CELL BIOLOGY

Multiplexed single-cell RNA-seq via transient barcoding for simultaneous expression profiling of various drug perturbations

Dongju Shin[1]*, Wookjae Lee[1]*, Ji Hyun Lee[2,3]†, Duhee Bang[1]†

The development of high-throughput single-cell RNA sequencing (scRNA-seq) has enabled access to information about gene expression in individual cells and insights into new biological areas. Although the interest in scRNA-seq has rapidly grown in recent years, the existing methods are plagued by many challenges when performing scRNA-seq on multiple samples. To simultaneously analyze multiple samples with scRNA-seq, we developed a universal sample barcoding method through transient transfection with short barcode oligonucleotides. By conducting a species-mixing experiment, we have validated the accuracy of our method and confirmed the ability to identify multiplets and negatives. Samples from a 48-plex drug treatment experiment were pooled and analyzed by a single run of Drop-Seq. This revealed unique transcriptome responses for each drug and target-specific gene expression signatures at the single-cell level. Our cost-effective method is widely applicable for the single-cell profiling of multiple experimental conditions, enabling the widespread adoption of scRNA-seq for various applications.

INTRODUCTION

Unlike conventional bulk measurements, single-cell RNA sequencing (scRNA-seq) permits analysis of the transcriptomes of individual cells (1–3), and this has shed light on the variations in cell populations, such as tumor heterogeneity. Platforms such as Drop-Seq (4), inDrop (5), and 10X Genomics Chromium (6) provide high-throughput single-cell information over thousands of cells. While the number of cells able to be profiled has increased and per-cell cost has dropped, challenges in scRNA-seq still include high sample preparation cost, ambiguous identification of true single cells, and sample-dependent batch effects (7), limiting the widespread adoption and scope of scRNA-seq. For experiments requiring the analysis of multiple single-cell samples (i.e., numerous samples of various conditions or samples from many patients), a separate scRNA-seq run must be conducted for each sample. Without the use of multiplexing, performing scRNA-seq for multiple samples is labor intensive and is limited by the high sample preparation cost rather than the per-cell cost in sequencing. Therefore, the demand for multiplexing samples in scRNA-seq is continuously increasing, requiring methods in which samples are pooled and subjected to a single scRNA-seq run.

To deal with the challenges stated above, a method has recently been developed for multiplexing samples from diverse patients by their endogenous genetic barcodes (8). In this approach, single-nucleotide polymorphisms in each patient serve as a sample barcode for determining the sample identity of each cell, enabling multiple samples to be pooled and sequenced simultaneously. While this approach can partially deal with multiplexing problems, applicable samples are limited to those that are genetically distinct. Still, multiplexing between samples genetically identical but in diverse conditions remains a challenge, necessitating a universal approach for barcoding and multiplexing samples regardless of the sample identities.

During drug discovery, gene expression profiling can be applied to annotate the function of small molecules (9) and to elucidate the mechanisms underlying a biological pathway (10, 11). While systematic approaches to profile gene expression for a large number of small molecules using microarray technology have been conducted (12, 13), they are limited to bulk measurements. To capture diverse responses of highly heterogeneous samples such as tumor cells, single-cell gene expression profiling is indispensable, although current technologies are not suitable for multiple screening. We envisioned that development of multiplexed scRNA-seq can provide simultaneous expression profiling of various drug perturbations in a very efficient manner.

Here, we designed a multiplexed scRNA-seq method that involved transient transfection of short barcoding oligos (SBOs) to label samples from various experimental conditions. We demonstrated that this method relying on simple transfection can be used for simultaneous single-cell transcriptome profiling for multiple drugs.

RESULTS

Design for transient barcoding method

SBO, a single-stranded oligodeoxynucleotides, consists of a sample barcode and a poly-A sequence (fig. S1A). Transient transfection of SBO allows a sample to be labeled with a unique barcode. The barcoded samples of various conditions are pooled and simultaneously processed for scRNA-seq (Fig. 1A). The poly-A sequence in the SBOs ensures that the mRNAs and SBOs are captured and reverse-transcribed together during the scRNA-seq process (fig. S1B). Computational analysis of digital count matrices of SBOs allows us to demultiplex and determine sample origins. This universal barcoding method, based on simple transfection, enabled sample multiplexing, identified multiplets and negatives, and reduced the preparation cost per sample.

### Validation of ability and accuracy for transient barcoding method

To demonstrate our method's ability and accuracy of multiplexing samples, we performed a 6-plex human/mouse species-mixing experiment. Two samples each of HEK293T and NIH3T3 cells carried

[1]Department of Chemistry, Yonsei University, Seoul, Korea. [2]Department of Clinical Pharmacology and Therapeutics, College of Medicine, Kyung Hee University, Seoul, Korea. [3]Department of Biomedical Science and Technology, Kyung Hee Medical Science Research Institute, Kyung Hee University, Seoul, Korea.
*These authors contributed equally to this work as co-first authors.
†Corresponding author. Email: duheebang@yonsei.ac.kr (D.B.); hyunihyuni@khu.ac.kr (J.H.L.)
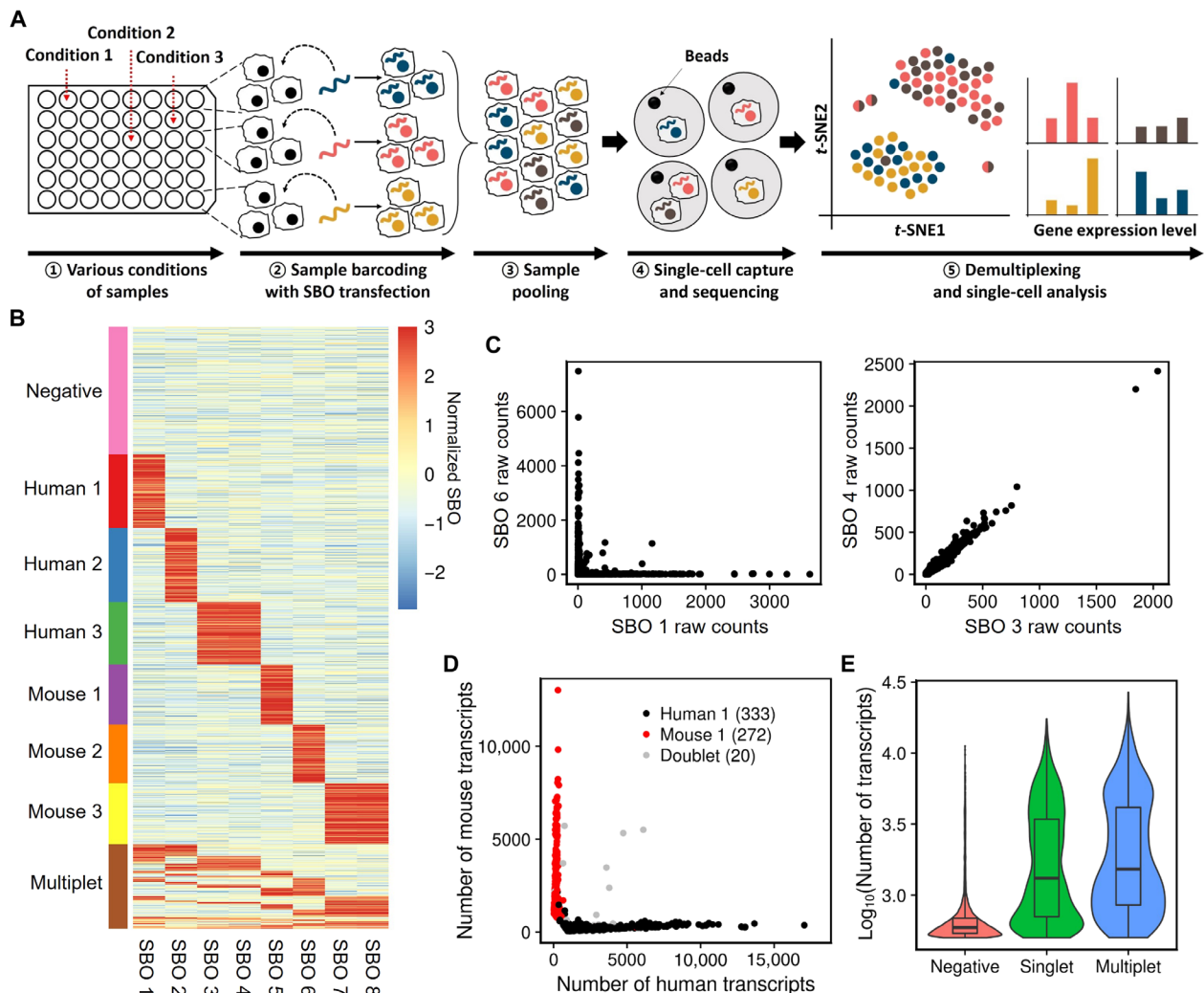
Fig. 1. Scheme and validation of transient barcoding method. (**A**) Scheme of multiplexed scRNA-seq by transient barcoding method using SBOs. (1) Samples with various conditions are prepared. (2) Each sample is transfected with SBO containing a unique sample barcode. (3) Barcoded cells are pooled together and processed for scRNA-seq (e.g., Drop-Seq). (4) Cells are lysed within droplets, and the released mRNAs and SBOs are captured, reverse-transcribed, and sequenced. (5) Cells are demultiplexed and assigned to their origins and processed for further analysis. (**B**) Heatmap of normalized SBO counts for 6-plex human/mouse species-mixing experiment. Rows represent cells, and columns represent SBOs. Cells are assessed whether they are positive for a particular SBO based on the SBO count matrix (see Materials and Methods). Cells were classified as singlets (positive for a unique SBO), multiplets (positive for more than one SBO), or negatives (not positive for any SBO) and ordered by their classifications. (**C**) Scatter plot showing raw counts between two SBOs. SBOs 1 and 6 were used to barcode different samples (Human 1, Mouse 2) (left). SBOs 3 and 4 were used to barcode the same sample (Human 3) (right). (**D**) Species-mixing plot of samples associated with SBOs 1 and 5. Cells were labeled according to their SBO classifications. Black dots indicate Human 1 sample barcoded with SBO 1, red dots indicate Mouse 1 sample barcoded with SBO 5, and gray dots indicate doublets that are positive for both SBOs. (**E**) Distribution of RNA transcript counts in cells between singlets (green), multiplets (blue), and negatives (red). Negatives, which imply beads exposed to ambient RNA, had the lowest number of transcripts. Multiplets had slightly more transcripts than singlets, indicating more RNA content within a droplet.

a single unique SBO, and one sample of each cell line carried a combination of the two SBOs (fig. S1C). We pooled all the samples together in equal proportions and performed a single run of Drop-Seq. Cells were deliberately overloaded during Drop-Seq to increase the chance of multiplets. We obtained 2759 cell barcodes, in which at least 500 transcripts were detected, and the cells were successfully assigned to their sample origins. Multiplets and negative cells were detected on the basis of the SBO count matrix (see Materials and Methods). Cells that were classified as singlets almost exclusively express their sample barcodes, while multiplets and negatives express multiple or no sample barcode, respectively (Fig. 1B). Scatter plots of SBO counts that

originated from two different samples showed an exclusive relationship, whereas SBO counts from the same sample showed a strong correlation in their expressions (Fig. 1C and fig. S2A). Species classification using SBOs was consistent with the transcriptome-based species-mixing plot results (Fig. 1D and fig. S2, B and C). We also observed a clear difference in the distribution of RNA transcripts between singlets, multiplets, and negatives as expected, indicating the unambiguous detection of multiplets and negatives (Fig. 1E). These results suggested that our method enabled sample multiplexing in single-cell experiments with high accuracy and specificity, and elimination of multiplets and negatives. We also verified that the SBO barcoding approach could be applied to

model heterogeneous samples without cell type–specific bias (fig. S3, A to E). Our data also demonstrate that transient transfection did not affect the gene expression profiles (fig. S3, F and G).

## Time-resolved expression profiles in drug perturbations

We envisioned that our method could be used when interested in screening multiple gene expression profiles in single cells subjected to drug perturbations. We performed a 5-plex time-course scRNA-seq in the K562 cell line, which is derived from chronic myeloid leukemia and expresses the *Bcr-Abl* fusion gene (*14*). Applying our multiplexing strategy, we investigated the single-cell transcriptional response of K562 cells to imatinib, a BCR–ABL–targeting drug (*15*),

over treatment time (see Materials and Methods). Unlike conventional scRNA-seq, by regressing out technical batch effects, multiplexed scRNA-seq applied here enables the detection of subtle transcriptional changes in the integrated analysis of multiple samples, facilitating more precise analysis. Following the drug treatments, samples were pooled, sequenced, and demultiplexed. After removing doublets and negatives, single cells were subjected to downstream analysis.

Pseudotime analysis of single cells in multiplexed samples collected from five time points showed a branched gene expression trajectory and a sequential progression in trajectory over drug treatment time (Fig. 2A). The branched trajectory showed that two transition
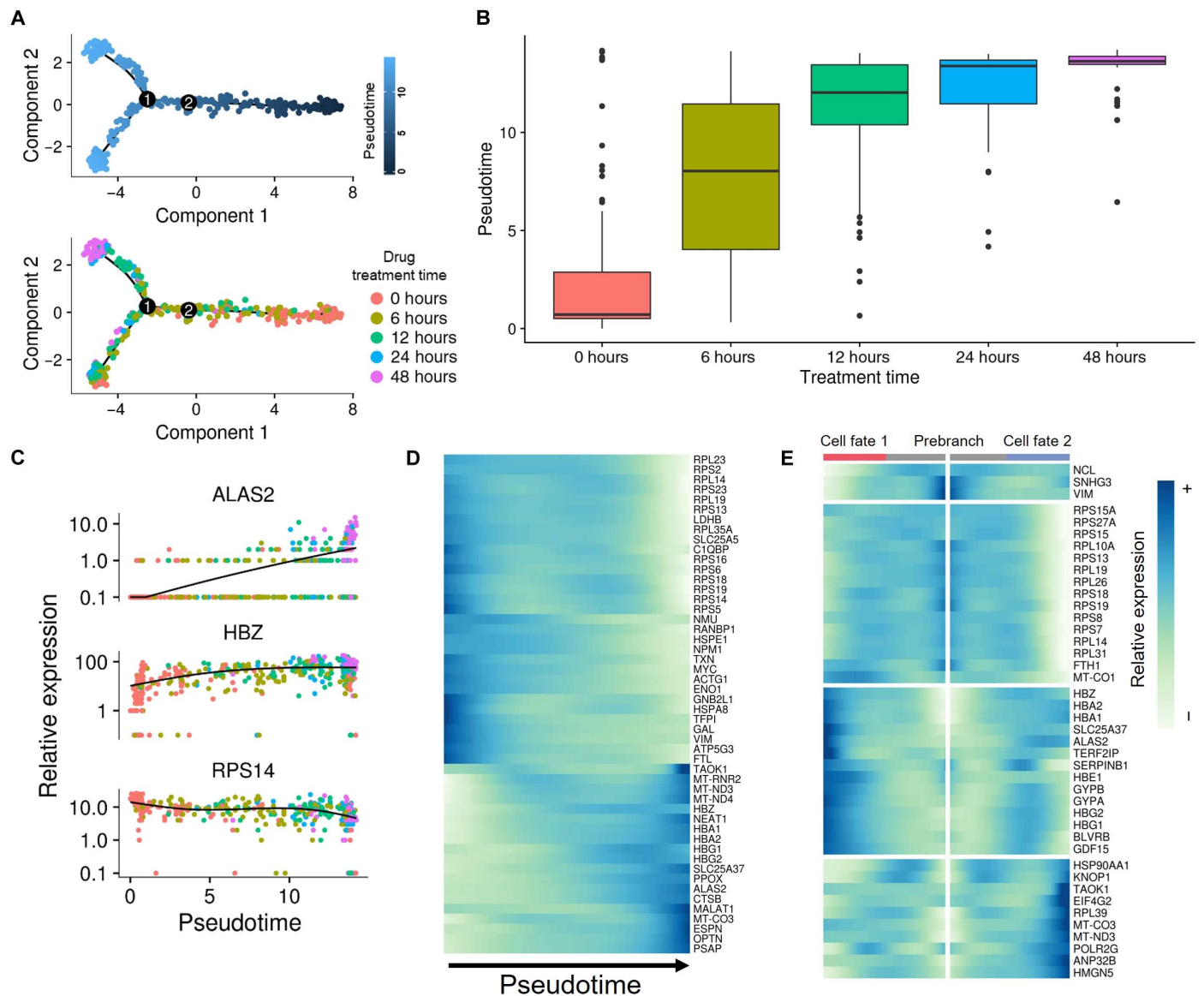
Fig. 2. Pseudotime analysis in 5-plex time-course experiment. (**A**) Monocle pseudotime trajectory of K562 cells treated with imatinib at different time points. Cells are labeled by pseudotime (top) and drug treatment time (bottom). The 0-, 6-, 12-, 24-, and 48-hour samples consist of 133, 109, 79, 49, 58, 52, and 90 cells, respectively. (**B**) Boxplot showing the distribution of pseudotime within each sample. (**C**) Prominent gene expression alterations in 5-plex time-course experiments of imatinib treatment. Note that the cells are labeled by drug treatment time and are not synchronously distributed over pseudotime. (**D**) Expression heatmap showing 50 genes with the lowest *q* values. (**E**) Expression heatmap showing DEGs between two transition states with $q < 1 \times 10^{-4}$. Prebranch refers to the cells before branch 1, Cell fate 1 refers to the cells of upper transition state, and Cell fate 2 refers to the cells in the lower transition state.
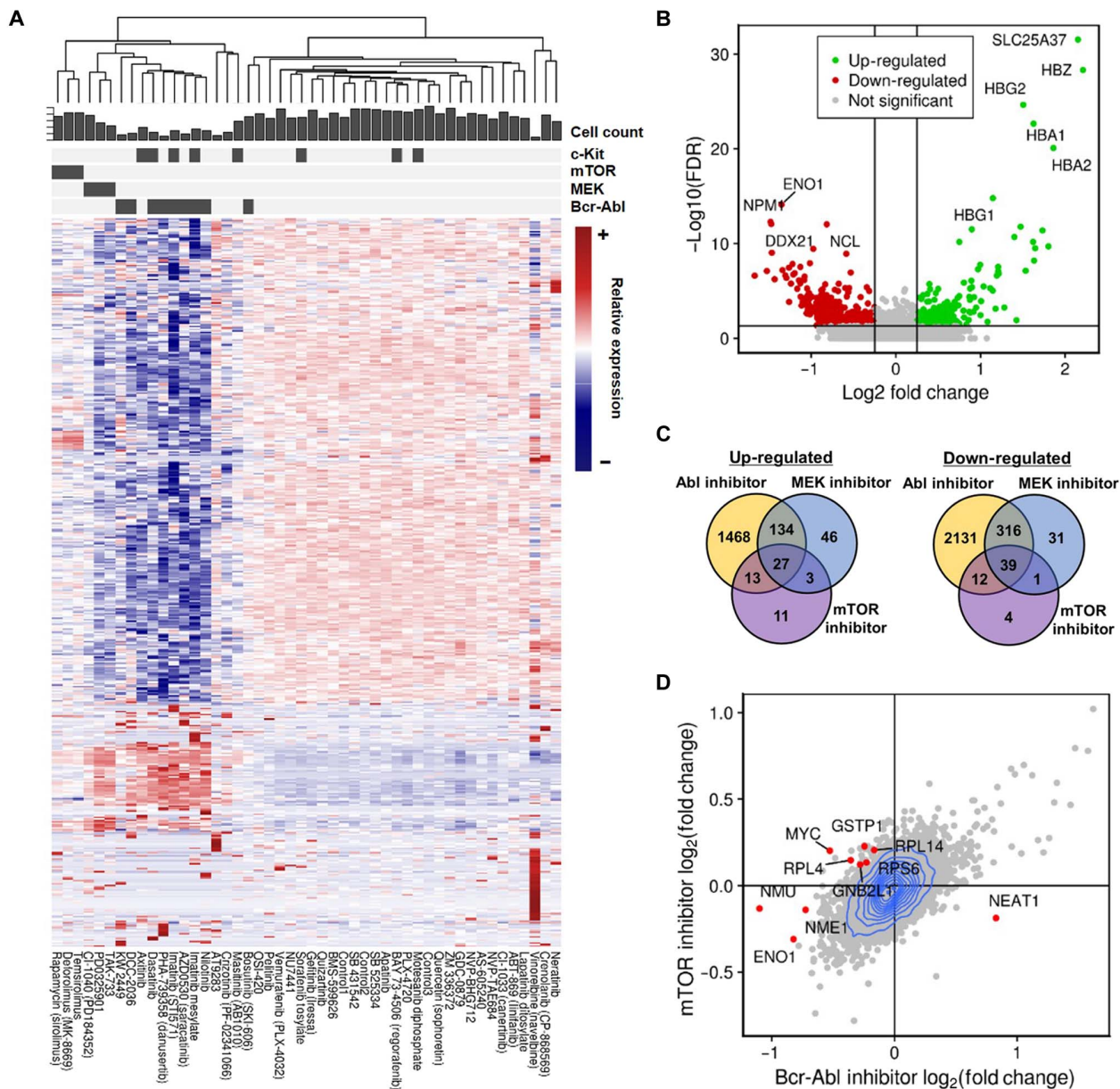
Fig. 3. Gene expression analysis in 48-plex drug treatment experiments. (**A**) Hierarchical clustered heatmap of averaged gene expression profiles for 48-plex drug treatment experiments in K562 cells. Each column represents averaged data in a drug, and each row represents a gene. DEGs were used in this heatmap. The scale bar of relative expression is on the right side. The ability of the drugs to inhibit kinase proteins is shown as binary colors (dark gray indicating positive) at the top. The bar plot at the top shows the cell count for each. (**B**) Volcano plot displaying DEGs of imatinib mesylate compared with DMSO controls. Genes that have a P value smaller than 0.05 and an absolute value of log (fold change) larger than 0.25 are considered significant. Up-regulated genes are colored in green, down-regulated genes are colored in red, and insignificant genes are colored in gray. Ten genes with the lowest P value are labeled. (**C**) Venn diagram showing the relationship between DEGs of three drug groups. Fourteen drugs are classified into three groups according to their protein targets (see Fig. 2C, top), and differential expression analysis is performed by comparing each group with DMSO controls. Relations of both positively (left) and negatively (right) regulated genes in each group are shown. (**D**) Plot showing a correlation between fold changes of expression in cells treated with mTOR inhibitors and BCR-ABL inhibitors compared with DMSO controls.

states existed as a result of imatinib treatment. Samples exhibited asynchronous patterns in pseudotime, although the average increased with drug treatment time (Fig. 2B). We noted that even in the zero-time sample, cells were highly heterogeneous in terms of pseudotime. In addition, we observed an accumulation on the upper transition state as the drug treatment time increased. Differential expression analysis over pseudotime identified several gene cohorts that change

during the transition (Fig. 2, C and D). Notably, the expression levels of erythroid-related genes such as *HBZ* and *ALAS2* had increased over pseudotime (Fig. 2C). This was consistent with previous studies that have shown increased expression of *HBZ* in imatinib-treated cells (*16, 17*). Differentially expressed genes (DEGs) between the two transition states were also identified, and different expression patterns between them were observed (Fig. 2E).

Simultaneous expression profiling of K562 subjected to various drug perturbations

Next, we assessed whether our approach could be used for simultaneous single-cell transcriptome profiling for multiple drugs in K562 cells. We selected 45 drugs, of which most were kinase inhibitors, including several BCR-ABL–targeting drugs. Three dimethyl sulfoxide (DMSO) samples were used as controls (table S1). A 48-plex single-cell experiment was performed by barcoding and pooling all samples after drug treatments. A total of 3091 cells were obtained and demul-

tiplexed after eliminating multiplets and negatives. The averaged expression profiles of each drug were visualized as a heatmap (Fig. 3A). Each drug exhibited its own expression pattern of responsive genes. Unsupervised hierarchical clustering of the averaged expression data for each drug revealed that the response-inducing drugs clustered together by their protein targets, whereas drugs that induced no response showed similar expression patterns with DMSO controls, indicating our method's ability to identify drug targets by expression profiles (Fig. 3A and fig. S4). In addition, we could evaluate cell toxicity
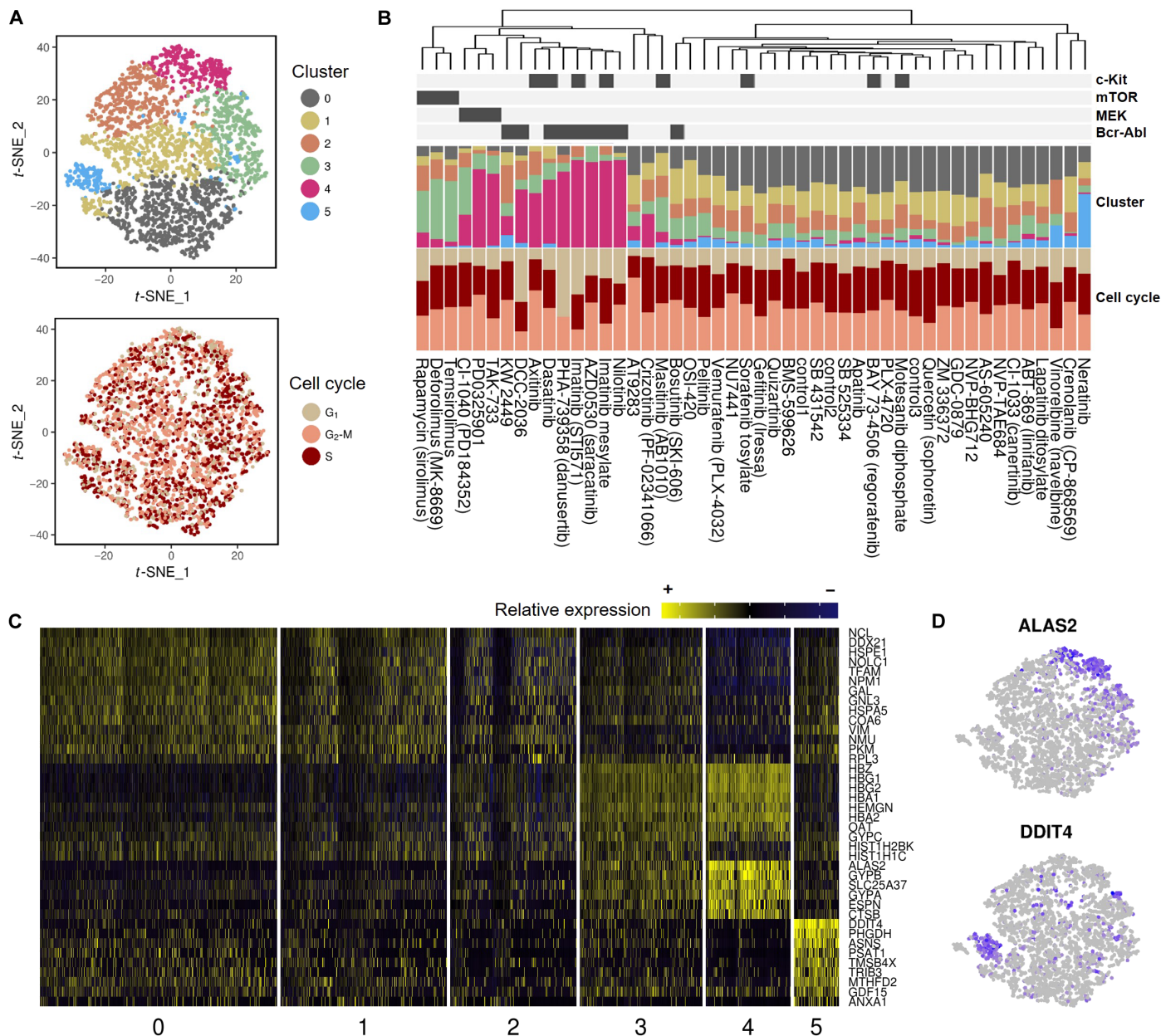
Fig. 4. Single-cell analysis in 48-plex drug treatment experiments. (**A**) The *t*-distributed stochastic neighbor embedding (*t*-SNE) plot of single cells in the 48-plex K562 samples. Plot shows six clusters (top), and additional *t*-SNE plot is labeled by cell cycle states (bottom). (**B**) Bar plots for 48-plex drug treatment experiments in K562 cells. The ability of the drugs to inhibit kinase proteins is shown as binary colors at the top (from Fig. 3A). The bar plot in the middle represents a relative fraction of cells in each *t*-SNE cluster [shown in (A)], and the bottom bar plot displays fractions of cell cycle states for every sample. Drugs are sorted by hierarchical clustering. (**C**) Expression heatmap showing the markers of the clusters. The numbers at the bottom represent cluster numbers. (**D**) Scaled expression of representative genes within the *t*-SNE plot. Intensity of the purple color determines expression levels, with higher intensity correlating with higher gene expression.

by examining the cell counts of each drug. Drugs that targeted BCR-ABL or ABL showed the strongest response and toxicity, and drugs that targeted MAPK kinase (MEK) or mammalian target of rapamycin (mTOR) showed relatively mild response. Differential expression analysis based on the single-cell gene expression data identified DEGs for each drug (Fig. 3B and fig. S5). We note that highly expressed erythroid-related genes such as *HBZ*, *HBA*, and *HBG* were up-regulated, and genes such as *DDX21*, *NCL*, *ENO1*, and *NPM1* were down-regulated in the sample treated with imatinib (Fig. 3B). Similar DEGs were identified for other drugs targeting BCR-ABL. Drugs such as vinorelbine and neratinib showed unique gene expression signatures and DEGs. We next grouped the drugs by their protein targets and performed differential expression analysis. The analysis showed different relationships between DEGs of each protein target (Fig. 3C). In addition, comparative analysis between mTOR inhibitors and BCR-ABL inhibitors revealed that ribosomal protein-coding genes including *RPL4*, *RPS2*, and *RPS3* and regulatory genes such as *MYC* and *GSTP1* are up-regulated in the mTOR inhibitor group (Fig. 3D).

To comprehensively analyze the drug screening data at a single-cell resolution, we performed unsupervised clustering analysis on all the single-cell datasets. We observed six clusters (Fig. 4A), which were not clearly separated possibly due to a highly complex transcriptional space. Nevertheless, for each drug, the relative abundance of cells assigned to each cluster was various (Fig. 4B and fig. S6). Most of the cells affected by BCR-ABL and MEK inhibitors were concentrated in cluster 4, whereas cells affected by mTOR inhibitors were mainly concentrated in cluster 3. Especially, most of the cells in cluster 5 belong to the neratinib-treated sample. Several markers associated with each cluster were verified by differential expression analysis (Fig. 4, C and D). Analysis of cell cycle states revealed no association between cell cycle states and specific clusters (Fig. 4A). The fraction of highly proliferative state (G_2 phase) was decreased in samples treated with BCR-ABL–targeting drugs possibly due to drug-induced cell cycle arrest (Fig. 4B) (18).

To validate the universal applicability of our methods, we performed a 48-plex drug screening experiment on the A375 cell line [*BRAF* V600E positive (19)] with an identical drug set. Similar to K562 cells, response-inducing drugs were clustered together in a target-specific manner in A375 cells (fig. S7). Our results showed that multiplexed scRNA-seq could be used to screen single-cell transcriptional responses to drugs in a high-throughput manner and drug targets could be estimated by their transcriptional patterns.

DISCUSSION
We have developed a novel method for multiplexing samples in scRNA-seq, in which samples were transiently transfected through SBOs containing their own barcodes, pooled, and simultaneously sequenced. This method offers several advantages over currently available scRNA-seq. Our barcoding approach has several advantages in terms of time and cost compared to running multiple individual scRNA-seq experiments. Except for the next-generation sequencing (NGS) cost, we believe that cost constraints occur in the scRNA-seq procedures and NGS preparation steps for each sample. For each sample, the cost of one Drop-Seq run and the corresponding NGS preparation process is approximately $160. In comparison, our SBO transfection method costs approximately $5 (e.g., oligos, transfection reagents, and SBO NGS preparation costs) for each additional sample. If multiple scRNA-seqs are individually processed, then each additional sample could consume an additional cost more than 30 times the barcoding approach of scRNA-seq with multiple samples. This cost saving in library preparation becomes substantial as the size of samples increases. In addition, "batch effect" is one of the major challenges in scRNA-seq (7). These technical noises can be critical and obscure true signals in integrated analysis for multiple samples from different preparations. By pooling and running all samples together, batch effects can be substantially reduced, enabling more precise analysis between single-cell samples.

We demonstrated that our method could also eliminate multiplets and negatives based on SBO count matrix, enabling filtering of true single cells. Identifying expression profiles of true single cells improves data quality and is advantageous for downstream single-cell analysis. In addition, the ability to eliminate multiplets and negatives has potential to increase throughput of scRNA-seq by using a high concentration of cells as an input and filtering single cells subjected to downstream analysis. Throughput of scRNA-seq can be increased beyond the experimental limit by the multiplexed RNA-seq.

Recently, a method for multiplexing samples using genetically natural barcodes has been developed (8). Genetically diverse samples are required in multiplexing by the demuxlet algorithm. Our method offers several advantages over the previous published methods for scRNA-seq (fig. S8). Particularly, our method is capable of multiplexing samples genetically identical but in different experimental conditions, whereas the method using the demuxlet algorithm is not capable of doing. More recently, a method for sample multiplexing using an antibody tagging has been reported (20). However, the method requires expensive reagents and surface markers, limiting the number of samples that can be practically applied.

Our method presented here is very simple and readily applicable to individual laboratories because of the easily accessible reagents and simple experimental process. In addition, because our method is based on liposomal transfection, it has a potential to be applied to nucleus samples. In addition, by using different combinations of SBOs, our method offers a high capacity for multiplexing. We expect that our multiplexing strategy will widely contribute to the adoption of scRNA-seq.

MATERIALS AND METHODS
Cell lines and cell culture
All cell lines were obtained from the Korean Cell Line Bank (KCLB) and maintained at 37°C with 5% CO_2. The human embryonic kidney HEK293T, the mouse embryo fibroblast NIH3T3, and the human malignant melanoma A375 cell lines were cultured in Dulbecco's modified Eagle's medium (DMEM; Gibco, USA) supplemented with 10% fetal bovine serum (FBS; Gibco, USA) and 1% penicillin-streptomycin (Thermo Fisher Scientific, USA). The human chronic myelogenous leukemia K562 cell line and the human colorectal adenocarcinoma SW480 cell line were cultured in RPMI 1640 (Gibco, USA) supplemented with 10% FBS and 1% penicillin-streptomycin.

Barcode design and transfection
The SBO contains a unique 8–base pair (bp) sample barcode, an amplification handle, and a poly-A tail. 5′-TCCAAGGTACAG-ACCTCTGACGNNNNNNNN(A)$_{30}$-3′ is the full SBO sequence. "TCCAAGGTACAGACCTATATCTGACG" is the amplification handle sequence, "NNNNNNNN" is the sample barcode sequence, and (A)$_{30}$ is the poly-A tail sequence. All SBOs were prepared by IDT

(Integrated DNA Technologies, USA) without any modifications. Four hours before Drop-Seq, SBO (28 pmol/ml) was transfected per well using Lipofectamine 3000 (Life Technologies, USA) according to the manufacturer's protocol.

Drop-Seq: NGS preparation of mRNA and SBOs

For each experiment, samples of various conditions were pooled together. The pooled cells were passed through a 40-μm filter and diluted at a final combined concentration of 100 to 400 cells/μl according to Drop-Seq protocol instructions (4). Droplets were generated and processed as previously described. Droplets were collected, and the recovered beads were processed for immediate reverse transcription, followed by exonuclease I treatment. The resulting complementary DNA (cDNA) was divided into appropriate number of tubes, amplified using the KAPA HiFi HotStart PCR Kit (Kapa Biosystems Inc., Switzerland). cDNA amplification was performed in 50 μl of polymerase chain reaction (PCR), which included 4 μl of 10 μM SMART PCR primer, 25 μl of KAPA HiFi DNA polymerase, and up to 21 μl of nuclease-free water. Then, PCR was performed using the following protocol: 3 min at 95°C; four cycles of 20 s at 98°C, 45 s at 65°C, 3 min at 72°C; nine cycles of 20 s at 98°C, 20 s at 67°C, 3 min s at 72°C; 5 min at 72°C. The PCR products were purified twice using 0.6× AMPure (Beckman Coulter, USA) beads according to the manufacturer's instructions. To obtain reverse-transcribed SBOs that are much shorter than cDNA, the first supernatant from AMPure purification step was further purified adding 1.4× homemade AMPure beads [using Sera-Mag SpeedBeads (Thermo Scientific, USA), hereafter Serapure beads (21)]. The cDNA products were fragmented and further amplified using the Nextera XT DNA Library Preparation Kit (Illumina, USA).

The SBO library preparation was performed using a two-step PCR protocol. One nanogram of the SBO cDNA product was loaded into 20 μl of the first adaptor PCR, which included 1 μl of 10 μM forward and reverse primers, 10 μl of KAPA HiFi DNA polymerase, and up to 8 μl of nuclease-free water. PCR was performed using the following protocol: 3 min at 95°C; eight cycles of 20 s at 95°C, 20 s at 64°C, 20 s at 72°C; 5 min at 72°C using the following primers: SMART+AC; P7-SBO hybrid. After 1.8× Serapure bead purification, 8 μl of the first PCR product was loaded into 20 μl of the second index PCR, which included 1 μl of 10 μM forward and reverse primers, and 10 μl of KAPA HiFi DNA polymerase. PCR was performed using the following protocol: 3 min at 95°C; six cycles of 20 s at 95°C, 20 s at 60°C, 20 s at 72°C; 5 min at 72°C using the following primers: New-P5-SMART PCR hybrid; Nextera index oligo. The second PCR product was purified using 1.2× Serapure beads. All primers were prepared by IDT. Sequencing was performed on an Illumina NextSeq 500 system using a NextSeq 500/550 High Output v2 kit (75 cycles) (Illumina, USA). The sequences of primers were provided in table S1. The sequencing depth and number of cells of each experiment are provided in fig. S9.

6-Plex human/mouse species-mixing experiment

HEK293T and NIH3T3 cells were prepared 1 day before Drop-Seq and plated on six-well plates (Techno Plastic Products, Switzerland) at approximately 70% confluency. Transfection of SBO (28 pmol/ml) was performed 4 hours before Drop-Seq, as described above. All cell samples were trypsinized using trypsin-EDTA (0.25%) and phenol red (Gibco, USA), pooled together, and washed four times with phosphate-buffered saline (PBS; Gibco, USA). The cells were then resuspended in 0.01% bovine serum albumin (BSA) + PBS, passed

through a 40-μm filter, counted using the LUNA Automated Cell Counter (Logos Biosystems, Korea), and diluted at a final combined concentration of 400 cells/μl. The diluted sample library was run once in Drop-Seq, and sample preparation and sequencing were performed as above. From one Drop-Seq run, about 77,000 beads were obtained and divided into 24 PCRs for cDNA amplification. Sample preparation was completed using two reactions of the Nextera XT DNA Library Preparation Kit (Illumina, USA).

SBO transfection efficiency and its effect on mixed cultures

To mimic heterogeneous samples, cell lines with different transfection efficiencies were mixed and then SBOs were transfected into the mixed cell line cultures to observe the transfection efficiency and effect on the gene expression profile. HEK293T, NIH3T3, A375, and SW480 cell lines were cultured in DMEM supplemented with 10% FBS and 1% penicillin-streptomycin. All cell lines were prepared 1 day before SBO transfection and plated into six-well plates at approximately 70% confluency with the same number of cells for each cell line. All subsequent steps were the same as described above in the 6-plex human/mouse species-mixing experiment. After the Drop-Seq run, the pooled beads were divided into 24 PCRs for cDNA amplification. Sample preparation was completed using three reactions of the Nextera XT DNA Library Preparation Kit.

We performed the same experiment to examine the effect of SBO transfection on the gene expression profile of K562 cells at the bulk level. These cells were cultured in six-well plates at approximately 30% confluency. After 4 hours of SBO transfection, total RNA of control and transfected K562 cells was extracted using an RNA extraction kit (RNeasy Mini Kit, Qiagen, USA). Total RNA (2 μg) of control and transfected K562 cells was used for cDNA synthesis. Then, sample preparation was completed using the Nextera XT DNA Library Preparation Kit.

5-Plex time-course experiment of drug treatment

K562 cells were plated on six-well plates at approximately 30% confluency. Imatinib (1 μM) was treated to K562 for five time points (0, 6, 12, 24, and 48 hours after treatment). Transfection of SBO (28 pmol/ml) was performed 4 hours before Drop-Seq, as described above, and the cells of each condition were pooled, washed four times with PBS, and resuspended with 0.01% BSA + PBS. After filtering and counting, the pooled cells were diluted at a final combined concentration of 100 cells/μl. The diluted sample library was run once in Drop-Seq. Sample preparation and sequencing were performed as above. From one Drop-Seq run, the pooled beads were divided into 24 PCRs for cDNA amplification. Sample preparation was completed using two reactions of the Nextera XT DNA Library Preparation Kit.

48-Plex drug screening experiment in K562 cells

K562 cells were plated on 24-well plates at approximately 30% confluency and treated with 1 μM of each drug. After 44 hours, transfection of SBO (28 pmol/ml) was performed. Four hours after the transfection, the cell samples from each drug treatment were pooled. The diluted sample library was run three times in Drop-Seq. All subsequent steps were the same as described above in the 5-plex timecourse experiment. After three Drop-Seq runs, the pooled beads were divided into 48 PCRs for cDNA amplification. Sample preparation was completed using three reactions of the Nextera XT DNA Library Preparation Kit.

48-Plex drug screening experiment in A375 cells

A375 cells were prepared 1 day before drug screening and plated on 24-well plates at approximately 30% confluency. All subsequent steps were the same as described above in the 48-plex drug treatment experiment. After three Drop-Seq runs, the pooled beads were divided into 48 PCRs for cDNA amplification. Sample preparation was completed using three reactions of the Nextera XT DNA Library Preparation Kit.

Single-cell transcriptome data processing

For each NextSeq sequencing run, raw sequencing data were converted to FASTQ files using bcl2fastq2 software (Illumina). Each sequencing sample was demultiplexed using Nextera N7xx indices. Tagging, trimming, alignment, and adding annotation tags were performed according to the standard Drop-Seq pipeline (http://mccarrolllab.org/dropseq/). Briefly, reads were first tagged according to the 12-bp cell barcode sequence and the 8-bp unique molecular identifier (UMI) in "read 1." Then, reads in "read 2" were aligned with the hg19 or hg19-mm10 concatenated reference depending on the experiments and collapsed onto 12-bp cell barcodes that corresponded to individual beads. A Hamming distance of 1 was used to collapse UMI within each transcript. Digital expression matrix was obtained by collapsing filtered and mapped reads for each gene by UMI sequence within each cell barcode.

Sample barcode (SBOs) processing

FASTQ files of SBOs were generated as described above. Raw sequencing reads were trimmed to remove PCR handles. Cell barcodes and UMIs were extracted from read 1, and sample barcodes were extracted from read 2. Reads were assigned to 8-mer of sample barcode reference (table S1) with a single-base error tolerance (Hamming distance = 1), and cell barcodes × sample barcodes count matrix (hereinafter referred to as SBO matrix) was generated with consideration to UMI de-duplication. All the processes were made by our homemade python scripts.

Merging sample barcode and transcriptome data

Independently obtained cell barcodes from the two matrices (SBO matrix and transcriptome matrix) were compared and merged on the basis of the cell barcodes from the transcriptome matrix. When merging, a Hamming distance of 1 was allowed. Last, the columns of the SBO and the transcriptome matrix consisted of the same cell barcodes.

Demultiplexing and classification of samples using SBO matrix

SBO matrix was normalized using a modified version of centered log ratio (CLR) transformation (22)

$$x_i' = \ln(x_i + 1) - \frac{1}{D} \sum_{j=1}^{D} \ln(x_i + 1)$$

x_i' denotes the normalized count for a specific SBO in cell i, x_i denotes the raw count, and D is the total cell number. In CLR transformation, the raw counts of SBO are divided by the geometric mean of individual SBO across cells and are log-transformed. We added the raw counts of SBO to 1 to avoid infinite values. We hypothesized that we can discriminate positive signals from negative (background) signals by fitting the distribution of negative signals of each SBO and thresholding the nor-

malized counts to a specific value of each SBO. Following the normalization, for each SBO, we excluded the cells with the highest expression of the SBO among all SBOs. We fitted a negative binomial distribution to the remaining cells to obtain a distribution of negative signals. Next, we calculated a quantile with 0.99 probability to get the threshold value of each SBO. Cells that have higher SBO counts than the threshold value were considered as positives for that SBO. Cells were demultiplexed and classified into singlets, multiplets, and negatives based on the above results.

6-Plex human-mouse species-mixing experiment analysis

Transcriptome and SBO data processing were performed as described above. We obtained 2759 of cell barcodes after filtering out cells with less than 500 transcripts. After the SBO matrix normalization and classification, we classified singlets as positive for one of SBOs, multiplets as positive for more than one SBO, and negatives as positive for none of SBOs. For species-mixing plots in Fig. 1D and fig. S4, only singlets and doublets of the two specified SBOs were used.

Pseudotime analysis

For pseudotime analysis of 5-plex time-course experiment, we applied the R package "Monocle 2" (23). After removing multiplets and negatives, samples were demultiplexed, quality-controlled, and analyzed. A single-cell trajectory was constructed by Discriminative Dimensionality Reduction with Trees (DDRTree) (24) algorithm using genes differentially expressed at different time points. Cells were ordered across the trajectory by setting the state containing 0-hour sample as a time zero, and pseudotime was calculated. To identify DEGs over pseudotime, a likelihood ratio test in the negative binomial model was performed and genes with a q value less than 0.01 were selected as DEGs. When drawing the heatmap, genes were clustered by their pseudotime expression patterns. Differential expression analysis between two transition states in branch 1 was performed using BEAM function in Monocle package.

48-Plex drug screening data analysis

Following the alignment of sequencing reads, downstream analysis of the 48-plex drug screening experiment was performed using the R package "Seurat" (25). After demultiplexing and removing multiplets and negatives, cells were quality-controlled on the basis of the mitochondrial reads fraction, number of UMI, and number of genes. We identified 3091 cells in which at least 500 transcripts and 300 genes were detected. RNA expression matrix was log-normalized and processed for the further analysis. To cluster the single cells, we ran principal components analysis (PCA) using the expression matrix of variable genes and then performed t-distributed stochastic neighbor embedding (t-SNE) using the first six PCA components. We identified six clusters using FindClusters function in Seurat with resolution = 0.6. We assigned cell cycle phase scores using cell cycle markers (26) and classified each cell to G_2-M, S, or G_1 phase. To draw a hierarchical clustered heatmap, we first identified DEGs for each drug with adjusted $P < 0.05$ by Wilcoxon rank-sum test and obtained 469 responsive genes by merging the DEGs altogether. Expression levels of each drug for the responsive genes were normalized, averaged, and scaled and were used for drawing the heatmap. To construct the dendrogram at the heatmap, hierarchical clustering was performed on the basis of correlations among the expression levels across drugs. Normalized and scaled gene expression data were used in the heatmap. To identify DEGs in Fig. 2D and fig. S7, we performed likelihood ratio test between single cells in each drug

and single cells in DMSO controls. To analyze samples by their protein targets, 14 drugs were classified into three groups (BCR-ABL inhibitors, MEK inhibitors, and mTOR inhibitors). Differential expression analysis between cells in each group and cells in DMSO controls was performed as described above. The analysis of drug screening experiment of A375 was performed in the same manner as K562.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/5/5/eaav2249/DC1

REFERENCES AND NOTES

1. A. K. Shalek, R. Satija, X. Adiconis, R. S. Gertner, J. T. Gaublomme, R. Raychowdhury, S. Schwartz, N. Yosef, C. Malboeuf, D. Lu, J. J. Trombetta, D. Gennert, A. Gnirke, A. Goren, N. Hacohen, J. Z. Levin, H. Park, A. Regev, Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236–240 (2013).
2. F. Tang, K. Lao, M. A. Surani, Development and applications of single-cell transcriptome analysis. *Nat. Methods* **8**, S6–S11 (2011).
3. Q. Deng, D. Ramsköld, B. Reinius, R. Sandberg, Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193–196 (2014).
4. E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev, S. A. McCarroll, Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
5. A. M. Klein, L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D. A. Weitz, M. W. Kirschner, Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
6. G. X. Y. Zheng, J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, M. T. Gregory, J. Shuga, L. Montesclaros, J. G. Underwood, D. A. Masquelier, S. Y. Nishimura, M. Schnall-Levin, P. W. Wyatt, C. M. Hindson, R. Bharadwaj, A. Wong, K. D. Ness, L. W. Beppu, H. J. Deeg, C. McFarland, K. R. Loeb, W. J. Valente, N. G. Ericson, E. A. Stevens, J. P. Radich, T. S. Mikkelsen, B. J. Hindson, J. H. Bielas, Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
7. S. C. Hicks, F. W. Townes, M. Teng, R. A. Irizarry, Missing data and technical variability in single-cell RNA-sequencing experiments. *Biostatistics* **19**, 562–578 (2018).
8. H. M. Kang, M. Subramaniam, S. Targ, M. Nguyen, L. Maliskova, E. McCarthy, E. Wan, S. Wong, L. Byrnes, C. M. Lanata, R. E. Gate, S. Mostafavi, A. Marson, N. Zaitlen, L. A. Criswell, C. J. Ye, Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* **36**, 89–94 (2018).
9. T. R. Hughes, M. J. Marton, A. R. Jones, C. J. Roberts, R. Stoughton, C. D. Armour, H. A. Bennett, E. Coffey, H. Dai, Y. D. He, M. J. Kidd, A. M. King, M. R. Meyer, D. Slade, P. Y. Lum, S. B. Stepaniants, D. D. Shoemaker, D. Gachotte, K. Chakraburtty, J. Simon, M. Bard, S. H. Friend, Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 (2000).
10. F. A. Middleton, K. Mirnics, J. N. Pierri, D. A. Lewis, P. Levitt, Gene expression profiling reveals alterations of specific metabolic pathways in schizophrenia. *J. Neurosci.* **22**, 2718–2729 (2002).
11. Y. Huang, A. de Reyniès, L. de Leval, B. Ghazi, N. Martin-Garcia, M. Travert, J. Bosq, J. Brière, B. Petit, E. Thomas, P. Coppo, T. Marafioti, J.-F. Emile, M.-H. Delfau-Larue, C. Schmitt, P. Gaulard, Gene expression profiling identifies emerging oncogenic pathways operating in extranodal NK/T-cell lymphoma, nasal type. *Blood* **115**, 1226–1237 (2010).
12. J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J.-P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, T. R. Golub, The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929–1935 (2006).
13. A. Subramanian, R. Narayan, S. M. Corsello, D. D. Peck, T. E. Natoli, X. Lu, J. Gould, J. F. Davis, A. A. Tubelli, J. K. Asiedu, D. L. Lahr, J. E. Hirschman, Z. Liu, M. Donahue, B. Julian, M. Khan, D. Wadden, I. C. Smith, D. Lam, A. Liberzon, C. Toder, M. Bagul, M. Orzechowski, O. M. Enache, F. Piccioni, S. A. Johnson, N. J. Lyons, A. H. Berger, A. F. Shamji, A. N. Brooks, A. Vrcic, C. Flynn, J. Rosains, D. Y. Takeda, R. Hu, D. Davison, J. Lamb, K. Ardlie, L. Hogstrom, P. Greenside, N. S. Gray, P. A. Clemons, S. Silver, X. Wu, W.-N. Zhao, W. Read-Button, X. Wu, S. J. Haggarty, L. V. Ronco, J. S. Boehm, S. L. Schreiber, J. G. Doench, J. A. Bittker, D. E. Root, B. Wong, T. R. Golub, A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* **171**, 1437–1452.e17 (2017).
14. J. V. Melo, The diversity of BCR-ABL fusion proteins and their relationship to leukemia phenotype. *Blood* **88**, 2375–2384 (1996).
15. F. X. Mahon, M. W. N. Deininger, B. Schultheis, J. Chabrol, J. Reiffers, J. M. Goldman, J. V. Melo, Selection and characterization of *BCR-ABL* positive cell lines with differential sensitivity to the tyrosine kinase inhibitor STI571: Diverse mechanisms of resistance. *Blood* **96**, 1070–1079 (2000).
16. A.-S. Espadinha, V. Prouzet-Mauléon, S. Claverol, V. Lagarde, M. Bonneu, F.-X. Mahon, B. Cardinaud, A tyrosine kinase-STAT5-miR21-PDCD4 regulatory axis in chronic and acute myeloid leukemia cells. *Oncotarget* **8**, 76174–76188 (2017).
17. L. Xiong, J. Zhang, B. Yuan, X. Dong, X. Jiang, Y. Wang, Global proteome quantification for discovering imatinib-induced perturbation of multiple biological pathways in K562 human chronic myeloid leukemia cells. *J. Proteome Res.* **9**, 6007–6015 (2010).
18. C. Gambacorti-Passerini, P. le Coutre, L. Mologni, M. Fanelli, C. Bertazzoli, E. Marchesi, M. Di Nicola, A. Biondi, G. M. Corneo, D. Belotti, E. Pogliani, N. B. Lydon, Inhibition of the ABL kinase activity blocks the proliferation of BCR/ABL+ leukemic cells and induces apoptosis. *Blood Cells Mol. Dis.* **23**, 380–394 (1997).
19. H. Davies, G. R. Bignell, C. Cox, P. Stephens, S. Edkins, S. Clegg, J. Teague, H. Woffendin, M. J. Garnett, W. Bottomley, N. Davis, E. Dicks, R. Ewing, Y. Floyd, K. Gray, S. Hall, R. Hawes, J. Hughes, V. Kosmidou, A. Menzies, C. Mould, A. Parker, C. Stevens, S. Watt, S. Hooper, R. Wilson, H. Jayatilake, B. A. Gusterson, C. Cooper, J. Shipley, D. Hargrave, K. Pritchard-Jones, N. Maitland, G. Chenevix-Trench, G. J. Riggins, D. D. Bigner, G. Palmieri, A. Cossu, A. Flanagan, A. Nicholson, J. W. C. Ho, S. Y. Leung, S. T. Yuen, B. L. Weber, H. F. Seigler, T. L. Darrow, H. Paterson, R. Marais, C. J. Marshall, R. Wooster, M. R. Stratton, P. A. Futreal, Mutations of the *BRAF* gene in human cancer. *Nature* **417**, 949–954 (2002).
20. M. Stoeckius, S. Zheng, B. Houck-Loomis, S. Hao, B. Z. Yeung, W. M. Mauck III, P. Smibert, R. Satija, Cell hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 224 (2018).
21. N. Rohland, D. Reich, Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.* **22**, 939–946 (2012).
22. J. Aitchison, Measures of location of compositional data sets. *Math. Geol.* **21**, 787–790 (1989).
23. C. Trapnell, D. Cacchiarelli, J. Grimsby, P. Pokharel, S. Li, M. Morse, N. J. Lennon, K. J. Livak, T. S. Mikkelsen, J. L. Rinn, The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
24. Q. Mao, L. Wang, S. Goodison, Y. Sun, paper presented at the Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia, 10 to 13 August 2015.
25. R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, A. Regev, Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
26. I. Tirosh, B. Izar, S. M. Prakadan, M. H. Wadsworth II, D. Treacy, J. J. Trombetta, A. Rotem, C. Rodman, C. Lian, G. Murphy, M. Fallahi-Sichani, K. Dutton-Regester, J.-R. Lin, O. Cohen, P. Shah, D. Lu, A. S. Genshaft, T. K. Hughes, C. G. K. Ziegler, S. W. Kazer, A. Gaillard, K. E. Kolb, A.-C. Villani, C. M. Johannessen, A. Y. Andreev, E. M. Van Allen, M. Bertagnolli, P. K. Sorger, R. J. Sullivan, K. T. Flaherty, D. T. Frederick, J. Jané-Valbuena, C. H. Yoon, O. Rozenblatt-Rosen, A. K. Shalek, A. Regev, L. A. Garraway, Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).