Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/5/5/eaav2249/DC1

Supplementary Materials for

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

Dongju Shin, Wookjae Lee, Ji Hyun Lee*, Duhee Bang*

*Corresponding author. Email: duheebang@yonsei.ac.kr (D.B.); hyunihyuni@khu.ac.kr (J.H.L.)

Published 15 May 2019, *Sci. Adv.* **5**, eaav2249 (2019)
DOI: 10.1126/sciadv.aav2249

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/5/5/eaav2249/DC1)

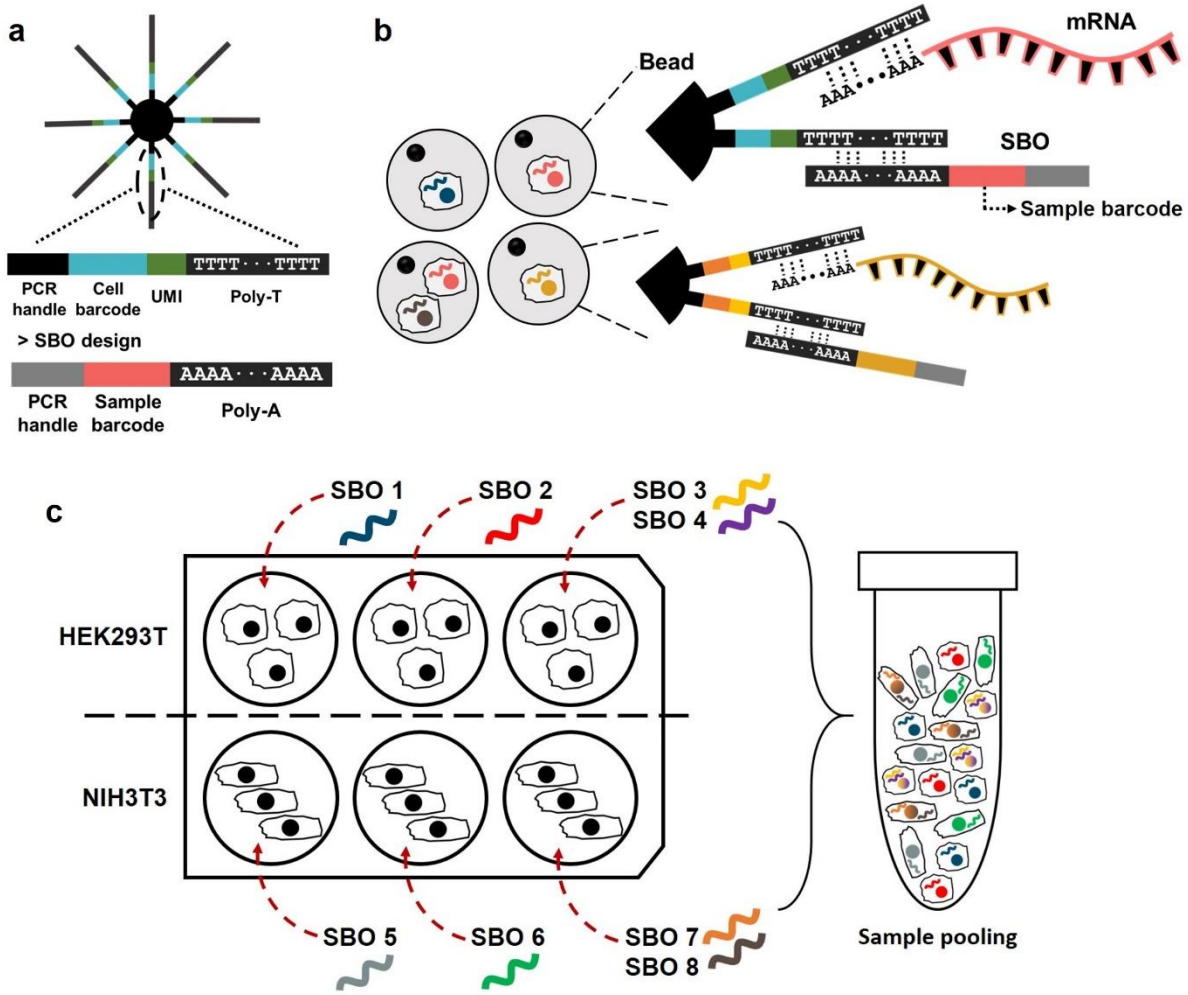

Fig. S1. Schematic figure of transient transfection using SBO. (a) Structure of Drop-Seq beads (top) and SBO (bottom). Drop-Seq beads contain the PCR handle (25 mer), cell barcode (12 mer), UMI (8 mer), and poly-T (30 mer). SBO contains the PCR handle (22 mer), sample barcode (8 mer), and poly-A (30 mer). **(b)** SBO and mRNA capture within droplets. After cell lysis, released mRNA and SBO are simultaneously captured by Drop-Seq beads by poly-A tail hybridization to poly-T sequence of beads. **(c)** Scheme of 6-plex human-mouse species-mixing experiment. Six cell samples (3 HEK293T samples and 3 NIH3T3 samples) are barcoded before sample pooling. Two samples of HEK293T and two samples of NIH3T3 are transfected with a single SBO. One sample each of HEK293T and NIH3T3 cells is transfected with a combination of two SBOs.

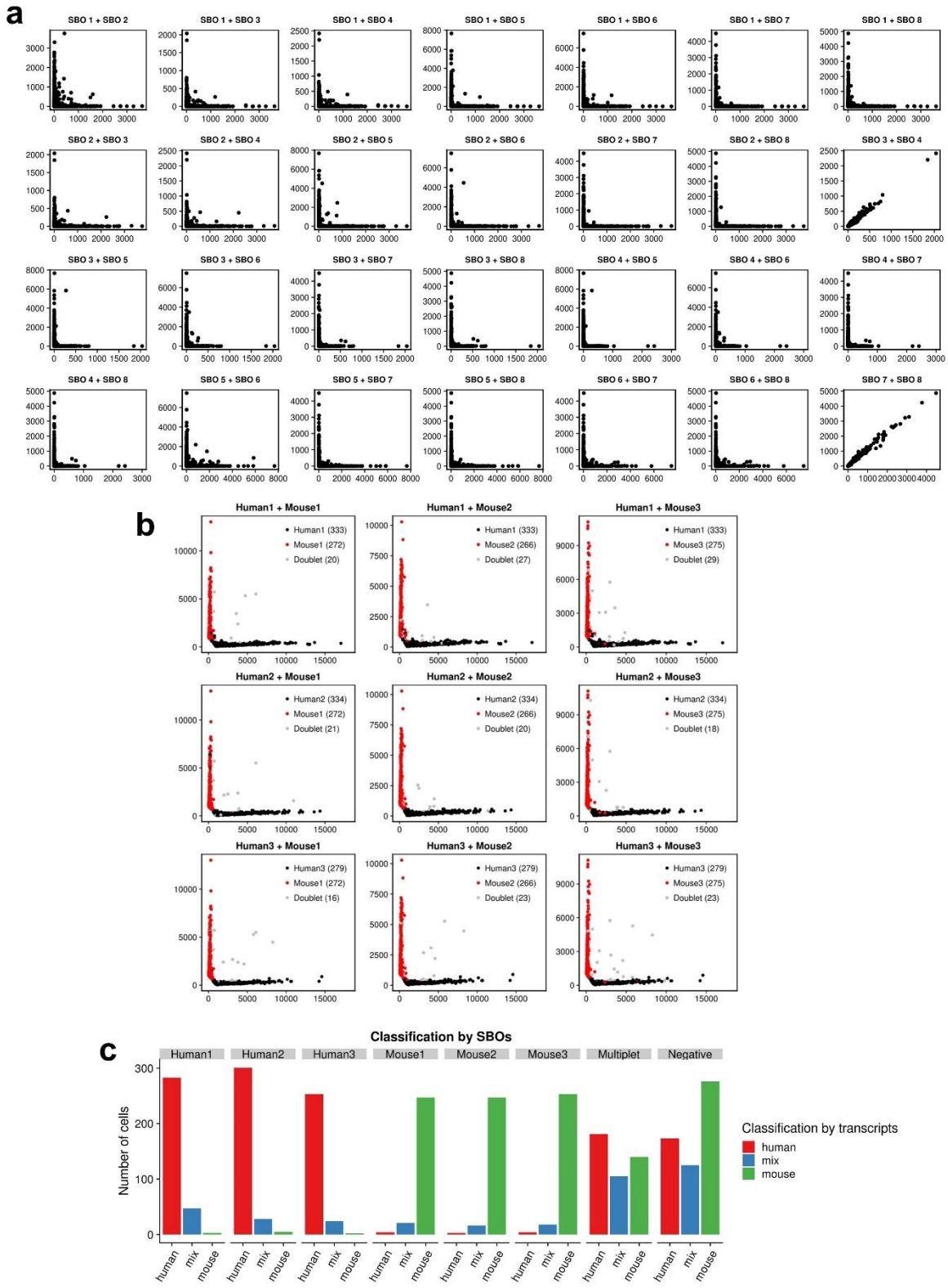

Fig. S2. Additional data for the species-mixing experiment. (**a**) Scatter plots of raw counts between two SBOs. Scatter plots of all possible combinations of two SBOs are shown. The x-axis denotes the former part of the subtitle. The y-axis denotes the latter part of the subtitle. Note that SBOs in samples (3+4 and 7+8) show a strong correlation in their counts, while the other combinations of SBOs show an exclusive relationship. (**b**) Species-mixing plot labeled by SBO classification. Cells are labeled according to their SBO classification. The x-axis denotes the number of human transcripts and the y-

axis denotes the number of mouse transcripts. Black dots indicate SBO-classified human samples, red dots indicate SBO-classified mouse samples, and grey dots indicate SBO-classified doublets that are positive for both human and mouse SBOs. (**c**) Bar plot showing the cell numbers of transcript-classified species within SBO-classified species.

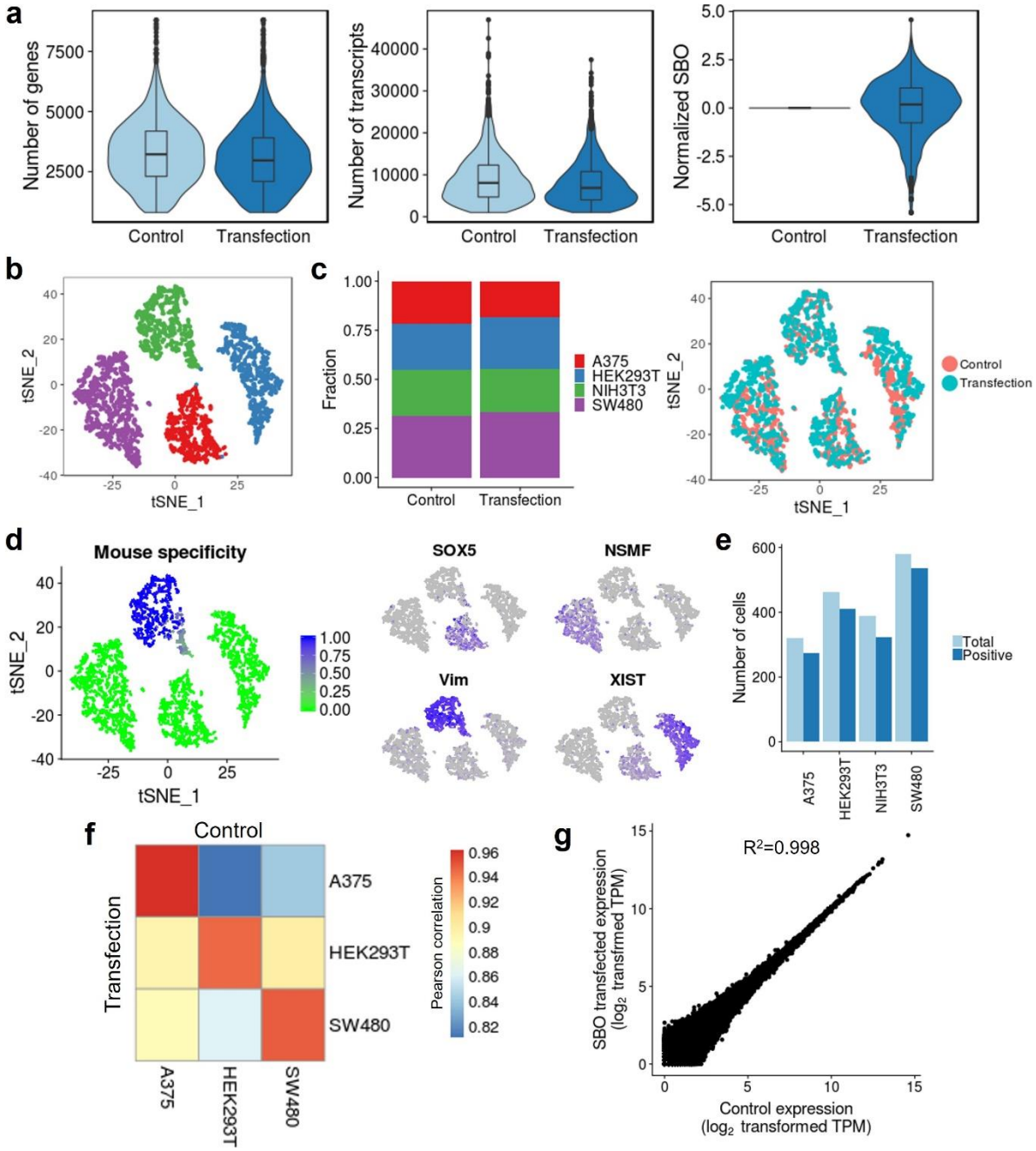

Fig. S3. SBO barcoding in heterogeneous cell samples and the effect of transient transfection. Heterogeneous samples derived from A375, HEK293T, NIH3T3, and SW480 cell lines were barcoded by transient transfection and compared against non-transfected (control) samples. Cell lines were mixed before barcoding. (**a**) Distribution of genes, transcripts, and SBOs in both control and transfection samples. (**b**) tSNE visualization of single-cells colored by cluster identity. Each cluster was annotated to the cell line by cell line-specific markers. (**c**) Cluster occupancy of each sample. Control and transfection samples successfully clustered together without bias. (**d**) Expression of cell line-specific markers and species specificity (for the NIH3T3 mouse cell line), indicating successful cell line annotation. (**e**) Transfection efficiency of all cell lines was 80 - 90%. The number of cells positive for the SBO barcode and the total number of cells in each sample are shown. (**f**) Gene expression correlation of human clusters in control and transfected samples. The scale bar indicates

the Pearson correlation coefficient. (**g**) Gene expression correlation of transfected and control samples verified in bulk RNA-seq. Bulk RNA-seq was performed using K562 cell line.

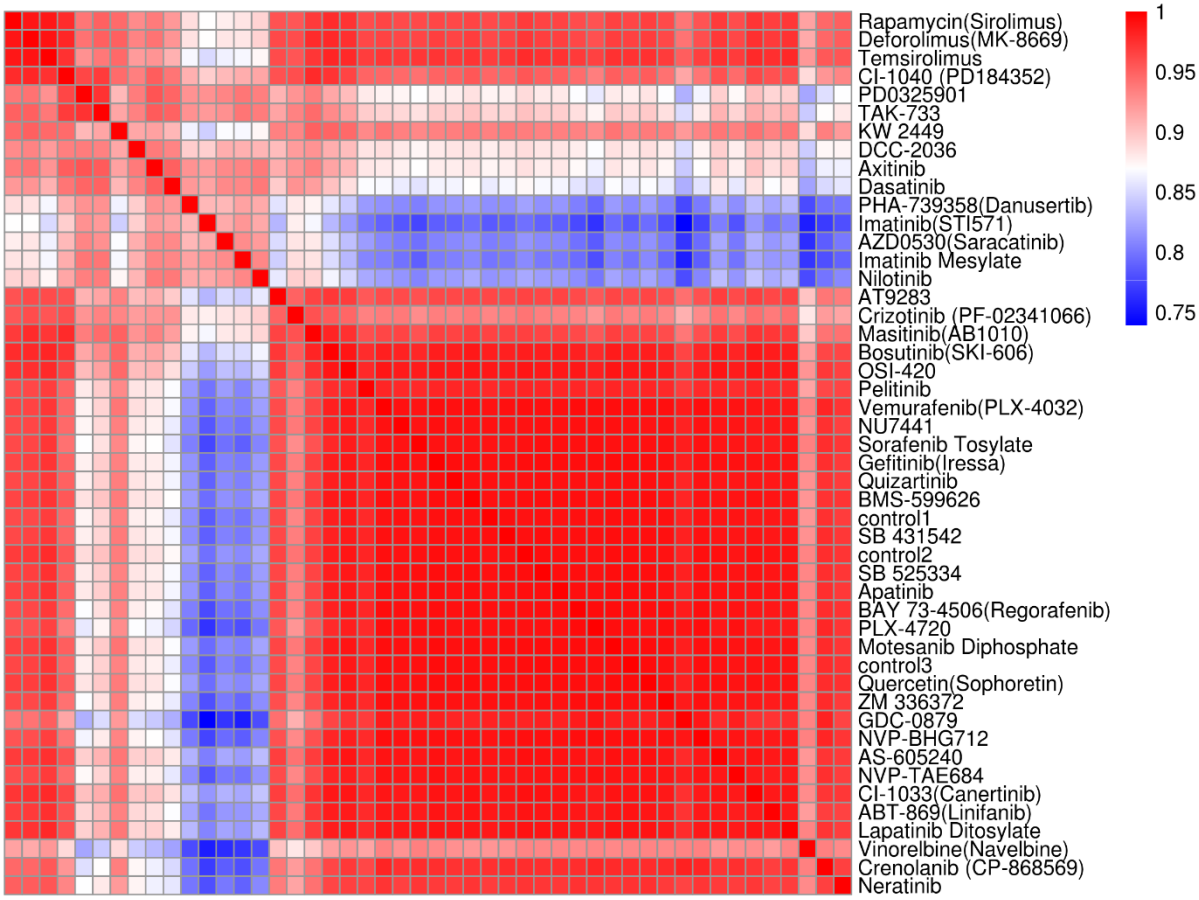

Fig. S4. Correlation heatmap of average gene expression across the drugs. Raw expressions of single-cells with each drug were averaged using genes used in Fig. 2c. Pearson correlation coefficients across 48 samples are calculated and expressed as a heatmap. Note that drugs that induced no response are highly correlated and response-inducing drugs correlate with their protein targets.

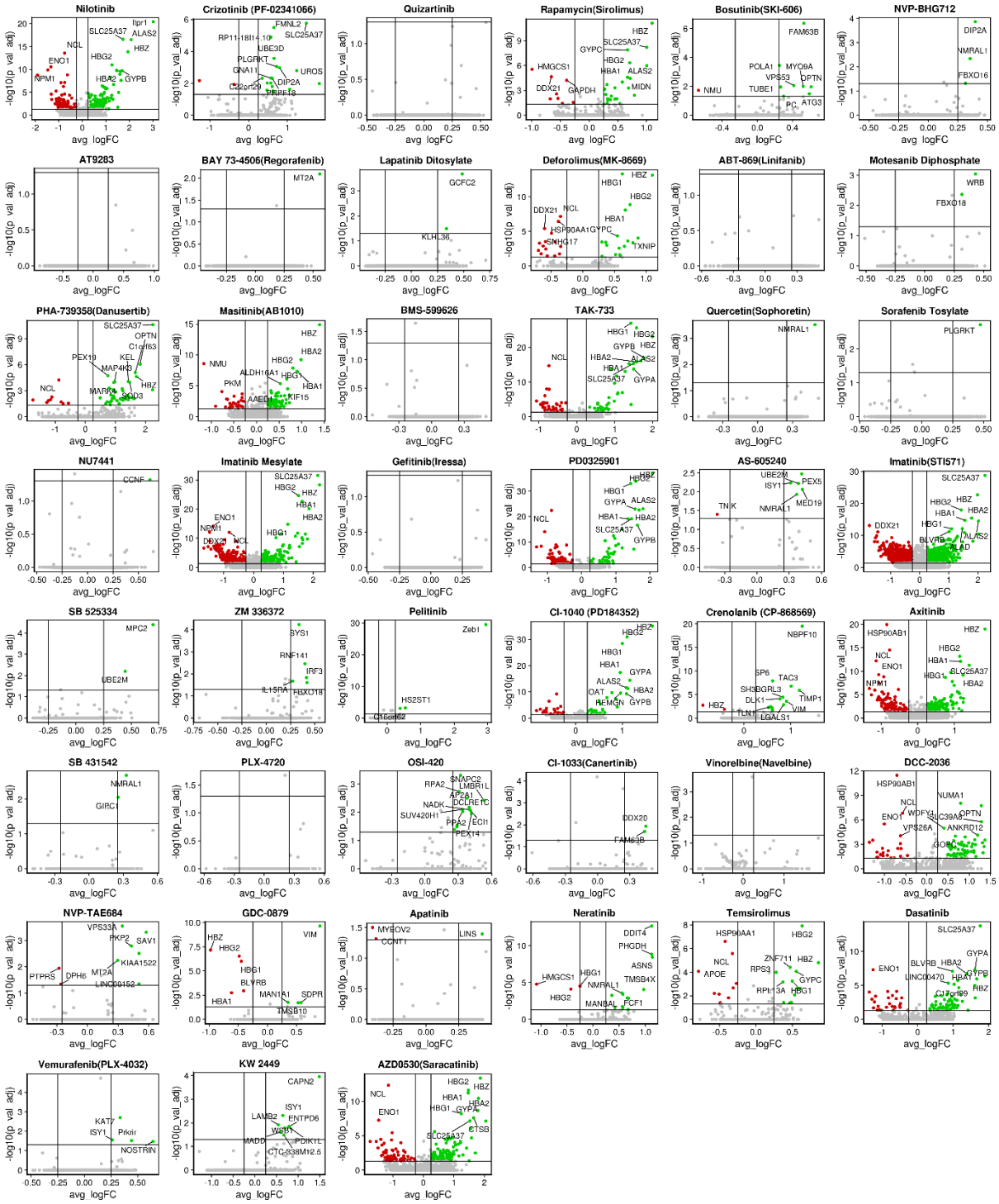

Fig. S5. Volcano plots showing DEGs for each drug. Volcano plots of 45 drugs showing differentially expressed genes. Genes that have a p-value smaller than 0.05 and an absolute value of log (fold change) larger than 0.25 are considered significant. Up-regulated genes are colored in green, down-regulated genes are colored in red. Insignificant genes are colored in grey. Ten genes with the lowest p-value are labeled.

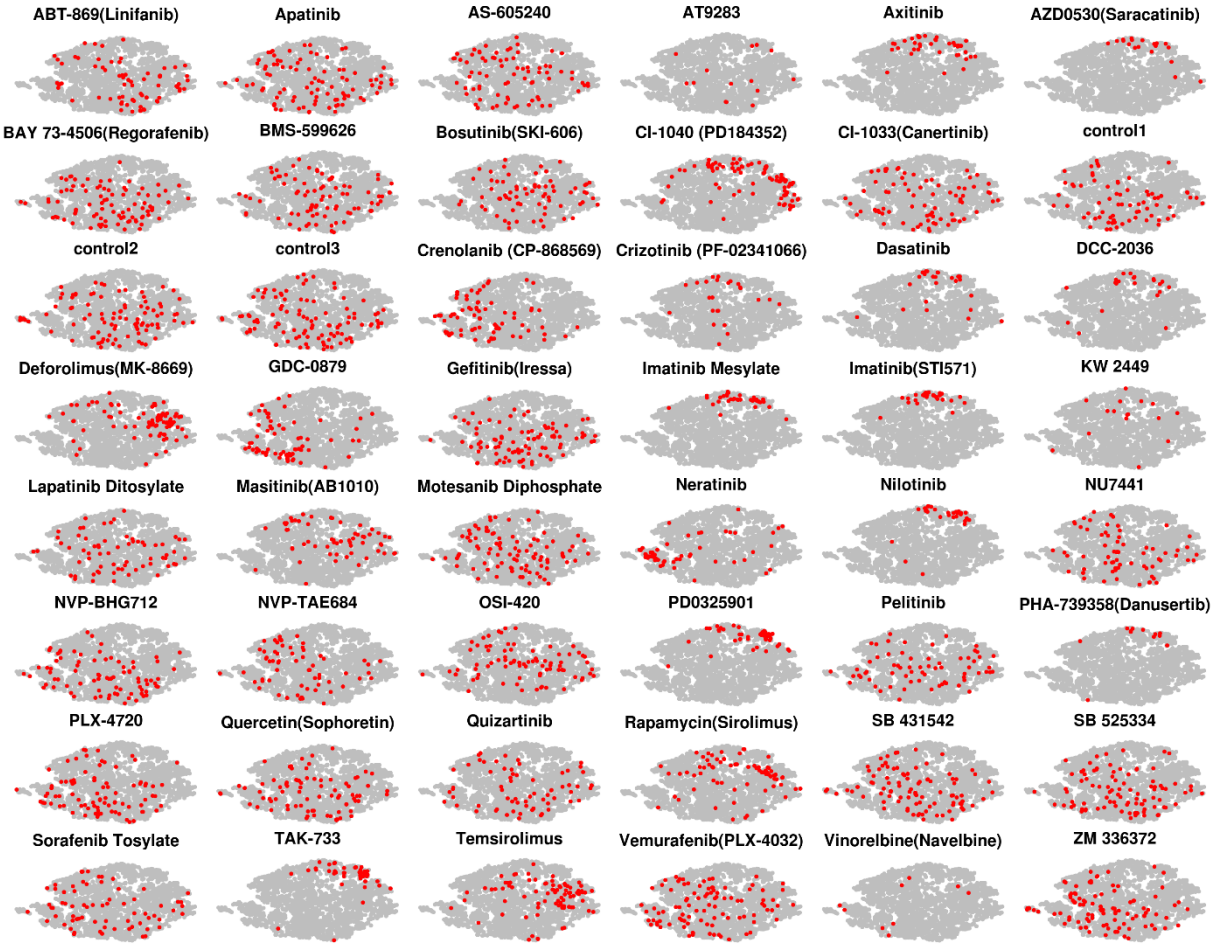

Fig. S6. Cell cycle analysis and *t*-SNE at a single-cell resolution. tSNE plots showing the distributions of samples in the plot.

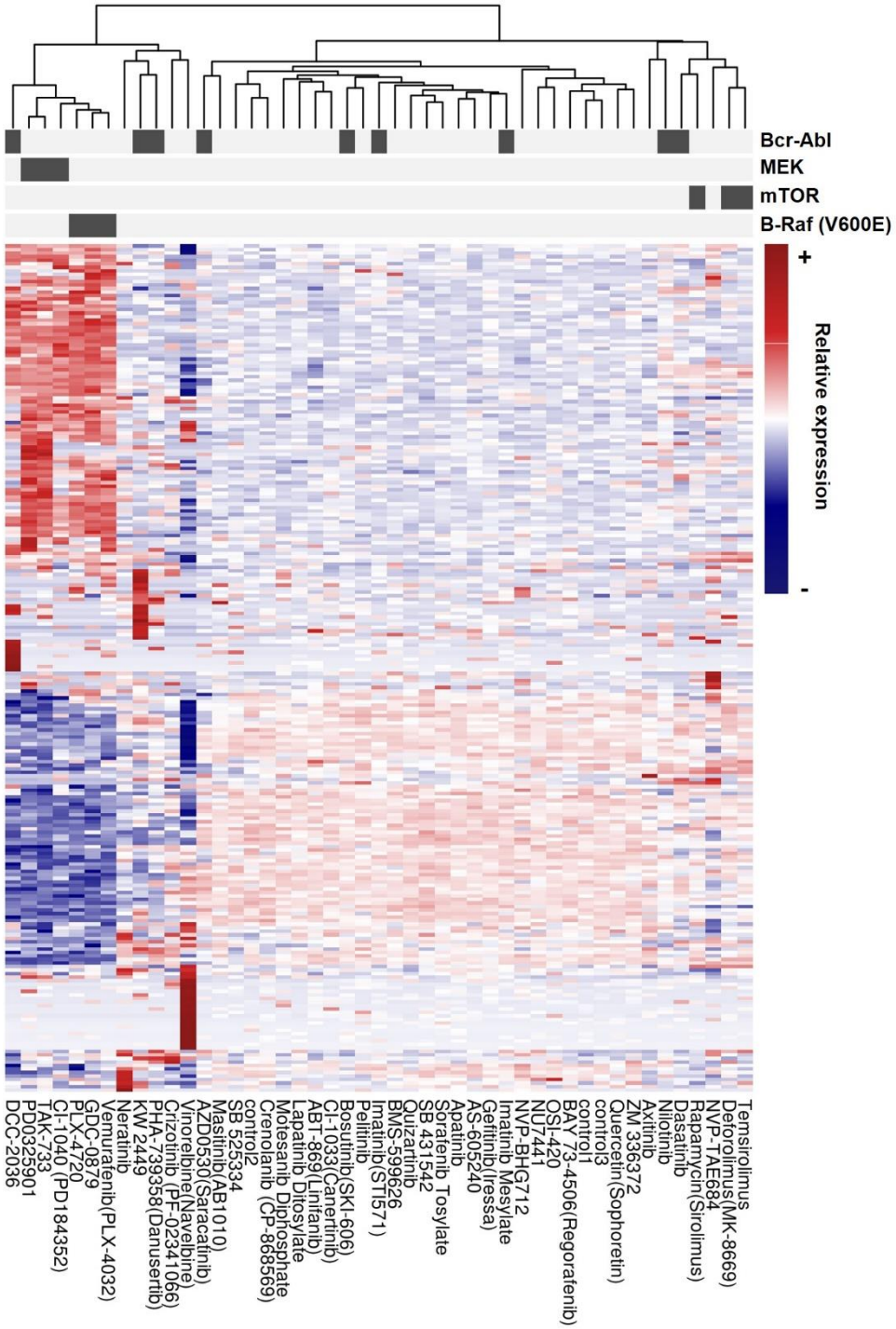

Fig. S7. Expression heatmap for A375 drug screening experiment. Hierarchical clustered heatmap of averaged gene expression profiles for 48-plex drug treatment experiments in A375 cells. Each column represents averaged data in a drug and each row represents a gene. Differential expressed genes were used 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 grey indicating positive) at the top.

	Previous scRNA-seq	Demuxlet†	Cell hashing‡	Our method
scRNA-seq method	Drop-seq, inDrop, 10× Chromium, etc	10× Chromium	10× Chromium	Drop-seq
Sample multiplexing	Unable	Able	Able	Able
Multiplexing method	-	Computational algorithm	Additional experimental method	Additional experimental method
Need for multiplexing	-	Information about single-nucleotide polymorphisms (SNPs) of individuals	Antibody-Oligo conjugation, Surface markers	Short barcoding oligo transfection
Sample barcode	-	Endogenous SNPs	Exogenous synthetic barcode	Exogenous synthetic barcode
Available sample types	Any sample	Samples with distinct genetic backgrounds	Samples with known surface markers	Any sample
Cell over-loading	Unable	Able	Able	Able
Multiple sample preparation cost	High	Low	Low	Low
Batch effect	Sample specific	Minimized	Minimized	Minimized

Fig. S8. Comparison between previous methods and our method.

Experiment	Species mixing	Time-course	Drug screening (K562)
Singlet	1,787	428	3,091
Multiplet	388	52	487
Negative	584	90	1,322
Total	2,759	570	4,900
Sequencing Depth (cDNA)	~30,000 reads per cell	~130,000 reads per cell	~50,000 reads per cell
Sequencing Depth (SBO)	~20,000 reads per cell	~5,000 reads per cell	~5,000 reads per cell

Fig. S9. Cell numbers and sequencing depth for each experiment.