and repressed in HL60, respectively. Only 9 TFs were shared (five were activated: IRF1, HIF1A, STAT1, NCOA3, and MTF1; four were repressed: DNMT1, MYB, PA2G4, and YBX1) (Figures 5B and 5C). The pseudotime-dependent genes in NB4 were only significantly associated with myeloid lineages, whereas those in HL60 were significantly associated with both myeloid and lymphocyte lineages (Figure 5D). Furthermore, at the single branch point in the NB4 subtrajectory, canonical granulocyte differentiation TFs such as SPI1, CEBPE, and STAT1 were again found to be upregulated along the axis of differentiation (Figure 5E). However, in the branch point of HL60, the branch-dependent TFs were YBX1, NR6A1, HIVP3, and NFAT5. Of them, only NFAT5 was previously implicated in T cell development (Trama et al., 2000). Thus, although many TFs associated myeloid and lymphocyte lineages were induced in HL60, they appeared not to be involved in the fate decision at a late phase of differentiation (i.e. after 3 days post ATRA treatment).

Cell-type-specific gene regulatory networks induced by ATRA

To further determine how gene regulatory networks were regulated by ATRA, we used the SCENIC (Single-Cell rEgulatory Network Inference and Clustering) algorithm to compute the activities of coregulated TF-target gene sets (regulons) in the scRNA-seq data (Aibar et al., 2017). In both NB4 and HL60, the differentiated cell clusters annotated by gene expression profiles were grouped together by the regulon specificity scores (RSS), consistent with the expectation that TFs are critical regulators of cell differentiation (Figures 6A, 6B, 6F, and 6G; see also Figures S5 and S6, Table S6).

Examination of top ATRA-induced regulons in both NB4 and HL60 cells indicated that those highly increased in differentiating NB4 cells were primarily involved in the myeloid lineage (e.g. regulons of CEBPE, JUNB, ETS1, and NFE2L2; Figure 6D), whereas those highly increased in differentiating HL60 cells

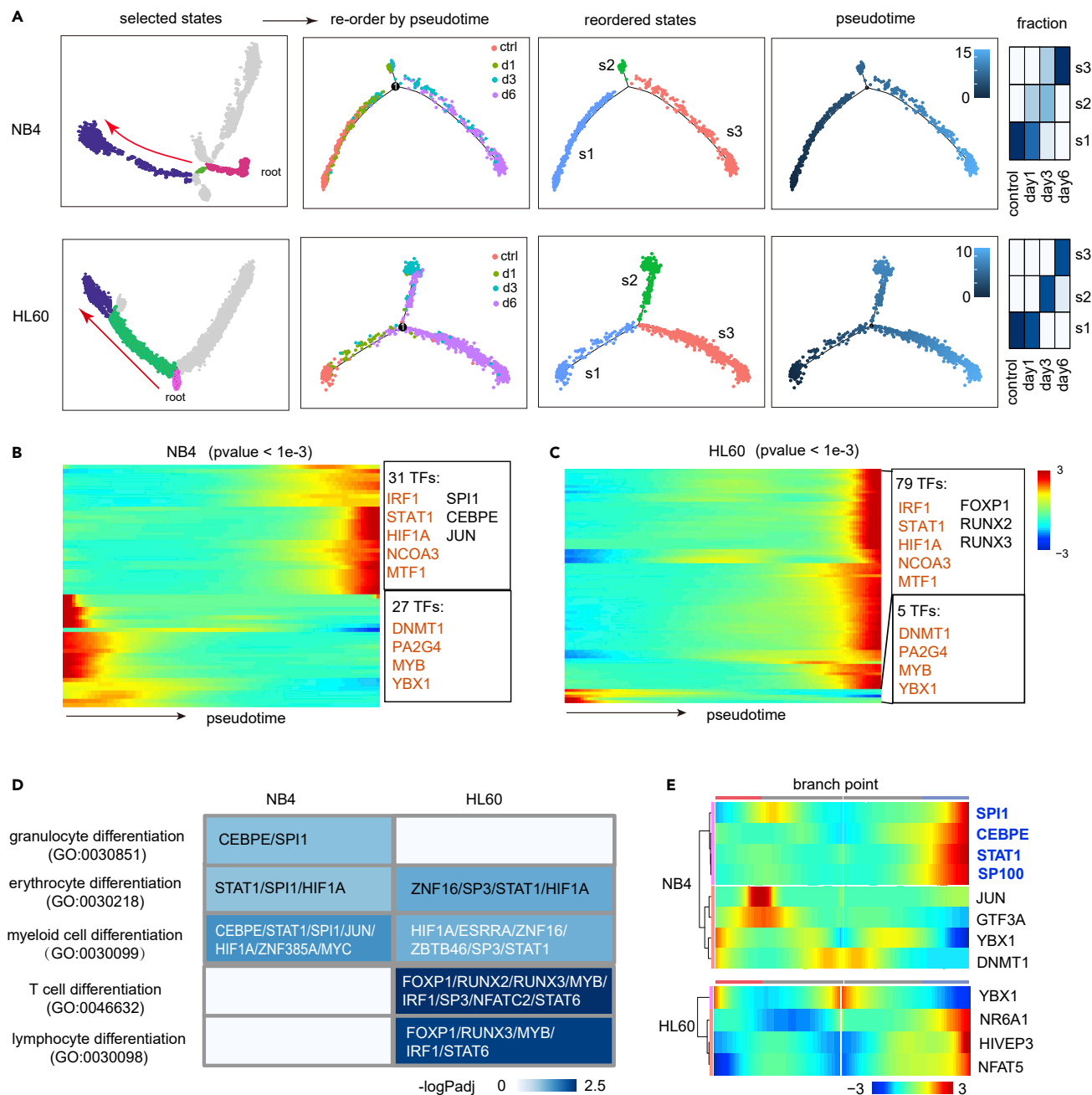


Figure 5. Comparison of the differentiation trajectories of ATRA-induced NB4 and HL60

(A) Cellular states along the differentiation path of NB4 and HL60 cells were selected and reordered by pseudotime. The resulted trajectories were colored by experiments, cellular states, and pseudotime values. Relative fraction of cells from different experiment in each state is shown on the right.

(B and C) Heat map of pseudotime-dependent expression of TFs in NB4 (B) and HL60 (C) (p value < 0.001). Select TFs are shown at right. Common branch-dependent TFs are colored in red.

(D) Heat map of GO terms related to myeloid and lymphoid differentiation. Pseudotime-dependent TFs of each GO term are shown.

(E) Heat map of branch-dependent TFs at the single branch points of the differentiation trajectories.

were primarily involved in the lymphoid lineage (e.g. regulons of RUNX2 and FOXP1; Figure 6I). Although the activities of many regulons were increased in both cell types, their induction levels notably differ. For example, in NB4, the top-ranked regulon was that of CEBPE, whose RSS increased by more than 5 times after ATRA induction, whereas its increase in HL60 cells was relatively small (Figures 6C and 6H). The

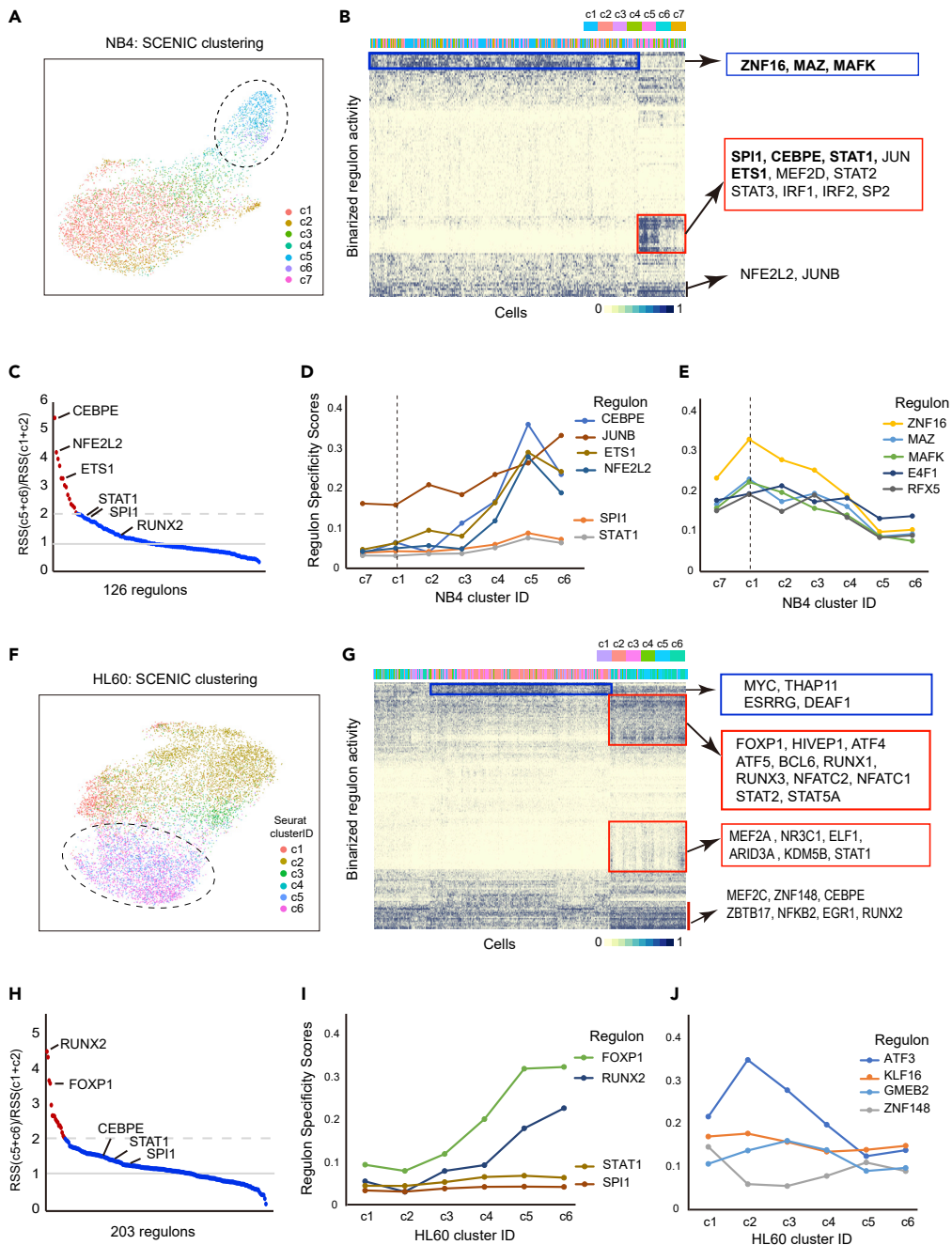


Figure 6. Single-cell gene regulatory network analysis of ATRA-induced NB4 and HL60 cells

(A) UMAP plot using the SCENIC regulonAUC scores of NB4 scRNA-seq data. Cells are colored by their Seurat cluster ID as in Figure 2A. Circle indicates clustered differentiated cell types (c5 and c6).

(B) Heatmap of binarized regulonAUC scores of SCENIC analysis of NB4 scRNA-seq data. X axis are cells; y axis are regulons (named by TF regulators).

(C) Dot-plot of fold changes of RSS of 126 regulons in differentiated (c5 and c6) versus undifferentiated (c1 and c2) clusters of NB4.

(D and E) Plots of regulon specificity scores (RSS) of selected regulons in each NB4 Seurat cluster.

(F) UMAP plot using the SCENIC regulonAUC scores of HL60 scRNA-seq data. Cells are colored by their Seurat cluster ID as in Figure 3A. Circle indicates clustered differentiated cell types (c5 and c6).

Figure 6. Continued

(G) Heatmap of binarized regulonAUC scores of SCENIC analysis of HL60 scRNA-seq data. X axis are cells; y axis are regulons (named by TF regulators).

(H) Plot of fold changes of RSS of 203 regulons in differentiated (c5 and c6) vs. undifferentiated (c1 and c2) clusters of HL60.

(I and J) Plots of regulon specificity scores (RSS) of selected regulons in each HL60 Seurat cluster.

converse pattern was observed for the RUNX2 regulon, which is the most highly increased in the HL60 data set, but its increase in NB4 is relatively small (Figures 6C and 6H). On the other hand, the activities of a number of regulons were decreased after ATRA-induced differentiation in both cell lines, which may reflect the loss of gene regulatory activities associated with the identity of parental cell lines (Figures 6E and 6J).

Dynamic changes of chromatin accessibility in response to ATRA

Finally, as an independent assay of ATRA-regulated transcriptional programs during differentiation of NB4 and HL60 cells, we performed the assay for transposase-accessible chromatin using sequencing (ATAC-seq) to profile the landscape of chromatin accessibility of each cell line cultured with 1 μ M of ATRA for 1/3/6 days. This assay uses the Tn5 transposase to preferentially digest the chromatin regions with low-histone-binding activities, allowing for identification of accessible chromatin regions (ACRs) enriched with active cis-regulatory elements such as enhancers and promoters (Buenrostro et al., 2015). Using these data, we identified 3956 and 3402 ATAC-seq peaks showing dynamic changes over time in NB4 and HL60, respectively (Figures 7A and 7B, see also Table S7).

To explore the dynamic patterns of these ACRs, we used an unsupervised clustering analysis to divided them into four groups (Figures 7A and 7B). In NB4, the accessibility about half of the dynamical ACRs increased (groups I and II), whereas that of the other half decreased (groups III and IV), during the course of ATRA treatment (Figure 7A). Genomic Regions Enrichment of Annotations Tool analysis indicated that the activated ACRs, but not silenced ones, were significantly associated with genes related to leukocyte activation, degranulation, migration, and inflammatory response; in addition, the DNA sequences of these ACRs were enriched with binding sites for the SPI and CEBP protein families (Figure 7A). Taking into consideration of the induction of SPI1 and CEBPE in ATRA-induced NB4 cells, these results are indicative of cis-regulatory reprogramming of granulocyte differentiation driven by these two canonical granulopoiesis factors. By contrast, the overall change of chromatin accessibility in ATRA-regulated ACRs in HL60 is relatively weaker (Figure 7B). Although the accessibility of 1790 ACRs (groups i and ii) showed consistently increase during the course of ATRA-treatment, they were not significantly associated with myeloid development, which is in spite of significant enrichment of DNA binding motifs for the SPI protein family. Moreover, the accessibility of one group (group iv) in HL60 were transiently increased at day 1 after the treatment but then decreased to the baseline level. Interestingly, these transiently activated ACRs were associated with genes related to leukocyte migration and chemotaxis (Figure 7B). These results are in agreement with the aforementioned transcriptome analyses that ATRA did not induce a complete myelopoiesis program in HL60 cells.

DISCUSSION

Here, we have compared the differentiation trajectories of two AML model cell lines that represented the APL/AML-M3-subtype (i.e. NB4) and AML-M2-subtype (i.e. HL60) respectively. We find that in NB4, ATRA induces three key granulopoiesis factors—SPI1, CEBPE, and STAT1—to direct a differentiation trajectory resembling the process of terminal granulopoiesis. By contrast, in HL60, the granulopoiesis program is only weakly induced; at the same time, a lymphopoiesis program—primarily regulated by RUNX2, RUNX3, and FOXP1—is activated, making the differentiation trajectory of ATRA-induced HL60 cells appear incomplete and promiscuous. The distinction between these differentiation trajectories is mainly delineated by expression- and pseudotime-based single-cell transcriptome analyses, using samples collected at discrete time points during the course of ATRA treatment. Although such analyses are subjected to influences from sampling time and the number of cells obtained from each time point, the very different modes of ATRA-induced differentiation processes are also supported by bulk transcriptome and cis-regulatory landscape analyses in this study.

The seeming activation of a lymphopoiesis program in HL60 is somewhat unexpected, considering that it is a promyelocyte cell line derived from a patient with AML (Collins et al., 1978; Dalton et al., 1988). Nevertheless, most, perhaps all, signal transduction pathways are pleiotropic in terms of regulating different sets of

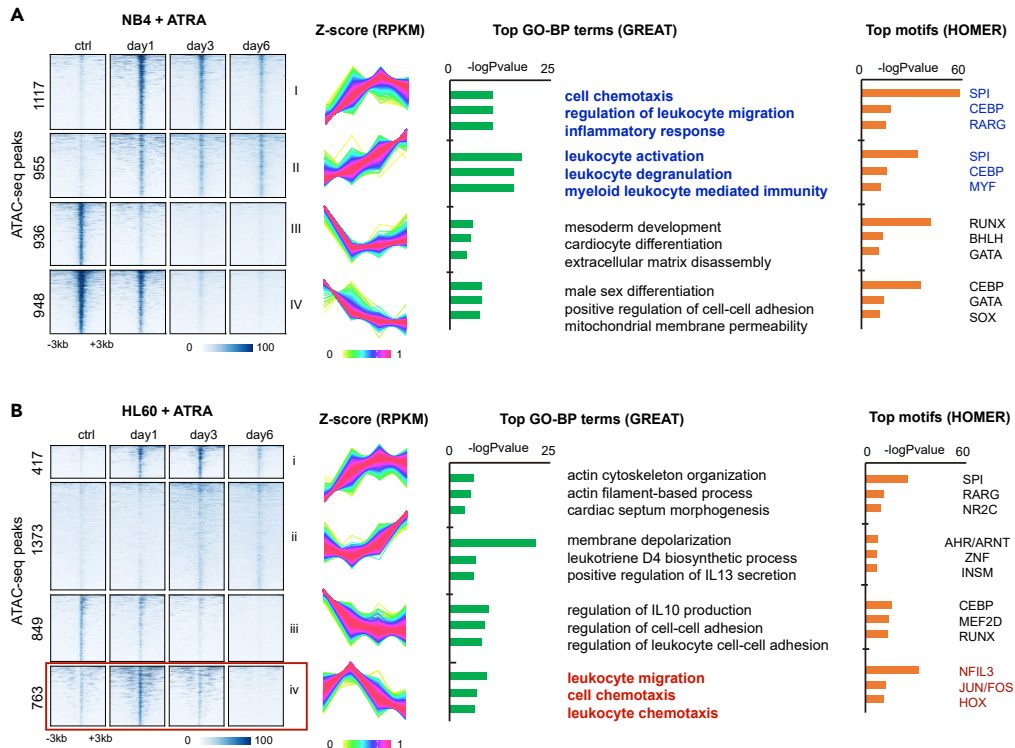


Figure 7. ATRA-regulated accessible chromatin during induced differentiation of NB4 and HL60

(A) ATAC-seq experiments of ATRA-treated NB4 cells. Dynamically changed accessible chromatin regions (ACRs) are shown, which are divided into four clusters by k-means clustering ($k = 4$). The heatmap of normalized peak intensity, plot of z-scores, top associated GO Biological Process (GO-BP) terms, and top enriched TF-binding motifs are shown from left to right. ACRs of groups I and II are associated with GO terms and TF related to myeloid development and leukocyte migration (color in blue).

(B) ATAC-seq experiments of ATRA-treated HL60 cells. Dynamically changed ACRs are also divided into four clusters by k-means clustering ($k = 4$). The heatmap of normalized peak intensity, plot of z-scores, top associated GO Biological Process (GO-BP) terms, and top enriched TF-binding DNA motifs are shown from left to right. ACRs of group iv (highlighted by red rectangle) are transiently activated during ATRA treatment and are associated with GO terms and TFs related to leukocyte migration (color in red).

target genes in different cellular context (Barolo and Posakony, 2002). For example, in early development of the hematopoietic system, the retinoid signaling pathway has been shown to be involved in the control of stem cell fate in addition to that of myeloid differentiation (Tang and Gudas, 2011). Thus, the retinoid pathway may have the potential to control some lymphopoiesis genes, which are normally silent but then inappropriately activated by ATRA in HL60. On the other hand, despite significant upregulation of many lymphopoiesis-related TFs, it appears that ATRA-induced HL60 cells did not undergo terminal differentiation along either the myeloid or lymphoid lineage. Instead, they seemed to be stalled at a cell fate decision point at about 3 days after ATRA treatment, whereby both canonical myelocyte- and lymphocyte-determining TFs ceased to driving further differentiation. The mechanism underlying the dissociation between these TFs and terminal differentiation in ATRA-treated HL60 cells is unclear, though one possibility is that the accessible chromatin landscape in HL60 may only favor myelopoiesis. This possibility is consistent with the observation that the dynamically changed accessible chromatin regions in both HL60 and NB4 are primarily enriched with binding motifs for myelopoiesis-TFs, but not lymphopoiesis-TFs.

Broadly speaking, the coinduction of lymphopoiesis and myelopoiesis programs in ATRA-induced HL60 cells could be referred to as a type of lineage promiscuity. Previously, the phenomenon of lineage promiscuity has been reported during early stage of the hematopoiesis, whereby some multipotent progenitors coexpress low levels of lineage-determining factors responsible for multiple cell types (e.g. myeloids and lymphoids) (Greaves et al., 1986; Miyamoto and Akashi, 2005). Developmental lineage promiscuity, however, is often transient and

disappears when the progenitor cells become committed toward differentiation into a given cell type (Laiosa et al., 2006). On the other hand, long-lasting lineage promiscuity has been shown in certain cancer cells, manifested as constitutive expression of markers of different cell types as a result of genetic mutations disrupting factors of gene regulation (Greaves et al., 1986). By contrast, the observed lineage promiscuity in HL60s is actively induced by ATRA. Therefore, it represents a type of therapy-induced lineage promiscuity, which becomes evident only after exposure to external stimulus (e.g. ATRA).

Our single-cell analysis also illustrated several layers of nuance in ATRA-induced cell differentiation. In the trajectories of both NB4 and HL60, there is at least one branch point bifurcating into two paths: one leads to differentiated cell states; the other leads to intermediate states showing downregulation of differentiation genes and upregulation of RNA processing and ribosome biogenesis genes, which might indicate a form of metabolically active state. Such metabolically active states appeared to be occupied by most cells treated by ATRA within the first three days (NB4 and HL60 alike), whereas differentiating/differentiated states become evident only after more than three days of continuous ATRA treatment. At present, it is not yet known whether cells on the path to differentiation originate from the high-metabolic cells or a small fraction of cells at earlier times that did not enter the high metabolic state(s). Furthermore, after six days of ATRA treatment, a small percent of NB4 cells appeared to remain undifferentiated (31 of 1133 ATRA day 6 cells in c1, 2.73%, Figure 2D). It is unclear whether these cells represented a group of rare ATRA-resistant cells that existed in the original cell culture or reflected intrinsic heterogeneous response to a given treatment through epigenetic reprogramming. Further studies using single-cell lineage tracing technologies may shed light on these issues (VanHorn and Morris, 2021).

In summary, our study uncovered two heretofore underappreciated side effects of ATRA's pleiotropic effects. First, it can activate a large number of genes related to ribosome biogenesis that do not appear to be directly linked with induced cell differentiation. Second, ATRA may simultaneously activate both myelopoiesis and lymphopoiesis programs in certain AML cells (i.e. HL60), leading to a mixed and incomplete differentiation phenotype. We suggest that induced lineage promiscuity may serve as a model to describe the lack of efficacy of ATRA-based differentiation therapy in non-APL cancer cells (e.g. HL60-like AML-M2 cells), and methods to restrict promiscuous gene expression programs may help enhance the efficiency of ATRA-dependent differentiation therapy in non-APL type of leukemia.

Limitations of the study

This study used two AML cell lines to investigate the trajectories of ATRA-induced differentiation. While NB4 cells has been shown by numerous studies to mimic the response of APL to ATRA-based differentiation therapy, it is not clear at present whether HL60 cells recapitulate the response of AML-M2 cells in patients *in vivo*. This is mainly because ATRA has not been approved for treating non-APL types of AML. Therefore, whether tumor cells from AML-M2 patients also show incomplete and lineage promiscuity remains to be verified. In future studies, this issue may be addressed by experiments using patient-derived leukemia cells cultured with ATRA *in vitro*. Furthermore, this study is performed using isolated cell cultures, which does not address potential influence of tumor cell microenvironment on the therapeutic effect of ATRA. Finally, the HL60 cell line has been commonly referred to as the AML-M2 subtype, based on the earlier morphology-based French-American-British classification system (Dalton et al., 1988). However, the current WHO classification system tentatively includes more than 20 AML subtypes, each characterized by distinct mutation profiles and pathophysiological features (Arber et al., 2016). The main genetic mutations of HL60 are known (MYC amplified; CDKN2A, NRAS, and TP53 mutated) (Dalton et al., 1988), but owing to significant heterogeneity in the mutation profiles between patients with AML and subtypes, it is still unclear which of the WHO-classified AML subtype HL60 might represent. Thus, cautions shall be taken in extrapolating this study to the interpretation of clinical responses of different AML subtypes to ATRA-based differentiation therapy.

Resource availability

Lead contact

Feng Liu (lf12034@rjh.com.cn)

Materials availability

No new or unique reagents were generated or used in this study.

Data and code availability

The next-generation sequencing data generated in this study are publicly available from The National Omics Data Encyclopedia (NODE; <https://www.biosino.org/node>) with the accession number OEP001921. NGS data analyses are performed using open-source tools as described in the [transparent methods](#) section in [supplemental information](#).

METHODS

All methods can be found in the accompanying [Transparent Methods supplemental file](#).

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2021.102410>.

ACKNOWLEDGMENTS

The authors thank Drs. Y Shen, JY Huang, Y Tan, YT Dai for critical reading and helpful comments on the manuscript. This study was supported by research funding from National Key Research and Development Program of China (2018YFA0107802), National Natural Science Foundation of China (81770143 and 81970130), Shanghai Commission of Science and Technology (17PJ1405800), Shanghai Municipal Education Commission Gaofeng Clinical Medicine Grant (20171902), Shanghai Pujiang Program (17PJ1405800), and Shanghai DF Scholarship.

AUTHOR CONTRIBUTIONS

F.L. S.C. Z.C. conceived and supervised the study. Y.T., X.T., F.L. conducted the experiments. Z.X. and J.C. provided technical assistance. F.L., Y.T., X.T., H.L., N.L. analyzed data and interpreted results. F.L. Z.C. S.C. Y.T. X.T. wrote the manuscript. All authors contributed to preparation, review, and approval of the manuscript.

DECLARATION OF INTERESTS

The authors declare no conflict of interest.

Received: January 12, 2021

Revised: March 16, 2021

Accepted: April 6, 2021

Published: May 21, 2021

REFERENCES

- Aibar, S., Gonzalez-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086.
- Altucci, L., Leibowitz, M.D., Ogilvie, K.M., de Lera, A.R., and Gronemeyer, H. (2007). RAR and RXR modulation in cancer and metabolic disease. *Nat. Rev. Drug Discov.* **6**, 793–810.
- Arber, D.A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M.J., Le Beau, M.M., Bloomfield, C.D., Cazzola, M., and Vardiman, J.W. (2016). The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* **127**, 2391–2405.
- Barolo, S., and Posakony, J.W. (2002). Three habits of highly effective signaling pathways: principles of transcriptional control by developmental cell signaling. *Genes Dev.* **16**, 1167–1181.
- Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21 29 21–21 29 29.
- Cacchiarelli, D., Qiu, X., Srivatsan, S., Manfredi, A., Ziller, M., Overbey, E., Grimaldi, A., Grimsby, J., Pokharel, P., Livak, K.J., et al. (2018). Aligning single-cell developmental and reprogramming trajectories identifies molecular determinants of myogenic reprogramming outcome. *Cell Syst.* **7**, 258–268 e253.
- Chen, Z., and Chen, S.J. (2017). Poisoning the devil. *Cell* **168**, 556–560.
- Collins, S.J., Ruscetti, F.W., Gallagher, R.E., and Gallo, R.C. (1978). Terminal differentiation of human promyelocytic leukemia cells induced by dimethyl sulfoxide and other polar compounds. *Proc. Natl. Acad. Sci. U S A* **75**, 2458–2462.
- Dalton, W.T., Jr., Ahearn, M.J., McCredie, K.B., Freireich, E.J., Stass, S.A., and Trujillo, J.M. (1988). HL-60 cell line was derived from a patient with FAB-M2 and not FAB-M3. *Blood* **71**, 242–247.
- de The, H. (2018). Differentiation therapy revisited. *Nat. Rev. Cancer* **18**, 117–127.
- Greaves, M.F., Chan, L.C., Furley, A.J., Watt, S.M., and Molgaard, H.V. (1986). Lineage promiscuity in hemopoietic differentiation and leukemia. *Blood* **67**, 1–11.
- Huang, M.E., Ye, Y.C., Chen, S.R., Chai, J.R., Lu, J.X., Zhao, L., Gu, L.J., and Wang, Z.Y. (1988). Use of all-trans retinoic acid in the treatment of acute promyelocytic leukemia. *Blood* **72**, 567–572.
- Jiao, B., Ren, Z.H., Liu, P., Chen, L.J., Shi, J.Y., Dong, Y., Ablain, J., Shi, L., Gao, L., Hu, J.P., et al. (2013). 8-CPT-cAMP/all-trans retinoic acid targets t(11;17) acute promyelocytic leukemia through enhanced cell differentiation and PLZF/RARalpha degradation. *Proc. Natl. Acad. Sci. U S A* **110**, 3495–3500.
- Kakizuka, A., Miller, W.H., Jr., Umesono, K., Warrell, R.P., Jr., Frankel, S.R., Murty, V.V.,

- Dmitrovsky, E., and Evans, R.M. (1991). Chromosomal translocation t(15;17) in human acute promyelocytic leukemia fuses RAR alpha with a novel putative transcription factor. *PML Cell* 66, 663–674.
- Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C., and Teichmann, S.A. (2015). The technology and biology of single-cell RNA sequencing. *Mol. Cell* 58, 610–620.
- Laiosa, C.V., Stadtfeld, M., and Graf, T. (2006). Determinants of lymphoid-myeloid lineage diversification. *Annu. Rev. Immunol.* 24, 705–738.
- Lanotte, M., Martin-Thouvenin, V., Najman, S., Balerini, P., Valensi, F., and Berger, R. (1991). NB4, a maturation inducible cell line with t(15;17) marker isolated from a human acute promyelocytic leukemia (M3). *Blood* 77, 1080–1086.
- Laurenti, E., and Gottgens, B. (2018). From haematopoietic stem cells to complex differentiation landscapes. *Nature* 553, 418–426.
- Lekstrom-Himes, J.A., Dorman, S.E., Kopar, P., Holland, S.M., and Gallin, J.I. (1999). Neutrophil-specific granule deficiency results from a novel mutation with loss of function of the transcription factor CCAAT/enhancer binding protein epsilon. *J. Exp. Med.* 189, 1847–1852.
- Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161, 1202–1214.
- Miyamoto, T., and Akashi, K. (2005). Lineage promiscuous expression of transcription factors in normal hematopoiesis. *Int. J. Hematol.* 81, 361–367.
- Moog-Lutz, C., Peterson, E.J., Lutz, P.G., Eliason, S., Cave-Riant, F., Singer, A., Di Gioia, Y., Dmowski, S., Kamens, J., Cayre, Y.E., et al. (2001). PRAM-1 is a novel adaptor protein regulated by retinoic acid (RA) and promyelocytic leukemia (PML)-RA receptor alpha in acute promyelocytic leukemia cells. *J. Biol. Chem.* 276, 22375–22381.
- Nerlov, C., and Graf, T. (1998). PU.1 induces myeloid lineage commitment in multipotent hematopoietic progenitors. *Genes Dev.* 12, 2403–2412.
- Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12, 453–457.
- Petrie, K., Zelent, A., and Waxman, S. (2009). Differentiation therapy of acute myeloid leukemia: past, present and future. *Curr. Opin. Hematol.* 16, 84–91.
- Ramirez, R.N., El-Ali, N.C., Mager, M.A., Wyman, D., Conesa, A., and Mortazavi, A. (2017). Dynamic gene regulatory networks of human myeloid differentiation. *Cell Syst.* 4, 416–429 e413.
- Tan, Y., Wang, X., Song, H., Zhang, Y., Zhang, R., Li, S., Jin, W., Chen, S.J., Fang, H., Chen, Z., et al. (2020). A PML/RARalpha direct target atlas redefines transcriptional deregulation in acute promyelocytic leukemia. *Blood* 137, 1503–1516.
- Tang, X.H., and Gudas, L.J. (2011). Retinoids, retinoic acid receptors, and cancer. *Annu. Rev. Pathol.* 6, 345–364.
- Tenen, D.G. (2003). Disruption of differentiation in human cancer: AML shows the way. *Nat. Rev. Cancer* 3, 89–101.
- Thiele, C.J., Reynolds, C.P., and Israel, M.A. (1985). Decreased expression of N-myc precedes retinoic acid-induced morphological differentiation of human neuroblastoma. *Nature* 313, 404–406.
- Trama, J., Lu, Q., Hawley, R.G., and Ho, S.N. (2000). The NFAT-related protein NFATL1 (TonEBP/NFAT5) is induced upon T cell activation in a calcineurin-dependent manner. *J. Immunol.* 165, 4884–4894.
- VanHorn, S., and Morris, S.A. (2021). Next-generation lineage tracing and fate mapping to interrogate development. *Dev. Cell* 56, 7–21.
- Wang, Z.Y., and Chen, Z. (2008). Acute promyelocytic leukemia: from highly fatal to highly curable. *Blood* 111, 2505–2515.
- Wu, S.F., Xia, L., Shi, X.D., Dai, Y.J., Zhang, W.N., Zhao, J.M., Zhang, W., Weng, X.Q., Lu, J., Le, H.Y., et al. (2020). RIG-I regulates myeloid differentiation by promoting TRIM25-mediated ISGylation. *Proc. Natl. Acad. Sci. U S A* 117, 14395–14404.
- Zheng, P.Z., Wang, K.K., Zhang, Q.Y., Huang, Q.H., Du, Y.Z., Zhang, Q.H., Xiao, D.K., Shen, S.H., Imbeaud, S., Eveno, E., et al. (2005). Systems analysis of transcriptome and proteome in retinoic acid/arsenic trioxide-induced cell differentiation/apoptosis of promyelocytic leukemia. *Proc. Natl. Acad. Sci. U S A* 102, 7653–7658.

iScience, Volume 24

Supplemental information

**Induced lineage promiscuity undermines
the efficiency of all-trans-retinoid-acid-induced
differentiation of acute myeloid leukemia**

Yijia Tang, Xin Tian, Zihan Xu, Junke Cai, Han Liu, Nan Liu, Zhu Chen, Saijuan Chen, and Feng Liu

Transparent Methods

Cell culture

NB4 cells were gift from M. Lanotte (Institut National de la Santé et de la Recherche Médicale, Paris) (Lanotte et al., 1991). HL60 (HL-60) cells were kindly provided by the Stem Cell Bank, Chinese Academy of Sciences. The identity of each cell line was independently authenticated by short tandem repeat (STR) DNA profiling and querying against the ATCC STR Database (Genewiz, Inc). Both cell lines were cultured in Iscove's Modified Dulbecco's Medium (IMDM) supplemented with 10% fetal bovine serum (MOREGATE), 100 units/mL penicillin and 100 mg/mL streptomycin (GIBCO) at 37 °C with 5% CO₂. ATRA was purchased from SelleckChem (Cat# S1653) and dissolved in DMSO as a stock solution at 1 mM. For cell differentiation induction experiment, the ATRA stock was diluted to 1 μM in cell culture media for indicated incubation time.

RNA-Seq

Total RNA was extracted from approximately 1 million cells. 1-2 μg of RNA per sample was used for sequencing library preparation using VAHTS Total RNA-seq Library Prep Kit for Illumina (Cat# NR601, Vazyme) following the manufacturer's recommendations. Multiplexed libraries were mixed and sequenced by Illumina Xten at the depth of 50-60 million 150 bp pair-end reads. The sequencing reads of each library were pseudoaligned to a transcriptome index file (based on Homo_sapiens.GRCh38.cdna.all.fa) using Kallisto (v0.46.0) (Bray et al., 2016). The resulted transcript compatibility counts were used for differential gene expression analyses using the DESeq2 R package (Love et al., 2014). CIBERSORT analysis was performed by uploading the gene expression matrix to the online tool server (<https://cibersort.stanford.edu/>), and LM22 signature gene file and 100 permutations were chosen for the analysis (Newman et al., 2015). The results of DEG analysis are summarized in Tables S1-S2. Genes associated with human myeloid and lymphoid differentiation are assembled from GO database (www.geneontology.org) and listed in Table S3.

Drop-seq

Single cell RNA-seq was performed using Drop-seq Library Protocol version 3.1 (12/28/15) according to Macosko et. al. (Macosko et al., 2015). Two biological replicates were performed for each experiment (control, ATRA_1d, ATRA_3d, and ATRA_6d) using NB4 or HL60 cells. Each library was sequenced in a flowcell lane by Illumina Xten using a 150bp paired-end sequencing kit with a custom read1 primer (Macosko et al., 2015). After sequencing, reads were filtered and sorted by their barcodes of origin and aligned to the reference transcriptome (GRCh38.gencode.v25.gene.annotation) using Drop-seq core computational protocol version 2.3.0 (<https://github.com/broadinstitute/Drop-seq/releases>). Mapped reads were quantified into UMI-filtered counts per gene. The resulted digital gene expression (DGE) matrices were imported to R programming language environment for clustering and pseudotime analyses.

Single cell transcriptome analysis

ScRNA-seq data analysis was performed as follows:

- 1) The DGE matrix of each experiment was imported to the Seurat R package (v3.1.5) (Butler et al., 2018) to filter out cells of low quality and doublets. Only cells expressing at least 200 genes and only genes that were expressed in at least 0.1% of all cells per sample were retained. Cells with more than 20% mitochondrial genes then removed. SCTransform was used to regress out variations of mitochondria gene expression, followed by principal component analysis using the RunPCA function. Based on the PCElbowPlot result, 8-12 principal components (PCs) were selected to identify clusters. Putative doublets were identified by the DoubletFinder (v2.0.2) R package assuming 10% doublet formation rate, after which highly expressed genes in each cluster were identified by the FindAllMarkers function (Wilcoxon rank sum test, `only.pos = TRUE`, `logfc.threshold = 0.1`). For each cluster, if the top 5 marker genes are mitochondria genes, then cluster was deemed as apoptotic and discarded. The DGE matrices containing the remaining cells were saved for subsequent data analysis.

2) the DGE matrices of each cell line (control, ATRA-day1, ATRA-day3, ATRA-day6) were merged and imported into Seurat for SCTransform and graph-based clustering as above.

Clusters with less than 1% of all cells in the dataset were discarded, and the remaining cells were subject to another round of SCTransform and graph-based clustering until all clusters contain more than 1% of all cells. These filtering and re-clustering steps were repeated once or twice to empirically remove putative poor quality cells that compose small and diffusive clusters without distinct cell type-specific markers. After the final round of cell filtering, marker genes for each cluster were identified by FindAllMarkers function (Wilcoxon rank sum test, `only.pos = TRUE`, `logfc.threshold = 0.2`). Clusters are grouped together if they express similar top marker genes and are in close contact with each other on the UMAP plot. The marker gene sets are summarized in Table S4.

3) The above post-clustering Seurat object was exported to the monocle R package (v2.16.0) for pseudotime analysis (Qiu et al., 2017). For cell ordering, we used the DEGs derived from the above bulk RNA-seq experiments instead of the highly expressed and highly variable genes computed from the scRNA-seq datasets, out of the concern that the latter was enriched with highly expressed house keeping genes but not developmentally important transcription factors that are generally expressed at low levels. After estimating the size factors and dispersions of the dataset, `reduceDimension (max_components=2, method='DDRTree')` was performed to build the trajectory tree. The state containing the largest fraction of cells in the control sample was designated as the root state. Branch analysis was performed using the BEAM function. Pseudotime-dependent genes were identified using `differentialGeneTest(fullModelFormulaStr="~sm.ns(Pseudotime)")`. Significant genes were those with $qvalue < 1e-3$. These results are summarized in Table S5.

4) SCENIC analysis was conducted using the pySCENIC workflow (version 0.9.18) (Van de Sande et al., 2020). Gene-motif ranking using 10 kb around the transcription start site (TSS) were used to determine the search space around the TSS, and a 20-thousand motif database was used for RcisTarget and GRNBoost2 to identify co-expression modules and regulons. The scores of the activity of each regulon were computed by AUCell (Van de Sande et al., 2020). Each cell's clusterID was imported from pre-processed Seurat objects. Regulon specificity

scores were the means of regulon activity scores of cells in each cluster. These results are summarized in Table S6.

ATAC-seq

ATAC-seq was performed using 50,000 cells according to the fast ATAC-seq protocol (Corces et al., 2017). The resulted libraries were mixed and sequenced by Illumina Xten at the depth of 40-60 million 150 bp paired-end reads per sample. De-multiplexed reads were mapped to the human genome (hg19) using Bowtie2 (v2.2.9) with default settings (Langmead and Salzberg, 2012). PCR duplicates were removed by SAMtools (v1.3.1) (Li et al., 2009). ATAC-seq peaks were called by MACS2 (v2.1.1.20160309) (Liu, 2014) using the following parameters: -f BAMPE -g hg -B -q 0.01. After removing ENCODE blacklisted regions (Amemiya et al., 2019), reproducible peaks between experimental repeats were used to assemble a merged peak set from all experiments. This merged peak set was used as input to generate a sequencing reads count table using the multicov function of bedtools (v2.25.0) (Quinlan and Hall, 2010). The count table was imported to the TCseq R package (v1.8.0) for time series analysis to identify peaks showing time-dependent changes of accessibility. Heatmaps of ATAC-seq peaks were generated by deepTools (v2.4.2) (Ramirez et al., 2016). Gene ontology enrichment analysis of the ATAC-seq experiments was performed using GREAT (v4.0.4) (McLean et al., 2010). Enriched TF binding motifs in each peak set were analyzed using the findMotifsGenome.pl function of HOMER (v4.1) (Heinz et al., 2010). The dynamically changed ATAC-seq peak sets are summarized in Table S7.

References

- Amemiya, H.M., Kundaje, A., and Boyle, A.P. (2019). The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Scientific reports* 9, 9354.
- Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology* 34, 525-527.
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature biotechnology* 36, 411-420.

Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., *et al.* (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nature methods* *14*, 959-962.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* *38*, 576-589.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods* *9*, 357-359.

Lanotte, M., Martin-Thouvenin, V., Najman, S., Balerini, P., Valensi, F., and Berger, R. (1991). NB4, a maturation inducible cell line with t(15;17) marker isolated from a human acute promyelocytic leukemia (M3). *Blood* *77*, 1080-1086.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* *25*, 2078-2079.

Liu, T. (2014). Use model-based Analysis of ChIP-Seq (MACS) to analyze short reads generated by sequencing protein-DNA interactions in embryonic stem cells. *Methods in molecular biology* *1150*, 81-95.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* *15*, 550.

Macosko, E.Z., Basu, A., Satija, R., Nemes, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., *et al.* (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* *161*, 1202-1214.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nature biotechnology* *28*, 495-501.

Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nature methods* *12*, 453-457.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. *Nature methods* *14*, 979-982.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* *26*, 841-842.

Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic acids research* *44*, W160-165.

Van de Sande, B., Flerin, C., Davie, K., De Waegeneer, M., Hulselmans, G., Aibar, S., Seurinck, R., Saelens, W., Cannoodt, R., Rouchon, Q., *et al.* (2020). A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nature protocols* *15*, 2247-2276.

Supplemental Figures

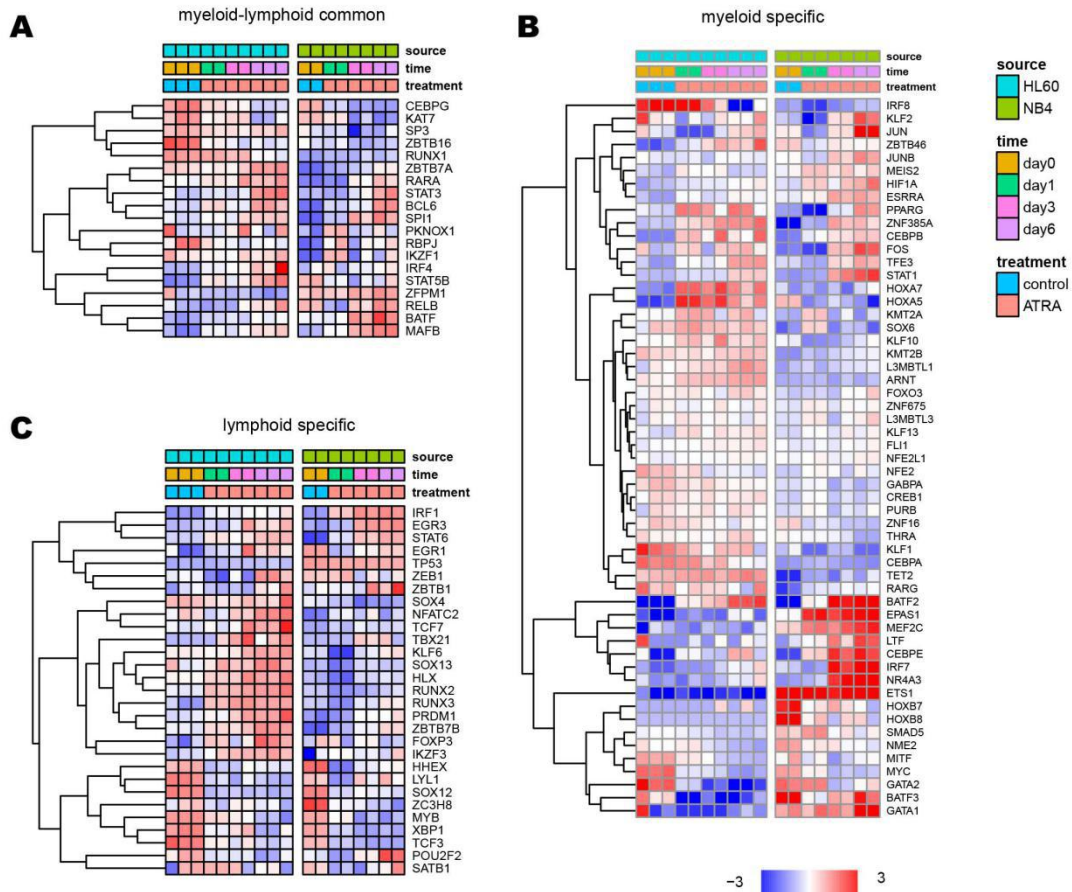


Figure S1. RNA-seq of TFs involved in myeloid and lymphocyte differentiation. Related to Figure 1.

Heatmap of RNA-seq data of genes associated with genes associated with both myeloid- and lymphoid-differentiation (A), with myeloid-differentiation only (B), or with lymphoid-differentiation only (C).

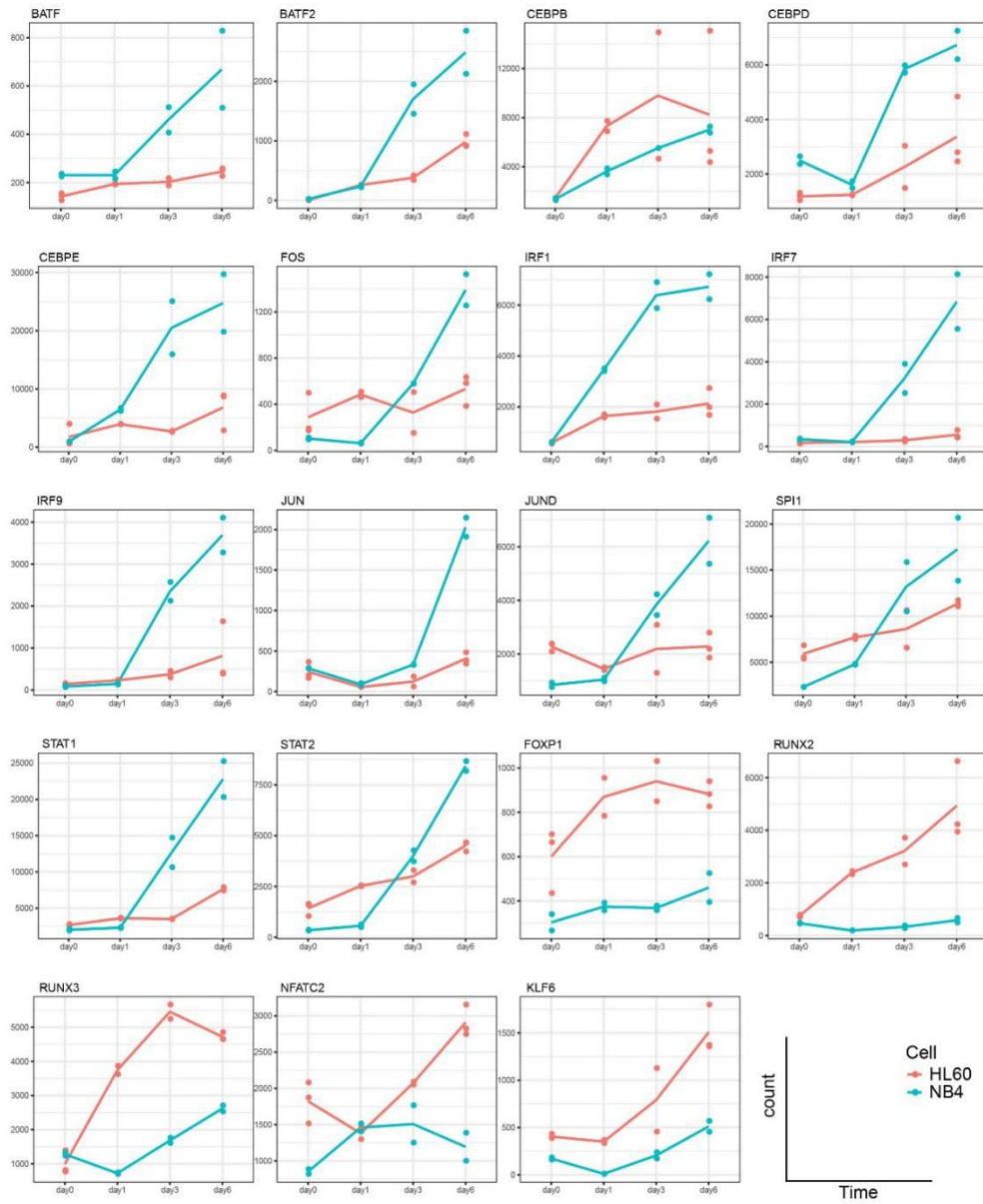


Figure S2. Expression of TFs regulating myeloid and lymphoid differentiation. Related to Figure 1.

RNA-seq data of in HL60 (red) and NB4 (blue) cells treated with 1 μ M ATRA for 0/1/3/6 days. Y-axis shows RNA-seq counts of each each gene. X-axis shows the time.

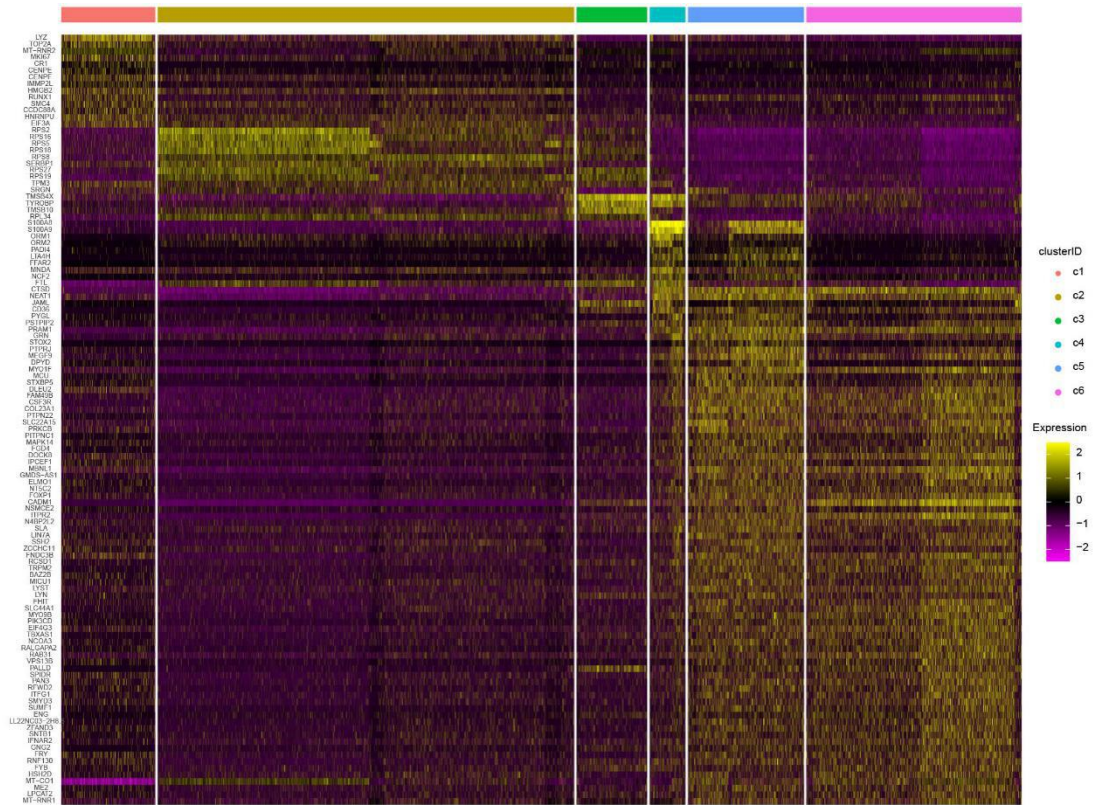


Figure S4. HL60 scRNA-seq analysis. Related to Figure 3.

Heatmap of cluster marker genes. Genes names are shown on the left. Cluster IDs are shown in the right.

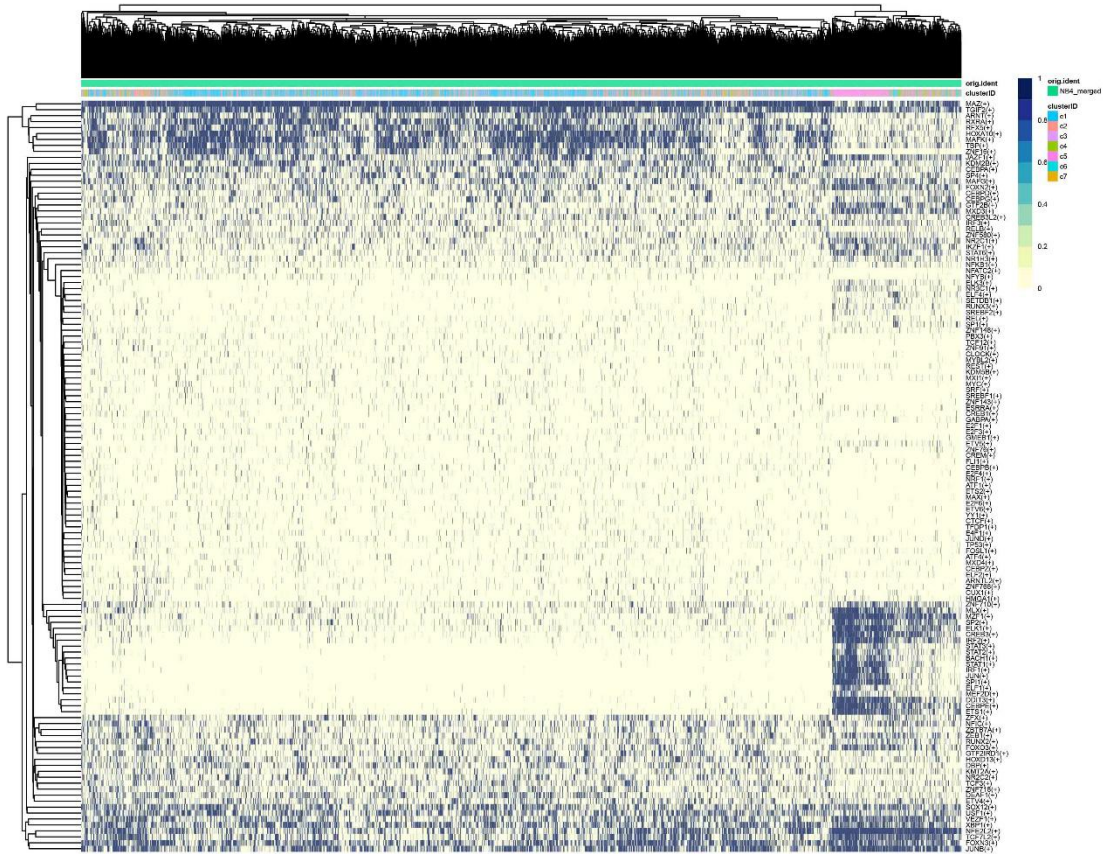


Figure S5. SCENIC analysis of NB4 scRNA-seq data. Related to Figure 6.

Heatmap binarized regulon activity scores of NB4. The names of TFs of each regulon are shown on the right.

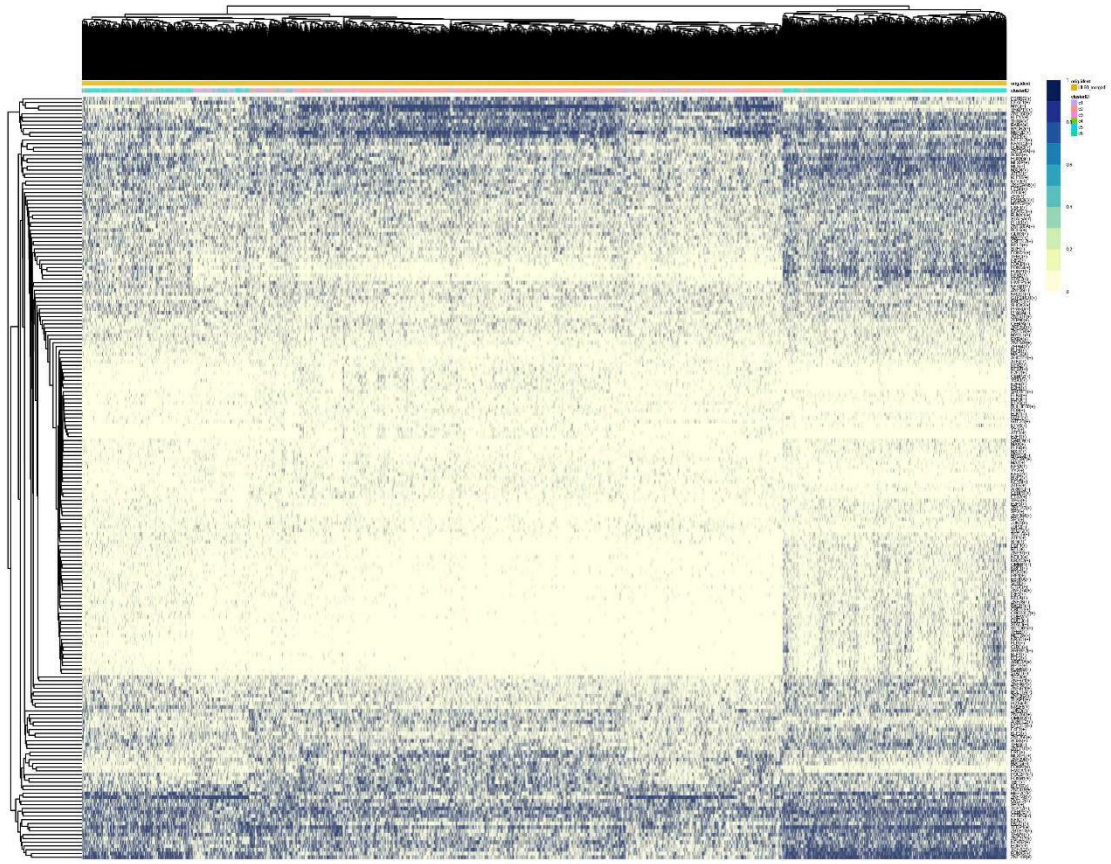


Figure S6. SCENIC analysis of HL60 scRNA-seq data. Related to Figure 6.

Heatmap binarized regulon activity scores of NB4. The names of TFs of each regulon are shown on the right.