# Supplementary Material

Highly Multiplexed Single-Cell RNA-seq for Defining Cell Population and Transcriptional Spaces

Jase Gehring[1,2], Jong Hwee Park[2], Sisi Chen[2], Matthew Thomson[2], Lior Pachter[2,3*]

1. Department of Molecular and Cell Biology, University of California, Berkeley

2. Division of Biology and Biological Engineering, California Institute of Technology

3. Department of Computing and Mathematical Sciences, California Institute of Technology

* Address correspondence to Lior Pachter (lpachter@caltech.edu)
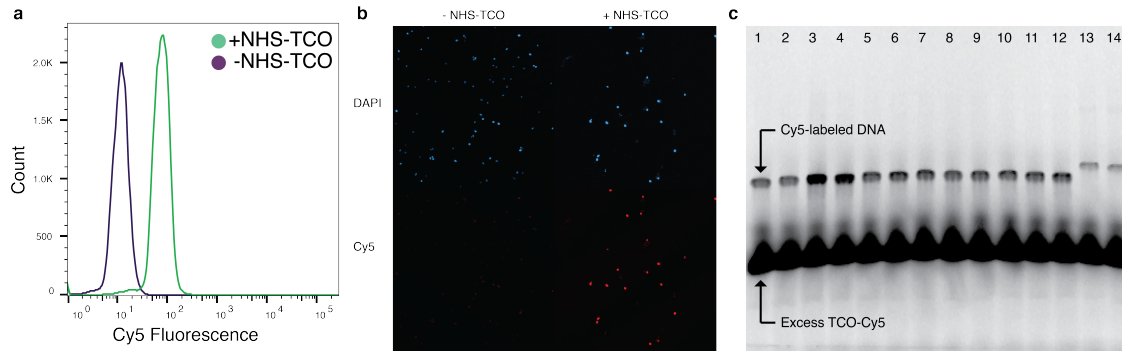
Supplementary Figure 1: Direct cell labeling with Inverse Electron-Demand Diels-Alder (IEDDA) chemistry (**a**) Yeast cells were fluorescently labeled in a one-pot, two-step reaction with NHS-TCO and MTZ-Cy5. Control reactions omitted NHS-TCO. (**b**) Fluorescence microscopy of yeast cells labeled with NHS-TCO and MTZ-Cy5 shows labeling only in the presence of NHS-TCO cross-linker. (**c**) Activity assay for panels of methyltetrazine-activated DNA sample tags. MTZ-DNAs were reacted with TCO-Cy5 and the products separated by polyacrylamide gel electrophoresis. Lanes 1-12 are 3'-amine modified, while lanes 13 and 14 are 5'-amine modified.
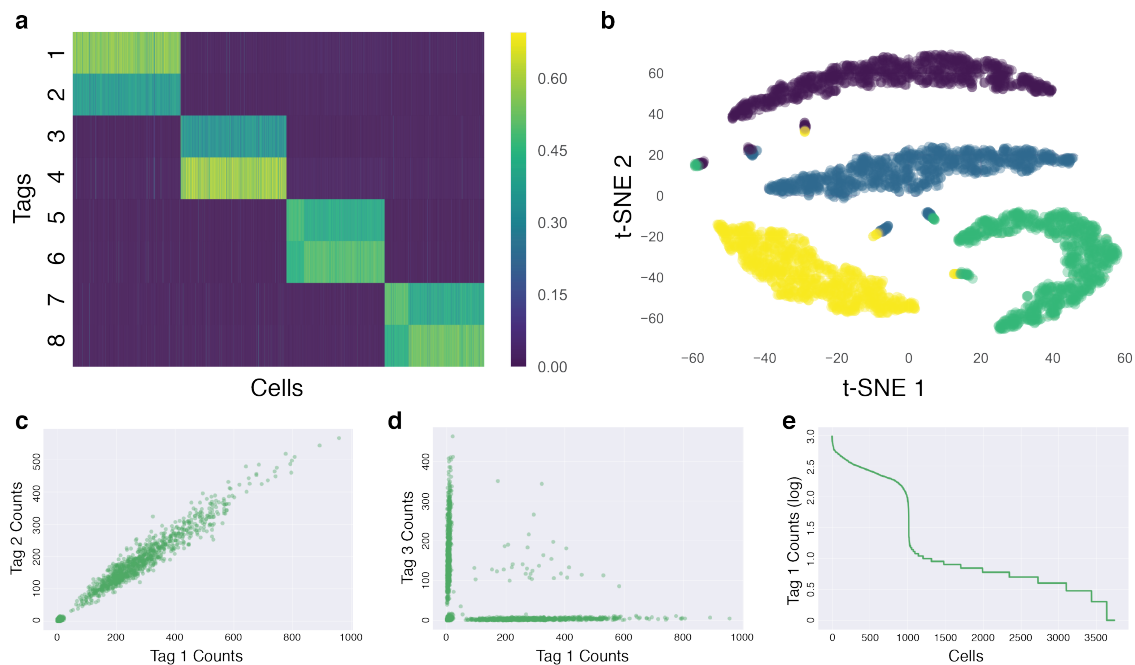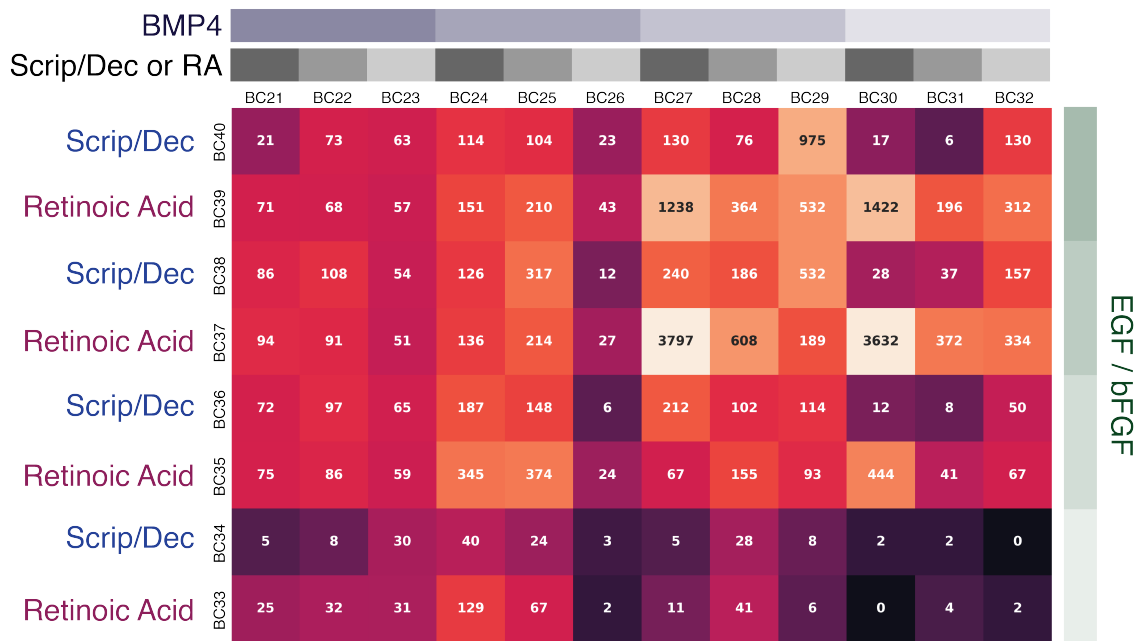

Supplementary Figure 2: Proof-of-concept sample tagging experiment (**a**) Heatmap showing 3,768 detected cells originating from four methanol-fixed samples each labeled by a pair of sample-specific tags. (**b**) t-SNE visualization of sample tag data colored by k-means clustering (k=4). Four main clusters are observed, corresponding to the four individual samples, as well as $6 = \binom{4}{2}$ small clusters corresponding to each possible combination of cell doublet originating from two different samples. (**c**) Scatter plot of counts for tags 1 and 2, which were used to label the same sample. The low-count population (bottom-left) is background from droplets not containing cells from the sample, while the high-count population corresponds to positive cells from the sample and shows a striking correlation between the two tag counts (Pearson's correlation coefficient r = 0.96). (**d**) "Barnyard plot" showing two tags from separate samples. Tags are clearly orthogonal, with doublets easily identified. (**e**) Counts for tag 1 from each cell in the experiment, ordered from highest to lowest and showing a clear inflection point between Tag 1 (+) and Tag 1 (-) cells.
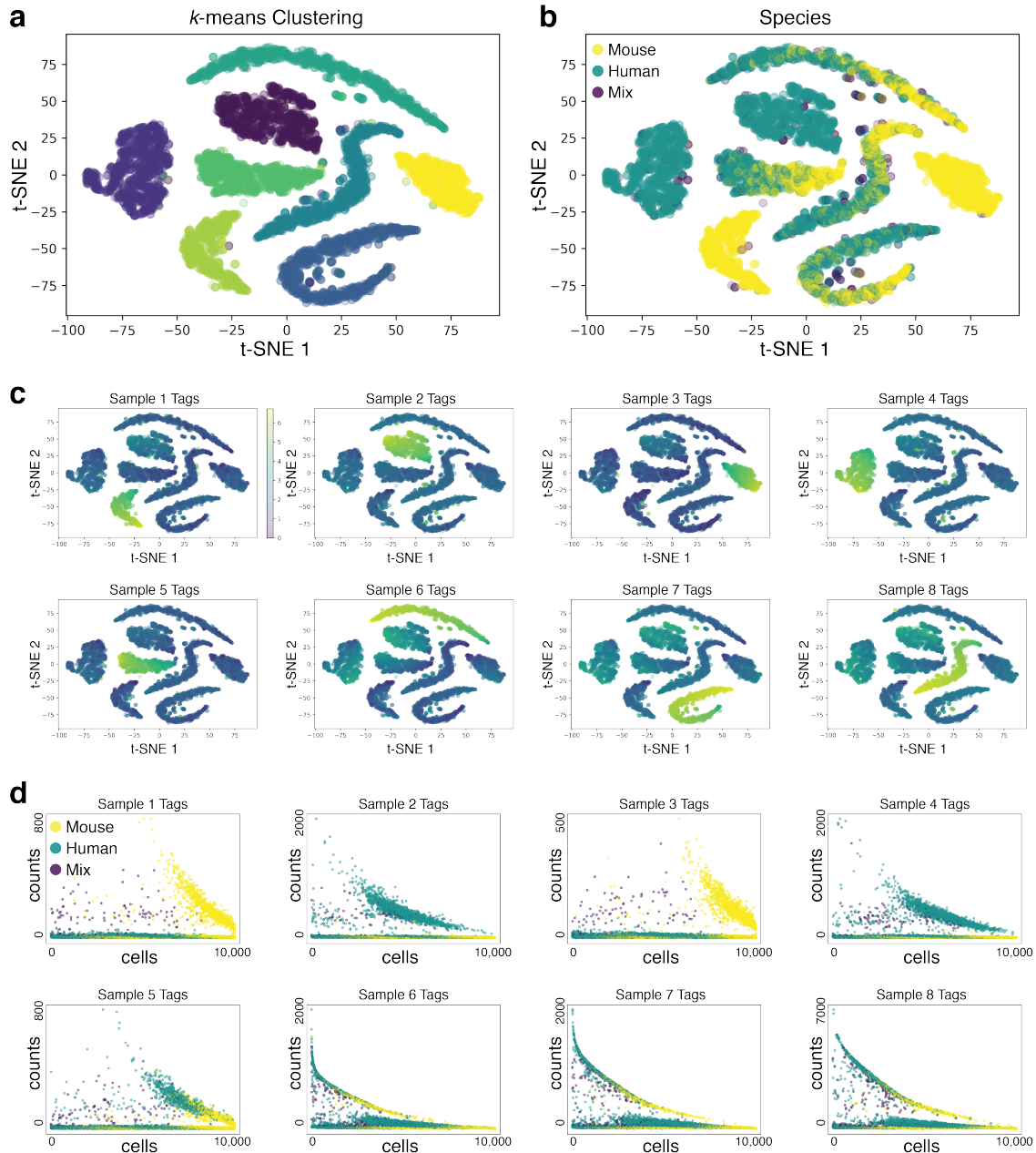
Supplementary Figure 3: Species mixing and tag multiplexing. Mouse neural stem cells and human HEK293T cells were labeled as follows: Sample 1: mouse, one label; Sample 2: human, one label; Sample 3: mouse, two labels; Sample 4: human, two labels; Sample 5: mix, two labels; Sample 6: mix, three labels; Sample 7: mix, 4 labels; Sample 8: mix, 5 labels (**a**) 10,054 cells were detected. t-SNE of sample tags x cells count matrix, colored by sample assignment from *k*-means clustering performed on a matrix normalized for tag numbers and counts per cell (see Methods). Eight major clusters are clearly identified. (**b**) t-SNE colored according to species assignment based on cDNA content. Four clusters represent a single species, and the remaining four are mixed, concordant with the experimental design. Cells identified as a mix of human and mouse are explained as cell doublets, and as expected fall outside of the major clusters, indicating a mix of sample tag signals as well. (**c**) t-SNE representation with detected cells colored as the logarithm of the sum of sample tags used in each of the eight experimental samples. (**d**) Sum of sample tag counts for each sample across all detected cells. Smaller NSCs present fewer sample tags than HEK293Ts from the same samples, indicating a correlation between cell size and the extent of labeling.
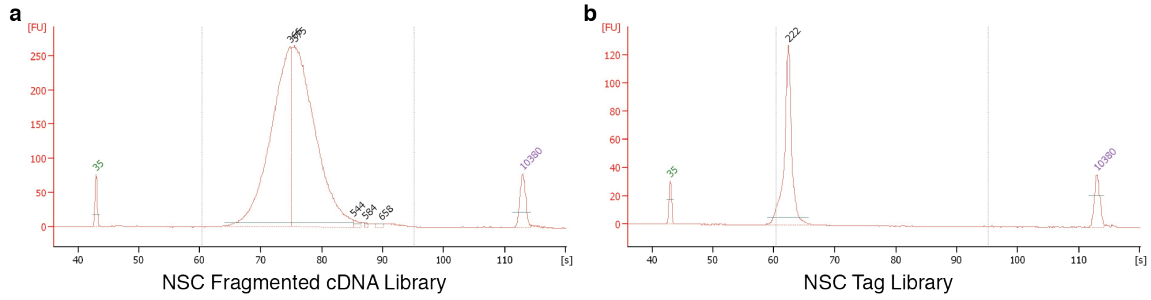
Supplementary Figure 4: Representative BioAnalyzer traces for (**a**) fragmented cDNA libraries and (**b**) sample tag libraries.

		BC21	BC22	BC23	BC24	BC25	BC26	BC27	BC28	BC29	BC30	BC31	BC32
Scrip/Dec	BC40	21	73	63	114	104	23	130	76	975	17	6	130
Retinoic Acid	BC39	71	68	57	151	210	43	1238	364	532	1422	196	312
Scrip/Dec	BC38	86	108	54	126	317	12	240	186	532	28	37	157
Retinoic Acid	BC37	94	91	51	136	214	27	3797	608	189	3632	372	334
Scrip/Dec	BC36	72	97	65	187	148	6	212	102	114	12	8	50
Retinoic Acid	BC35	75	86	59	345	374	24	67	155	93	444	41	67
Scrip/Dec	BC34	5	8	30	40	24	3	5	28	8	2	2	0
Retinoic Acid	BC33	25	32	31	129	67	2	11	41	6	0	4	2

Supplementary Figure 5: Organization of 96-plex perturbation experiment. Matrix entries correspond to the number of cells recovered from each sample.