

Cite as: I. Mazelis *et al.*, *Science*
10.1126/science.ady7209 (2025).

Multistep genomics on single cells and live cultures in subnanoliter capsules

Ignas Mazelis¹, Haoxiang Sun¹, Arpita Kulkarni², Theresa L. Torre¹, Allon M. Klein^{1*}

¹Department of Systems Biology, Harvard Medical School, Boston, MA, USA. ²Single Cell Core, Harvard Medical School, Boston, MA, USA.

*Corresponding author. Email: allon_klein@hms.harvard.edu

Single-cell sequencing methods uncover natural and induced variation between cells. Many functional genomic methods, however, require multiple steps that cannot yet be scaled to high throughput, including assays on living cells. Here we develop capsules with amphiphilic gel envelopes (CAGEs), which selectively retain cells and large analytes while being freely accessible to media, enzymes and reagents. Capsules enable high-throughput multistep assays combining live-cell culture with genome-wide readouts. We establish methods for barcoding CAGE DNA libraries, and apply them to measure persistence of gene expression programs in cells by capturing the transcriptomes of tens of thousands of expanding clones in CAGEs. The compatibility of CAGEs with diverse enzymatic reactions will facilitate the expansion of the current repertoire of single-cell, high-throughput measurements and their extension to live-cell assays.

Single-cell genomics encompasses a set of methods whereby hundreds to millions of cells are individually subjected to highly-multiplexed assays including DNA sequencing, chromatin accessibility or modification, RNA sequencing, and combinations thereof (1, 2). These methods enable unbiased, systematic discovery of cellular phenotypes and their dynamics, including immune cell responses, transitional states in tissue development, and changes in disease or upon perturbation. However, integrating these tools with each other or with functional cell screens faces limitations because the amount of material that can be obtained from each single cell is small and often requires reaction conditions incompatible with each other (1). Balancing sensitivity, scale, and genome-wide breadth with live culture capabilities imposes trade-offs between assay type, performance, and scalability.

Single-cell sequencing methods are currently implemented via three technological modalities: in wells (e.g., SMART-seq3) (3, 4), in micro-compartments (droplets or nano-wells) (5–7), or by in-cell barcoding (ICB) of cross-linked cells (sciRNA-seq and SPLiT-seq) (8, 9). These methods each offer different trade-offs between sensitivity and number of cells that can be analyzed, and each suffer different limitations regarding the complexity of assays they enable. Well-based methods are typically the most sensitive, possibly because each cell is processed in multiple steps in a manner resembling bulk methods. These methods also enable carrying out functional or live-imaging assays on living cells prior to genomic analysis. However, well-based methods scale poorly. Micro-compartment-based methods achieve increased throughput by physically isolating individual cells together with individual DNA-barcoded beads, enabling cell-specific barcoding of material from thousands of cells together within one sample. However, these methods are limited to a single reaction step once cells are compartmentalized. ICB methods

scale to exponentially large numbers of cells, but require cell cross-linking, which leads to material loss and reduced measurement sensitivity (8, 9). Neither droplets nor ICB methods allow parallelizing functional live-cell assays. Accordingly, approaches have been developed to pre-process cells prior to their analysis by these modalities, such as single cell ELISA (10, 11) or DNA tagging of cell populations (12, 13).

A technology that may unify advantages of these modalities is semi-permeable capsules (14, 15). Capsules are subnanoliter spherical liquid compartments surrounded by a resilient shell. The chemical composition of the shell can be chosen to be selectively permeable to small molecules including tissue culture media, small proteins such as DNA polymerases and ligases, and oligonucleotides, while selectively retaining larger macromolecules like DNA or mRNA (Fig. 1A). Like tissue-culture wells, they have the potential to allow culturing cells in isolation prior to analysis, with free exchange of media and nutrients. They also allow carrying out optimized molecular biology reactions on compartmentalized cells or analytes by enabling the repeated exchange of buffers and reagents. Like droplets, capsules can be produced at scale. Their resilient shell allows sorting to enrich cells or colonies with desired features (15). These properties expand the range of genomic assays that are feasible for single cells and live-cell cultures and present a versatile technological platform that may allow for the parallelization of assays that, until now, have been difficult and costly to carry out at high throughput (Fig. 1B). Here, we report the development of capsules with appropriate permeability and stability for genomic assays. We implement a split-pool barcoding-based approach to generate single-capsule genomic and transcriptomic libraries, and we apply these developments to perform a massively-parallel live-cell assay characterizing the diversity and persistence of transcriptome-wide expression fluctuations in

clones. The utility of capsules for multistep genomics is also shown in an accompanying study (16) exploring different applications.

Developing capsules with amphiphilic gel envelopes (CAGEs)

Capsules are generated by forming concentric layers of core and shell components (17), with one approach using liquid-liquid phase-separation (LLPS) between two polymers inside microfluidic droplets (14, 15). The outer polymer can then be cross-linked to form a stable shell. For biological applications, several studies have reported production of capsules consisting of a polyethylene glycol diacrylate (PEGDA) shell (14, 15). However, PEGDA fails to form stable shells under changes in buffer pH, salinity, and concentration, thereby limiting the ability to control capsule permeability (18, 19), (fig. S1A). Amphiphilic molecules should more reliably wet the water-oil interface of a droplet than PEG, so we speculated that such molecules would be more suitable for producing stable concentric shells through spontaneous LLPS. We tested the addition of Pluronic F127, a biocompatible amphiphilic block-copolymer, and found that it substantially improved the production of PEGDA capsules (fig. S1, B and C). This led us to use Pluronic F127 functionalized with diacrylate (F127DA) directly as a shell building block, without any PEGDA. The production of Pluronic diacrylate capsules occurs through LLPS between F127DA and a core polymer, dextran, inside microfluidic droplets. Once formed, the F127DA is covalently cross-linked into a physical gel using a photoinitiator and brief exposure to 405nm light (Fig. 1C). The resulting capsules are optically clear (Fig. 1D), tunable by varying polymer concentrations (fig. S1D), can be reliably formed at scale ($90 \pm 9\%$ recovery, table S1), are monodisperse ($CV < 3\%$, fig. S1E), stable over time (fig. S1F), and can be used to capture cells inside during production ($79 \pm 6\%$ capture rate, table S1 and movie S1). Additionally, the strategy is generalizable: capsule production is possible with other amphiphilic copolymers of different lengths, including other Pluronics (P123DA, F68DA, and F108DA) and PPPDA (acrylate-polycaprolactone-PEG-polycaprolactone-acrylate) (Fig. 1E), enabling the formation of capsules with varying shell properties. We refer to these capsules as CAGEs: Capsules with Amphiphilic Gel Envelopes.

We next optimized the selective permeability of CAGEs. We focused on using F127DA because Pluronic F127 forms a hydrogel consisting of a packed micellar array, which would set a shell pore size comparable to the micelle diameter (radius 10-20 nm), (20, 21). Indeed, scanning electron cryo-microscopy (cryo-SEM) imaging shows that the F127DA shell has a foam-like structure, with surface pores 10-50 nm in diameter (Fig. 1F). Genomic assays utilize small enzymes (gyration radii 3-5nm) and oligonucleotide primers of up to ~100

nt in length (gyration radii < 10 nm) (22) that should diffuse through these pores, while the double-stranded DNA (dsDNA) generated during library preparation should not (persistence length > 50 nm for products > 150 base pairs (23)). We generated CAGEs with a shell composed of F127DA and other copolymers, and measured mean lifetime (T_{Diff}) associated with the diffusion of fluorescently-labeled proteins, DNA and RNA through the CAGE shell (Fig. 1G and fig. S1, G to I). We identified a polymer composition for CAGE formation for which dsDNA of length ≥ 300 bp showed no diffusion through the CAGE shell (diffusion lifetime $T_{\text{Diff}} > 14$ days) in any buffer tested, while immunoglobulins and 100 nt single-stranded DNA (ssDNA) were able to diffuse into the CAGEs ($T_{\text{Diff}} \sim 5$ min) (Fig. 1, H and I, and table S2). The diffusion time remained stable over at least 10 days, and following centrifugation and incubation at different temperatures (fig. S1, J and K). The diffusion rate across the shell slowed with decreasing salt concentration (Fig. 1I) and with increasing temperature (fig. S1L). This makes F127DA CAGEs particularly suitable for high temperature reactions, such as polymerase chain reaction (PCR) (fig. S1G), where retention of molecules at high temperature is desirable. The dependence of diffusion on salt concentration also suggests a strategy for modifying buffers for washing, storing or loading CAGEs. Further, F127DA offers a stable base for capsule formation, allowing diffusion rates to be tuned by altering polymer composition (fig. S1M).

CAGEs can be handled with common lab equipment, can be analyzed by flow cytometry, and can be sorted through a commercial fluorescence-activated cell sorting (FACS) instrument to selectively fractionate them based on fluorescence (Fig. 1J). Taken together, these results demonstrated that CAGEs offer a physically robust platform with a tunable permeability threshold. They enable exchange of oligonucleotides and proteins, while retaining large macromolecules, making them suitable for multistep molecular biology reactions and thus high-throughput genomics.

Barcoding DNA in CAGEs

We reasoned that CAGEs could enable massively parallel analyses by first compartmentalizing live cells or macromolecules, and subsequently performing sequential enzymatic reactions. This approach would require labeling the analytes in each compartment through DNA barcoding that later identifies their capsule-of-origin by sequencing.

Using CAGEs, we developed a flexible split-and-pool barcoding (8, 9) approach that allows indexing of double-stranded DNA molecules—the final form of all major analytical libraries analyzed by sequencing (Fig. 2A). The approach first amplifies the DNA derived from different assay reactions, such as reverse-transcription (RT) or tagmentation, to overcome inefficiencies in subsequent barcoding. This

amplification is carried out by PCR with a primer that contains a deoxyuridine (dU) residue, which can be efficiently cleaved to generate a 3' overhang and a 5' phosphate for subsequent barcoding (Fig. 2B and fig. S2A). This process can be iteratively repeated after each barcode addition to remove linker DNA sequences and to generate a 5'-P ligation site for the next round. In this work, CAGEs are subject to two rounds of split-pool ligation and one round of PCR DNA barcoding (schematic in Fig. 2A) after which the prepared libraries are readily released by dissolving the CAGEs (fig. S2, B and C). Additional barcodes can be added during initial library construction (through PCR, RT, or tagmentation) as well as indexing after release from CAGEs, scaling the number of barcodes. The number of capsules that can be profiled scales exponentially with the number of reaction steps (fig. S2, D and E). We refer to this method for indexing DNA libraries in capsules as inC-seq.

We tested five-step inC-seq barcoding on CAGEs. We encapsulated plasmid DNA and then barcoded the CAGEs with $14 \times 48 \times 48 \times 48 \times 16 = 24.7 \times 10^6$ potential barcodes. Applying this protocol to 350 μ L of packed CAGEs, we obtained 355,418 capsule barcodes (1 μ L corresponding to ~ 1000 capsules) with each of the possible barcode elements uniformly represented across all five rounds of barcode addition, yielding an effective theoretical barcoding space (perplexity) of 21.8 million barcodes (Fig. 2C). We next tested inC-seq on single-cell derived material by encapsulating cells carrying transgenes encoding red or green fluorescent proteins (RFP or GFP) into CAGEs and then amplifying these loci from genomic DNA (gDNA) (Fig. 2D and fig. S2F). The DNA libraries produced from cells demonstrated that the cells and genomic material within each CAGE were effectively compartmentalized through the entire multistep barcoding protocol, with $99 \pm 5\%$ (median $\pm$ SD) of reads from each CAGE mapping specifically to single GFP or RFP sequences, and with a doublet rate consistent with expectations for Poisson loading (Fig. 2D).

Single-cell genomic assays in CAGEs

DNA barcoding in capsules enables diverse analytical approaches to study DNA and RNA. We implemented single-capsule transcriptomics (inC-RNA-seq) and, separately, a single capsule assay for transposase-accessible chromatin (24) (inC-ATAC-seq) (Fig. 3, A and B, and fig. S2E). We used a mixture of human and mouse cells to confirm the effective isolation of cellular material in each CAGE, with a cell doublet rate again limited by Poisson loading (Fig. 3, C and D, and fig. S3A). For inC-ATAC-seq, we observed a strong enrichment of fragments at transcription start sites (17-fold enrichment) with low mitochondrial reads (<3%) and high fraction of reads in peaks (fig. S3, C to E). inC-ATAC-seq is not yet optimized and sensitivity was low (median 2.3k fragments/cell, fig. S3, F and G) but nonetheless identified differentially-

accessible regions in a mixture of three cell lines (fig. S3H).

We then focused on optimizing methods for high sensitivity single-cell inC-RNA-seq, using an optimized well-based approach (SMART-seq3) as a starting point (3). The flexibility of CAGEs allowed us to optimize sequential steps of lysis, RT and pre-amplification in cell lines (HEK293T and NIH3T3), resulting in a protocol that captures as many transcripts and genes as a commercial droplet-microfluidic system (10x Genomics, GEMCode v3.1) (Fig. 3E, and fig. S4, A and B). We observed high reproducibility between independently prepared and barcoded CAGEs (fig. S4C, log-correlation $R > 0.99$) and between CAGEs and GEMCode (fig. S4D, log-correlation $R = 0.95$). The porosity of capsules raises a question of whether smaller mRNA transcripts might be under-represented, but notably we find no apparent bias in the length of detected mRNA transcripts in CAGEs, in line with our diffusion assessment data (fig. S4E). Further, we observed comparable transcript counts and low length-bias when performing inC-RNA-seq with different sample preparations: (i) cells fixed with formaldehyde for which cross-linking was reverted in CAGEs prior to RT (Fig. 3F and fig. S4F), (ii) material cryopreserved after cell lysis in CAGEs (Fig. 3G and fig. S4G), and (iii) cells encapsulated while still in tissue culture media (Fig. 3H and fig. S4H). Notably, culture media can inhibit RT, yet performance remained comparable to buffer-washed samples. These results indicate the versatility of CAGEs for complex assay design, and suggest their utility for profiling difficult-to-process samples, such as microbes (14, 15), cells from marine organisms (25) or clinical formaldehyde-fixed samples.

To complete benchmarking inC-scRNA-seq, we profiled human peripheral blood mononuclear cells (PBMCs), which have been frequently analyzed by other methods. Starting with CAGEs containing an estimated 60,000 PBMCs, we obtained transcriptomes for 43,665 PBMCs spanning the full landscape of PBMC states as seen by a uniform manifold approximation and projection (UMAP) plot (Fig. 3, I to K), including rare circulating CD34+ progenitors, plasmoblasts, and a clear partitioning of subsets of T cells and natural killer (NK) cells. Our method captured comparable numbers of transcripts and genes as the GEMCode v3.1 and SMART-seq3 (3) methods (Fig. 3, L and M, and fig. S4, I and J), and considerably more transcripts than reported with split-and-pool ICB methods (9, 26). Overall, inC-seq approaches provide a versatile platform for high-throughput genome-wide measurements, with the performance matching established transcriptomic profiling methods, and with the added flexibility of conducting multistep reactions.

Massively parallel live-cell assays in CAGEs

Capsules have the potential to enable highly parallel functional assays in which cells are allowed to grow and/or interact prior to molecular profiling. To enable cell culturing in

CAGEs, we introduced dithiothreitol and bovine serum albumin to buffer against reactive oxygen species (ROS) formed during shell photopolymerization (27), which otherwise led to large-scale death of encapsulated cells (fig. S5, A and B). We also removed excess fluorinated oil, which acts as an oxygen reservoir. With these changes, cells captured in CAGEs are viable and clonally expand as demonstrated with several cell lines (L1210, K562, Jurkat, HEK293T, L929) (Fig. 4, A and B), with division times matching those of cells grown in plates (Fig. 4C). Culturing in CAGEs did not induce major transcriptomic changes (log-correlation $R^2 > 0.95$; < 10 genes with > 2 -fold change at 5% false discovery rate fig. S5C) or stress (fig. S5D). CAGEs also support viable growth of primary mouse bone marrow-derived hematopoietic stem cells and human induced pluripotent stem cells, with the latter forming cyst-like structures (Fig. 4A) and maintaining expression of pluripotency markers (fig. S5, C and E). CAGEs can be produced at > 2 million/hour and thus enable high-throughput expansion of isolated colonies at scale (Fig. 4B and fig. S5F), all of which can then be analyzed with inC-RNA-seq.

We utilized this ability to grow cells in CAGEs in order to measure persistent clonal heterogeneity in gene expression. In several cancer types, cells occupy persistent epigenetic states, which are thought to underlie non-genetic heterogeneity in drug resistance (28, 29). Drugs targeting the inheritance of epigenetic modifications (DNA methyltransferase (DNMT) inhibitors (30); and histone deacetylase (HDAC) inhibitors (31)) have been proposed as tools to activate tumor suppressor genes that are silenced, and they are used in combination with more traditional chemotherapies (32–34). Whether these drugs broadly reduce epigenetic persistence in gene expression, however, has not been explicitly tested. To measure the duration of epigenetic persistence, other studies have used RNA-seq to identify clonally-variable gene expression patterns – an approach so far carried out by culturing isolated cells in wells, with analysis restricted to a few dozen clones (35, 36). Using CAGEs, one can parallelize this live-cell assay to carry out tens of thousands of clonal growth experiments. We did so with cell clones derived from a mixture of human erythroleukemia (K562) and mouse lymphoblastoma (L1210) cells, and evaluated changes in clonal heterogeneity in the presence or absence of the DNMT inhibitors decitabine (Dec) and 5-aza-cytidine (Aza), and the HDAC inhibitor vorinostat (Vor) (Figs. 4, D to H, and 5).

As shown schematically in Fig. 4D, cells were pre-incubated with Dec, Aza, Vor or vehicle alone at sub-lethal doses for two days (approximately 25% reduction in growth rate, see fig. S6A for dose-survival curves). They were then encapsulated and grown in CAGEs over a period of 6 days with continued drug treatment. At progressive time points, the clones were sampled, lysed in CAGEs and then frozen prior to inC-

RNA-seq. In total, we analyzed transcriptomes from 134,805 CAGEs representing single cells sampled at day 0 ($n=37,902$) and expanded clones grown 2, 4, or 6 days ($n=96,903$) in different treatment conditions. An equivalent analysis of this number of isolated colonies in wells would require 1,010 96-well plates. In both K562 and L1210 clones, we identified genes with above Poisson coefficients of variation in gene expression (CV, or standard deviation/mean; shown for untreated cells at days 4 and 6 in fig. S6, B and C). The most variable genes between clones defined modules of gene expression that persisted over time, as seen from clustering gene-gene correlations and comparing the results to those from mock clones generated by randomly combining single cell transcriptomes (Fig. 4E and fig. S6D). Persistent clonal heterogeneity can also be appreciated by UMAP embedding of the clones (fig. S6E).

To identify persistent gene expression programs varying between clones, we decomposed their expression by nonnegative matrix factorization (NMF). Approximately 12–15 programs explained variation in gene expression above random for both cell lines [eigenvalue crossover method (5), fig. S6F]. Several of these were readily identifiable from their gene loadings, including cell cycle-associated programs (K562 programs 12,13 and L1210 program 14), erythroid differentiation (K562 program 3), putative epithelial (*KRT8/18*-enriched) and mesenchymal (*VIM*-enriched) programs (K562 programs 4, 14) as well as myeloid programs expressing the transcription factors *Irf8* and *Myc*, interleukin receptor *Il7r* (L1210 program 0); and myeloid transcription factor *Klf6*, M-CSF (*Csf1*) and osteopontin (*Spp1*) (L1210 program 7) (Fig. 4F, fig. S6G, and table S4). These programs remained variable after clonal expansion beyond that expected from mock clones (Fig. 4G and fig. S6H). Several programs also showed mutually exclusive patterns of expression: for example, K562 day 6 colonies were either enriched in program 14 (*VIM*-hi) or program 3 (erythroid, *HBE1/HBA2/ALAS2*-hi) but not both ($p < 10^{-50}$, Fisher's exact test, adjusted to correct for false discovery). By contrast, mock colonies showed mixing of the two programs (Fig. 4H and fig. S6, I and J).

Next, we explored how Aza, Dec or Vor treatments alter gene expression dynamics. Drug treatment that alters epigenetic inheritance could bias the fraction of cells that express a program, or alter the rate at which a program switches states, or both (Fig. 5A). To quantify these two effects, we introduced a statistical inference framework that fits the observed distribution of NMF program usages across clones over time to a stochastic model of state switching, parameterized by the persistence times of cells in a 'low' state ($1/r_{01}$) and 'high' state ($1/r_{10}$) for each program (Fig. 5, B and C, and supplementary text 1). From these fits we obtained changes in program persistence times (the relaxation time-scale to steady-state, $1/(r_{01}+r_{10})$) (Fig. 5D and fig. S7, A to C), and

steady-state bias (the fraction of time that a program is active, $r_{01}/(r_{01}+r_{10})$) (Fig. 5E and fig. S7D). These analyses revealed that all three drugs triggered changes in program persistence and bias in K562 cells, with the dynamics of some programs slowing down (higher persistence) and others accelerating. In K562 cells, all three drugs triggered differentiation toward the erythroid state (program 3 high state), with increased persistence in the differentiated state (Fig. 5, C to E). The changes in persistence times were correlated but distinct from changes in mean expression of individual genes (Fig. 5F). In L1210 cells, the DNMT inhibitors (Aza, Dec) showed distinct responses from the HDAC inhibitor (fig. S7D), but again the drugs led to both increases and decreases in state persistence times. Thus, contrary to what might be expected, the DNMT and HDAC inhibitors did not reduce clonal heterogeneity over several cell divisions. This observation is consistent with prior work that showed only modest changes in gene expression heritability after genetic ablation of DNA methyltransferases (36). The specific nature of the effects of these drugs, and their failure to globally reduce persistence, may explain why they sensitize chemotherapy in some patients but not others (37). More generally, the approaches taken here for live-cell genomic analysis following perturbations and growth may be generalizable to other colony assays, including on gastruloids and tissue-derived organoids.

Discussion

In summary, we have established a versatile platform for high-throughput single-cell and single-colony genomic analysis through the development and use of capsules with amphiphilic gel envelopes—CAGEs. The small compartments are biocompatible, quick to produce and have tunable chemistry. Alternative compositions can be evaluated through biophysical assays to define new permeability, mechanical, adhesive, or other functional properties. The approaches reported here for barcoding DNA libraries in capsules (inC-seq) and to study living cells can be adapted to a range of problems and may be further combined with sorting, fixing and staining steps, each of which is demonstrated here.

CAGEs also have some limitations. First, when CAGEs are used for single-cell assays, many are empty and increase the volume of subsequent reactions. This problem is alleviated in experiments where multiple cells are encapsulated, or by enriching for CAGEs containing cells. Second, while CAGEs are compatible with most molecular biology reactions, they may require slight changes in conditions to maximize efficiency and yield. The inC-RNA-seq workflow described in this work, for example, was optimized by changing reagent concentrations and reaction times, while inC-ATAC-seq is demonstrated but not yet optimized. Lastly, the F127DA CAGEs described in this work are not suitable for live-cell recovery, as they can only be solubilized using harsh reaction

conditions. Alternative CAGEs may be amenable to enzymatic solubilization.

Overall, CAGEs present a powerful tool for live-cell functional and genomic assays. They will likely enable the profiling of difficult-to-process samples, and may prove useful in studying perturbation outcomes that depend on cell–cell interactions, such as developmental or immune-cell interactions, and for performing colony-forming assays or studies of organoids at scale. Capsules may also allow the evaluation of libraries of signaling molecules, or may be used as vehicles for clonal evolution or selection followed by genomic profiling. These in turn may provide functional information that could support the development of predictive, data-driven models of cellular dynamics and interactions.

Materials and methods

CAGE synthesis

The following sections lay out the basic protocol for synthesizing CAGEs. Several modifications are made to the core and shell mixes and to incubation and wash conditions in subsequent methods sections. The production of CAGEs requires a droplet microfluidic system, and no additional equipment other than a 405nm lamp and standard molecular biology lab equipment.

Basic CAGE synthesis protocol

Capsules with amphiphilic gel envelopes (CAGEs) were prepared using a co-flow droplet microfluidic PDMS device (Darwin Microfluidics, DG-MCN-C4) by flowing shell and core solution at 170 μ L/h into fluorinated HFE7500 carrier phase supplemented with 2.5% (w/v) 008-FluoroSurfactant (Ran biotechnologies) infused at 580 μ L/h. The shell polymer solution consisted of 8% (w/v) F127DA (Creative PEGWorks, PPO-121-5 g), 1% (w/v) PPPDA (Sigma Aldrich, 929611-500MG) in 1x DPBS (Thermo Fisher Scientific, 14190136), unless stated otherwise. The core solution consisted of a 13% (w/v) dextran (Mw = 500,000Da) (Sigma Aldrich), 0.1% (v/v) of F68 (Thermo Fisher, 04196546SB) and 0.2% (w/v) of LAP (lithium phenyl-2,4,6-Trimethylbenzoyl phosphinate) (Sigma Aldrich, 900889-1G) in 1x DPBS, unless stated otherwise. Generated droplets were collected on ice into a 1.5mL Eppendorf tube.

CAGE polymerization

After collection, the tube containing the droplet emulsion was incubated on ice for 5 min, followed by incubation at 30°C for 3 min and removal of the bottom HFE7500 fraction, unless stated otherwise. Next, the tube was hand shaken 5 times and incubated at 30°C for an additional 2 min. The tube was carefully transferred to a bottomless tube holder and exposed to 405 nm light from below for 35s, unless stated otherwise.

Oil removal

After polymerization, the CAGEs were purified by removing the bottom HFE7500 phase, adding 100 μ L 20% (v/v) PFO (Sigma Aldrich, 370533-25G) in HFE7500, hand shaking, and spinning down for 5s. One additional wash with HFE7500 was performed after removal of PFO/HFE7500 followed by three 5s spins to remove any remaining HFE7500 droplets. The resulting hydrogel CAGE pellet was resuspended in a wash buffer as described below.

Modification 1: PEG shells

PEG-shell capsules (fig. S1A) were synthesized as described above, but replacing the shell mix with 8% (w/v) PEGDA in 1X DPBS (Mn = 700, Sigma Aldrich, 455008) and keeping other conditions as described above.

Modification 2: CAGEs made from other shell polymers

For Fig. 1E, CAGEs were synthesized as described above, but using a shell mix consisting of 1X DPBS and 8% (w/v) of the polymers [PPDPA (Sigma Aldrich, 929611-500MG); F127DA (Creative PEGWorks, PPO-121-5 g); P123DA (Sigma Aldrich, 915858-1G); F68DA (Sigma Aldrich, 914428-1G), F108DA (Sigma Aldrich, 914665-5G)]

Modification 3: cell encapsulation

Experiments encapsulating cells used the same shell mix as in the *Basic CAGE synthesis protocol*, but with 8mM DTT and 1% (w/v) BSA added to the core mix. This was used for experiments described in Figs. 2D and 3, C, and E to G; and figs. S3B and S4, A to H.

Modification 4: cell encapsulation in media

Experiments encapsulating cells used the same shell mix as in the *cell encapsulation* protocol above, but the base buffer in the core mix was changed from 1X DPBS to IMDM and polymerization time extended to 45s. This was used for experiments described in Figs. 3, H to M, 4, and 5; and figs. S3, A, and C to F; S4, I and J; S5; S6; and S7.

CAGE washing

CAGE washing allows replacing buffers or removing enzymes and cofactors, and it is carried out at multiple points in the experiments described in this paper.

Basic CAGE wash protocol

A basic protocol involves resuspending up to 300 μ L of packed CAGEs in 1000 μ L of wash buffer in a DNA Lo-Bind 1.5mL Eppendorf tube, then pipetting using a P1000 pipettor and pelleting the CAGEs by centrifugation for 60s. The centrifugation is carried out at 300 g when CAGEs contain live cells, and at 800 g otherwise. The supernatant is then aspirated leaving a small volume above the packed CAGEs so as not to

disturb them. For larger volumes, CAGEs were washed in 15mL falcon tubes with 5mL of buffer and centrifugation time extended to 3 min.

Wash buffers

Different CAGE wash buffers were used as described below with corresponding composition: **IFLB**: IMDM with 1% (w/v) BSA, 0.1% (w/v) Pluronic F127 and 0.1% (w/v) Pluronic L31; **IFL**: IMDM with 0.1% (w/v) Pluronic F127 and 0.1% (w/v) Pluronic L31; **TI**: 10 mM Tris pH 8.0, 0.1% Igepal CA-630 (Millipore Sigma, I8896-50ML); **TI10**: 10 mM Tris pH 8.0, 10mM NaCl, 0.1% Igepal CA-630; **TI100**: 10 mM Tris pH 8.0, 100mM NaCl, 0.1% Igepal CA-630; **TIS**: 10 mM Tris pH 8.0, 0.1% Igepal CA-630, 0.05% SDS; **TEIS**: 10 mM Tris pH 8.0, 0.1mM EDTA, 0.1% Igepal CA-630, 0.1% SDS; **HI10**: Nuclease free water, 10mM NaCl, 0.1% Igepal CA-630; **3xSSCI**: Nuclease free water, 15%v/v 20x Sodium saline citrate buffer (pH 5.9) (Thermo Scientific, J60839.K3), 0.1% Igepal CA-630.

Basic post PCR wash steps

After PCR, CAGEs were washed three times with TI100 (pH 8.0) buffer facilitating faster diffusion, then twice with detergent-containing TIS (pH 8.0) buffer to denature enzymes, and twice with a low-salt buffer TI10 (pH 8.0) before resuspending in a buffer of interest.

Post dU PCR wash steps

After DNA amplification with dU primers, CAGEs were washed three times with TI100 (pH 8.0), twice with TIS (pH 8.0), three times with TI100 (pH 8.0) and twice with TI10 (pH 8.0). This procedure is as for *basic post-PCR wash steps* but including three additional washes.

Post reverse-transcription wash step

After RT, CAGEs were washed three times with TI100 (pH 8.0), twice with TIS (pH 8.0), three times with ice-cold TI100 (pH 8.0) with 2 min incubations on ice. Finally, CAGEs were washed twice with TI10 (pH 8.0).

Dextran core digestion

Core dextran digestion was performed by resuspending up to 300 μ L of CAGEs in 1mL IFLB wash buffer (see *Wash buffers*) and adding 0.5% (v/v) solution of dextranase from *Chaetomium erraticum* (Sigma Aldrich, D0443-50ML). The reaction was incubated at 37°C for 1 min with 700 rpm shaking. Immediately after, CAGEs were spun down at 300 g for 1 min. The supernatant was removed and CAGEs were resuspended in 1 mL of IFLB buffer. CAGEs were washed once more with 1mL IFLB before proceeding to the next step.

CAGE yield and cell capture efficiency

This section describes tests carried out to evaluate the yield

of CAGEs and the fraction of cells in an input sample that are successfully encapsulated (supporting table S1). We define the number of CAGEs recovered as N_{CAGE} and the mean occupancy of cells per CAGE as λ_{cells} . The CAGE fractional yield was calculated as $N_{CAGE}/N_{CAGE}^{(expected)}$, cell recovery as $\lambda_{cells}/\lambda_{cells}^{(expected)}$, and absolute cell capture as $N_{encap.}/N_{encap.}^{(expected)}$, where $N_{encap.} = N_{CAGE} \times \lambda_{cells}$ is the total number of cells encapsulated. To evaluate these quantities, CAGEs were synthesized using 60 μ L of core solution with cells at a concentration of 1.2 M/mL (see *CAGE Synthesis, cell encapsulation*), followed by dextranase treatment and three washes with 1mL IFLB wash buffer as described above. This corresponds to an expectation of generating $N_{CAGE}^{(expected)} = 72 \cdot 10^4$ CAGEs, with an average occupancy $\lambda_{cells}^{(expected)} = 0.10$ cells/CAGE and a total $N_{encap.}^{(expected)} = 72 \cdot 10^3$ cells encapsulated. Then, the actual number of CAGEs recovered (N_{CAGE}) and the mean number of cells per CAGE (λ_{cells}) were estimated manually using a hemocytometer (numbers in table S1) and the formulae above were used to estimate the final yields.

Cryo-SEM imaging of CAGEs

CAGEs were stored in TI buffer (see *wash buffers*) prior to imaging. A 10 μ L sample was loaded onto each of three gold planchets (Leica #16770132; Carrier, Dia. 3.0x0.8mm, Cu-Au, dome) and rapidly frozen using slushy nitrogen (made using Edwards Plate Degasser PD3; Model RV3, Serial No. 056188208, Code No. A652-01-903) and transferred to the Leica EM VCM loading station (SN 614252, #16771611104). The positioned frozen planchets were next transferred from the Leica EM VCM loading station to the Leica ACE 600 sample preparation chamber (SN 601238, #16771525) with the help of the Leica EM VCT 500 vacuum cryo-transfer system maintained at -150°C using liquid nitrogen. In the chamber, the samples were knife-fractured at -150°C, sublimated at -105°C for 9 min and 10 s, then re-cooled to -150°C. The samples were sputter-coated with a 10nm layer of Platinum/Palladium (80:20) and transferred to the Zeiss Gemini 360 Field Emission Scanning Electron Microscope (FESEM) (GeminiSEM360 349599-9100-010, SN 8217010169), with the help of the VCT 500 vacuum cryo-transfer system. Imaging was performed at -150°C on a temperature-controlled FESEM stage at 3 keV using the Everhart Thornley detector.

Analyte cross-shell diffusion assays

Rates of analyte diffusion through the CAGE shell were assessed using time-lapse confocal microscopy on an A1R Nikon Confocal laser-scanning microscope using a 20X objective

(fig. S1H). Fluorescent analytes, loaded into CAGEs as described below, generate a bright signal in the CAGE core that decreases over-time as molecules diffuse out into the non-fluorescent surrounding buffer. CAGEs were imaged in 96-well glass bottom plates (MatTek, P96G-1.5-5-F) until the fluorescence signal from the core decreased below 1/e (36.8%) of initial value or up to 14 days. Acquired time-lapse movies were processed by segmenting CAGEs from each field of view using a custom PyTorch convolutional neural network (38) and average core fluorescence intensities were extracted. Average fluorescence time-series were fitted with double exponential decay and time-scales were extracted as time required to reach 1/e of the initial intensity. Values for analyzed buffers and analytes are provided in table S2.

Small analyte diffusion

Cross-shell diffusion half-lives of 63-100 nt ssDNA, 100bp DNA and IgG were measured as described above after first incubating empty CAGEs in a buffer containing the analytes (fig. S1H, left). For this step, CAGEs were synthesized without analytes (*Basic CAGE synthesis protocol*, above) and then washed 3 times with a fast-diffusion (high-salt) buffer TI100 (pH 8.0) for loading analytes. Subsequently, 5 μ L of packed CAGEs were transferred to 1.5mL tubes. 1 μ L of fluorescently labeled analyte solution (10 μ M Cy5-63nt, 10 μ M FAM-100nt, 5 mg/mL FITC Anti-mouse CD45 IgG (BioLegend, 110705)) was added onto 5 μ L of packed CAGEs and left on ice for 2 hours. See table S4 (diffusion) for sequences of the oligonucleotides Cy5-63nt, FAM-100nt. After incubation, CAGEs loaded with analytes were washed with TI (pH 8.0) buffer twice followed by a single wash with a buffer containing different salt concentrations (TI, TI10, TI100). After removal of supernatant, the packed CAGEs were transferred to a 96 well glass bottom plate and resuspended to a final volume of 40 μ L in a buffer of interest for imaging.

dsDNA diffusion

dsDNA that is too large to quickly load into CAGEs by diffusion was generated within CAGEs using a PCR reaction (fig. S1H, middle). First, CAGEs were generated following the normal procedure (*Base protocol*, above), but with plasmid DNA added into the dextran core mix at a final concentration of 10pg/mL prior to encapsulation. CAGEs were polymerized and washed three times with fast-diffusion (high-salt) buffer TI100 (pH 8.0), twice with TI10 (pH 8.0) and twice with TI (pH 8.0). The analyte was then generated in CAGEs in a 25 μ L PCR reaction consisting of 12.5 μ L of packed CAGEs, 2.5 μ L of 10 μ M universal Cy3 labeled primer, 2.5 μ L of 10 μ M Xbp_RV primer (table S4, diffusion), 1 μ L of 10mM dNTPs (NEB, N0447L), 5 μ L of 5x Phusion HF reaction buffer (NEB, E0553L), 1 μ L of 10% Igepal CA-630 and 0.5 μ L of Phusion polymerase (NEB, E0553L). Here, X= 300, 500 and 1000

corresponds to primers generating different size amplicons (see table S4, diffusion). For each of the four primer pairs, the PCR reaction was carried out by first incubating the tubes on ice for 5 min, followed by a PCR cycling program: 98°C for 30s; 15 cycles of 98°C for 10s, 65°C for 20s, 72°C for 30s; and a final extension of 72°C for 5 min and hold at 25°C. After PCR, CAGEs were washed as described in *Basic Post PCR wash* and once with a buffer of interest. After removal of supernatant, the packed CAGEs were transferred to a 96 well glass bottom plate and resuspended to a final volume of 40 µL in a buffer of interest for imaging.

ssRNA diffusion

300/500/1000nt RNA molecules were generated in CAGEs by carrying out an IVT reaction (fig. S1H, right). First, dsDNA of appropriate size was amplified in CAGEs by PCR and then washed as described above in *dsDNA diffusion*, but with only 8 PCR cycles, and using a primer containing the T7 RNA Polymerase promoter sequence (table S4, diffusion). IVT was then carried out in 21 µL reactions, consisting of 1X T7 reaction buffer (NEB, M0251S), 8mM of ATP, GTP, CTP (NEB, N0466L), 6 mM of UTP (NEB, N0466L), 0.16mM of Cy3-UTP (APEX-BIO, B8330), 4U of T7 polymerase (NEB, M0251S) and 1.06U RiboLock RNase Inhibitor (Thermo Fisher, EO0381) for 60 min. After IVT, CAGEs were washed twice with TIS (pH 8.0), twice with TI100 (pH 8.0), twice with TI (pH 8.0) and once with the buffer of interest. After removal of supernatant, the packed CAGEs were transferred to a 96 well glass bottom plate and resuspended to a final volume of 40 µL in a buffer of interest for imaging.

CAGE flow cytometry

FITC fluorescent CAGEs used for optimization of fluorescence-activated CAGE sorting (FACS) were synthesized as described above (*Base protocol*, above), with Fluorescein isothiocyanate-dextran (FITC-dextran) (Sigma Aldrich) added to the core mix at 169 µg/mL. Non-fluorescent CAGEs were synthesized as above. After CAGE polymerization and HFE7500 removal, CAGEs were washed once with 300 µL of hexane. FITC+ and FITC- CAGEs were mixed so that the final fraction of FITC+ was 0.05-0.1. CAGEs were transferred to a 15 mL conical tube and spun down at 800 g for 1 min. Supernatant was removed and the CAGEs were then resuspended in TI buffer (*wash buffers*, above) to a final concentration of 500 - 1,000 CAGEs/µL in the sample tube via pipetting. CAGEs were sorted using a 200 µm nozzle on the Beckman Coulter MoFlo Astrios EQ high speed cell sorter. “Enrich” or “Purify” sort modes were used in addition to a “1-2 Drop” Drop Envelope. Drop Delay values were manually calibrated using FlowCheck beads (Beckman Coulter, 6605359). The Drop Delay value determined using the FlowCheck beads was offset by +/- 0.5 units. To check Drop Delay calibration using

CAGEs, 50 CAGEs were sorted onto a glass slide and quickly assessed for purity and recovery using fluorescence microscopy. Once calibrated, CAGEs were sorted into 1.5 mL Eppendorf tubes (Eppendorf). After sorting, the collection tubes were rinsed with TI buffer to dilute the sheath fluid and dislodge CAGEs from tube walls. Sorted sample purity and recovery were assessed by sampling the CAGEs and imaging them on a Nikon Ti1000 wide-field fluorescent microscope. Purity and recovery were estimated by imaging the sorted CAGEs on a microscope (Nikon A1R confocal, 20X magnification) and counting the number of fluorescent and non-fluorescent CAGEs.

CAGE dU PCR

Base protocol

A 50 µL volume PCR reaction utilizing a dU-containing primer (dU PCR) was carried out as follows: a reaction mix was assembled consisting of 10 µL of 5X Q5U reaction buffer (NEB, M0515L), 2 µL of 10mM dNTPs (NEB, N0447L), 5 µL of 10 µM FW primer (protocol specific), 5 µL of 10 µM dU universal RV primer (table S4), 25 µL of packed CAGEs, 1 µL of Q5U High-Fidelity DNA polymerase (NEB, M0515L), and 2 µL of 10% Igepal CA-630 (Millipore Sigma, I8896-50ML). To set up the reaction, CAGEs were transferred to PCR tubes on ice and the remaining PCR reagents were added ensuring proper mixing by pipetting. The reaction was carried out by first incubating the tubes on ice for 5 min, followed by a PCR cycling program: 98°C for 30s; 6 cycles of 98°C for 15s, 68°C for 20s, 72°C for 4 min; a final extension of 72°C for 2min and hold at 25°C. After the DNA amplification, CAGEs were washed three times as described above in *CAGE washing - Post dU PCR wash steps*.

Modifications

For barcoding libraries of plasmids (Fig. 2C), the PCR reaction was carried out in parallel in 14 wells after first distributing the capsules, and then using a different dU-containing primer in each well. These primers contained an additional barcode (see table S2). For other genomics assays (inC scRNA-seq and scATAC-seq) the CAGE dU PCR steps used a universal dU primer that contains no barcodes (here, barcodes were introduced prior to PCR during reverse-transcription and transposition reactions). For Fig. 2B, CAGEs were imaged during this process on a Nikon Ti2000 epifluorescent microscope at 20X magnification.

USER-based dU excision

For dU excision, a reaction was performed in TI10 (pH 8.0) buffer with 0.01U/µL of USER™ enzyme (NEB, M5505L) in a shaking heat block at 37°C for 20 min. Following the reaction, CAGEs were washed twice with TI100 (pH 8.0), once with TIS (pH 8.0) and TI10 (pH 8.0). To fully remove the cleaved DNA

sequence, CAGEs were washed in 1 mL of TI (pH 8.0) buffer with a 5 min incubation at 60°C, twice, with a room temperature TI (pH 8.0) wash in between.

CAGE split-and-pool barcoding

Barcoding is carried out through several steps, which add the Ligation-1 barcode, Ligation-2 barcode, and PCR-3 barcode:

1) USER digestion and washes

dsDNA barcoding in CAGEs was carried after a dU PCR reaction, USER excision and washing as described above. In the final two washes using TI (pH 8.0), the buffer was supplemented with 200mM MgCl₂. All protocols described in this paper utilized a 5' CGTGATGCTGTACTTACATGTGACdUG-3' PCR handle.

2) Preparation of barcoding plates

Work done in this paper used a 48-barcode format. Plates containing Ligation-1 and Ligation-2 barcodes were prepared in advance by mixing 50 μ L of 100 μ M forward and reverse oligos (table S4, Ligation-1, Ligation-2) at 1:1 ratio. Barcodes were annealed by heating the plates at 95°C in a PCR machine for 5 min and then letting the plate cool at room temperature for 1 hour, followed by incubation on ice until a reaction is carried out. A third plate for PCR-3 barcoding was also prepared in advance, with wells each containing 10 μ M barcoding primers (table S4, PCR-3). All barcoding plates were stored frozen at -20°C.

3) Ligation-1 barcode addition

CAGEs barcode-1 ligation was performed in a 96-well plate. Reaction volumes were scaled based on the volume of CAGEs being barcoded as follows. For every 2 μ L of CAGEs/well, a reaction was carried out in a 10 μ L volume with: 5 μ L of 2x StickTogether DNA Ligase Buffer (NEB, M0318L), 2 μ L of CAGEs in TI with 200mM MgCl₂, 2 μ L of 50 μ M annealed barcode duplex and 1 μ L of T7 DNA ligase. (NEB, M0318L). Before setting up the reaction, CAGEs were pelleted in TI (pH 8.0) with 200mM MgCl₂ and the majority of the buffer was aspirated. The amount of buffer to leave behind is determined by rounding the packed CAGEs volume to a convenient working volume (for example, 170 μ L of packed CAGEs would be aspirated to 200 μ L using another 1.5mL tube loaded with buffer as a visual standard). Packed CAGEs were left on ice for at least 5 min. A fresh Multiplate™ 96-well PCR plate (BioRad, M0318L) was placed on ice and then 1/50th of the total CAGEs volume was transferred to each well using a pipette (in the case of 200 μ L, 4 μ L would be added to each well). Next, a ligation master mix composed out of 2x StickTogether DNA Ligase Buffer (NEB, M0318L) and T7 DNA ligase (NEB, M0318L) was prepared for 50 reactions in the same tube that housed the aliquoted CAGEs and 1/50th of the

volume was added directly to CAGEs in each well without pipetting. Finally, thawed and well mixed ligation-1 DNA barcodes were added to wells housing CAGEs and ligation master mix using a 12-tip multichannel pipette (volume defined above), followed by pipetting with the same tip to mix the reagents within each well. The barcoding plate was spun down, covered with a silicone mat (Millipore Sigma, Z374938-10EA) and incubated on a heat block at 25°C for 30 min.

After incubation, the reaction was stopped, and CAGEs were pooled by adding 100 μ L of TIS (pH 8.0) into each well and then transferring the volume from all wells into a 15mL falcon tube. Each well was washed twice, and the plate was inspected under the inverted microscope to make sure no CAGEs are left behind. CAGEs were transferred to a 1.5mL tube and were washed four times with TI10 (pH 8.0).

4) Overhang generation by dU excision

Next, DNA overhangs for Ligation-2 barcode addition were generated by performing a *USER-based dU excision* as described above. After the reaction, CAGEs were washed twice with TI10 (pH 8.0) and twice using TI (pH 8.0) with 200mM MgCl₂ without the 60°C heating step.

5) Ligation-2 barcode addition

CAGEs Ligation-2 barcode addition was performed in a 96-well plate by repeating the steps described in *Step 3* above.

6) PCR-3 barcode addition

CAGEs PCR-3 barcode addition was performed in a 96-well plate. Reaction volume scales with the volume of CAGEs being barcoded. For 1-2.6 μ L of CAGEs, a standard 10 μ L reaction houses 5 μ L of 2x NEBNext® High-Fidelity 2X PCR Master Mix (NEB, M0543L), 0.4 μ L of 10% Igepal CA-630 (Millipore Sigma, I8896-50ML), 1-2.6 μ L of CAGEs in TI (pH 8.0), 1 μ L of 10 μ M universal PCR primer, targeting the other end of DNA molecule, 1 μ L of 10 μ M PCR barcode (table S4, PCR-3) and water up to 10 μ L. Before setting up the reaction, CAGEs were pelleted, resuspended and distributed over the wells of the barcoding plate as described in *Step 3*. Next, a PCR master mix, composed out of 2x NEBNext® High-Fidelity 2X PCR Master Mix (NEB, M0543L), Igepal CA-630 (Millipore Sigma, I8896-50ML), universal PCR primer and water (if used). was prepared for 50 reactions in the same tube that housed the aliquoted CAGEs and 1/50th of the volume was added directly to CAGEs in each well without pipetting. Finally, thawed and vortexed PCR-3 DNA barcodes were added to wells housing CAGEs and PCR master mix as described in *Step 3*, using a 12-tip multichannel pipette (volume defined above). The reaction plate was spun down, covered with adhesive PCR film (Thermo Fisher, AB0558), incubated on ice for 5 min and followed by a PCR cycling program: 98°C for 45s; 5 cycles of 98°C for 10s, 68°C for 20s, 72°C for 4 min; a

final extension of 72°C for 2min and hold at 25°C.

After the PCR-3 barcode addition, CAGEs were pooled by adding 100 μ L of TIS (pH 8.0) into each well and transferring all the volume into a 15mL falcon tube. Each well was washed twice, and the plate was inspected under the inverted microscope to make sure no CAGEs are left behind. CAGEs were transferred to a 1.5mL tube and were washed three times with TI100 (pH 8.0), three times with TI10 (pH 8.0), and twice with TI (pH 8.0). At this point, CAGEs were split into 1.5mL tubes for DNA purification and further downstream processing.

CAGE dissolution and DNA purification

CAGEs were dissolved in the presence of 0.5M NaOH, 0.5% Triton X-100 and 0.5% SDS in a shaking heat block set to 40°C for 20min in 1.5mL tubes. The solution was neutralized by adding an equimolar amount of HCl and 200mM Tris (pH 8.0) buffer to a final volume of 100 μ L. DNA from solubilized CAGEs was extracted using Qiagen MinElute DNA purification kit (Qiagen, 28004) or AMPure XP magnetic beads (Beckman Coulter, A63881).

Post-purification barcoded DNA amplification

Barcoded and purified DNA was amplified using PCR. A standard 50 μ L reaction was composed out of 25 μ L 2x NEBNext® High-Fidelity 2X PCR Master Mix (NEB, M0543L), 2.5 μ L of 10 μ M P7_rv primer (table S4), 2.5 μ L of 10 μ M FW primer (protocol depended), 20 μ L of purified barcoded DNA. Reactions were performed using a PCR cycling program: 98°C for 30s; 10 cycles of 98°C for 15s, 68°C for 20s, 72°C for 2min; a final extension of 72°C for 2min and hold at 4°C unless stated otherwise.

Fragmentation-based Illumina library construction

To construct Illumina cDNA sequencing libraries, 50 ng of amplified cDNA was fragmented, blunted and dA-tailed by mixing it with 3.5 μ L of NEBNext Ultra II FS Reaction buffer (NEB), 1 μ L of fully mixed NEBNext Ultra II FS Enzyme mix (NEB) and nuclease free water to a final volume of 17.5 μ L. The reaction mix was vortexed for 10 s, spun down and placed into a preheated PCR machine for incubation at 37°C for 10 min; 65°C for 30 min; hold 4°C (lid at 75°C). Fragmented DNA was purified using 0.8X AMPure XP beads (Beckman Coulter, A63881) and eluted into 16 μ L of water. Single-stranded DNA adapter, containing Illumina's TruSeq read 1 (Adp. TruSeq R1 FW and RV) sequences were annealed together at the final concentration of 1.5 μ M by mixing 1.5 μ L of forward ligation primer and reverse ligation primer (each at 100 μ M) with 97 μ L of nuclease-free water and heating the solution to 95°C for 2 min and cooling it at room temperature for 5 min. 16 μ L of the purified fragmented DNA was mixed with 1.25 μ L of 1.5 μ M annealed adapters, 1.5 μ L of 10X T4 DNA ligase buffer, 15 μ L of NEBNext Ultra II Ligation Master Mix

(NEB, E7805S kit) and 0.5 μ L of NEBNext Ultra II Ligation Master Enhancer (NEB, E7805S kit). The ligation reaction was carried out by incubating in the thermocycler for 15 min at 20°C (lid at 30°C). After ligation, the reaction mix was diluted to 100 μ L and was purified using AMPure XP beads (Beckman Coulter, A63881) with 0.8X volume ratio and eluted into 40 μ L nuclease-free water. 20 μ L of this DNA was then used in the *Library indexing PCR* while the rest was frozen down at -20°C.

Library indexing PCR

Indexing PCR was carried out by mixing purified DNA with 25 μ L of 2x NEBNext® High-Fidelity 2X PCR Master Mix (NEB, M0543L) and 2.5 μ L of 20 μ M P7_rv, 2.5 μ L of 20 μ M indexing primer that houses an 8bp library index (table S4, indexing) and water to 50 μ L. PCR cycling program: 98°C, 2 min; 3 cycles of 98°C, 20 s; 55°C, 30 s; 72°C, 20 s and 9 cycles 98°C, 20 s; 65°C, 30 s; 72°C, 20 s; with a 1 min final extension at 72°C, and 4°C, hold at the end.

Plasmid DNA sequencing

CAGEs containing plasmid DNA were generated as described in *Basic CAGE synthesis protocol*, except the core solution contained 1ng/ μ L of plasmid DNA. After polymerization, CAGEs were washed three times with TI100 (pH 8.0), twice with TI10 (pH 8.0) and twice with TI (pH 8.0). Next, 14 reactions of CAGE dU PCR were performed in 50 μ L volumes as described in *CAGE dU PCR* using 2.5 μ L of 20 μ M GFP/RFP FW primer (table S4), 2.5 μ L of 20 μ M dU barcoded CMV [1-14] RV primer (table S4, PCR) and 27.5 μ L of packed CAGEs. The reaction was carried out by first incubating the tubes on ice for 5 min, followed by cycling: 98°C for 30s; 14 cycles of 98°C for 15s, 68°C for 20s, 72°C for 4 min; a final extension of 72°C for 2min and hold at 25°C. After PCR, CAGEs were washed as described in *Basic Post PCR wash steps* and pooled into a single tube. The CAGEs were then subjected to a second PCR reaction performed in 50 μ L volumes as described above, except 2.5 μ L of 20 μ M GFP/RFP FW, 2.5 μ L of labeled dU universal RV primer and 30 μ L of packed CAGEs with no extra water were used with the same cycling conditions for 8 cycles. After the DNA amplification, CAGEs were washed as described in *Post dU PCR wash steps* and were subjected to split-and-pool barcoding as described above.

Following split-and-pool barcoding, CAGEs were split into 16x 1.5mL tubes and were dissolved as described above. DNA was purified using MinElute PCR purification kit (Qiagen, 28004) and barcoded DNA was eluted into 20 μ L of the elution buffer. Barcoded and purified DNA was amplified using PCR as described in *Post-purification barcoded DNA amplification*, using 2.5 μ L of TruSeq_R1_GFP/TruSeq_R1_RFP or TruSeq_R1x_GFP/TruSeq_R1x_RFP primer mix (table S4) to increase library diversity. After the reaction, amplified

barcoded DNA was purified using 0.8X AMPure XP beads (Beckman Coulter, A63881) and was eluted into a 15 μ L Elution Buffer. DNA concentration was measured using a Qubit dsDNA HS Assay (Thermo Fisher, Q32854). Final sequencing library index PCR was carried out using 20 ng of purified DNA product 2.5 μ L of 20 μ M TruR1_x_p5 primer as described in *Library indexing PCR* step. Following final library indexing, DNA was double-side purified using 0.6x-0.8x SPRIselect magnetic beads (Beckman Coulter, B23318) and eluted into 15 μ L of water.

CAGE reverse-transcription reaction

Reverse-transcription reaction was set on ice and performed in PCR strip tubes. A standard 100 μ L reaction was composed out of 60.5 μ L of packed CAGEs, 2.5 μ L of 10 μ M RT poly(T) primer housing a 10nt UMI and 6nt barcode (table S4, RT), 3 μ L of 10% Igepal CA-630 (Millipore Sigma, I8896-50ML), 20 μ L of 5X SMARTseq3 RT buffer (100mM Tris pH 8.3, 150mM NaCl, 12.5mM MgCl₂, 5mM GTP, 40mM DTT), 5 μ L of 10mM dNTPs (NEB, N0447L), 2.5 μ L of 40U/ μ L Recombinant RNase Inhibitor (Takara Bio, 2313A), 1.5 μ L of 50 μ M TSO, 5 μ L of Maxima H- reverse transcriptase (Thermo Fisher, EP0753), unless stated otherwise. Reaction volumes were scaled according to the amount of CAGEs used and number of reactions performed. To set up the reaction, CAGEs, RT primer and Igepal CA-630 (Millipore Sigma, I8896-50ML) were mixed on ice in a PCR tube, with each housing a different RT barcode. CAGEs were incubated on ice for 10 min to allow the RT primer to diffuse in, followed by heating at 70°C for 5 min and cooling down on ice. Next, RT master mix, composed out of the remaining reagents was added to each tube, ensuring a thorough mixing of CAGEs with pipetting. Tubes were incubated on ice for 5 min and the reaction was carried out at 42°C for 90 min, followed by cycling between 50°C and 42°C incubation of 2 min for 15 cycles. Reactions were left at 20°C overnight and processed the next day.

Biotinylated DNA pulldown

Biotin labeled DNA pulldown was performed using Pierce Streptavidin Magnetic Beads (Thermo Fisher, 88816). 25 μ L of magnetic beads were washed three times using 30 μ L of 5mM Tris (pH 7.4) with 0.5mM EDTA and 1M NaCl while placed on a magnetic rack. After washing, beads were resuspended in 60 μ L of 10mM Tris (pH 7.4) with 1mM EDTA and 2M NaCl and mixed with DNA sample diluted to 60 μ L using water. Beads were incubated at room temperature for 30 min, placed on a magnetic rack. Supernatant was transferred to a new tube and was used for purifying non-biotinylated DNA with AMPure XP beads (Beckman Coulter, A63881). Remaining beads with bound DNA were washed three times with 50 μ L of 5mM Tris (pH 7.4) with 0.5mM EDTA and 1M NaCl and biotinylated DNA was eluted by resuspending 20 μ L of

95% formamide with 10mM EDTA (pH 8.2) and heating at 65°C for 5 min. Eluted DNA was purified using 1X AMPure XP beads (Beckman Coulter, A63881) and eluted into 20 μ L of water.

CAGE single-cell gDNA amplicon-seq

K562-RFP and L1210-GFP (kindly donated by Sean McGeary and Yuyang Chen) cell suspension was mixed at a ratio of 1:1. CAGEs housing single-cells were prepared as described in *CAGEs synthesis for cell encapsulation*, except core solution contained cell suspension at a final concentration of 1.2e6 cells/mL in 1x DPBS. Resulting CAGEs were resuspended in ice-cold 3x SSCI (pH 5.9). The subsequent steps were performed on ice: CAGEs were washed 6 times with 3x SSCI, once with TEIS (pH 7.5), three times with 3x SSCI and three times with H₂O.

While the paper analyzes only gDNA amplicons, the experiment reported in the paper was simultaneously processed for joint amplicon-seq and scRNA-seq. To this end, two 50 μ L reverse-transcription reactions were set on ice in two PCR strip tubes as described in *CAGE reverse-transcription reaction* and were washed as described in *post reverse-transcription wash steps*.

The CAGEs containing gDNA and cDNA were then processed to amplify target gDNA regions by carrying out 16 cycles of CAGE gDNA amplicon PCR. The PCR reactions were performed in 50 μ L volume as described in *CAGE dU PCR* using 5 μ L of 10 μ M RFP_R1 FW primer, 5 μ L of 10 μ M GFP_R1 FW primer, 5 μ L of 10 μ M universal dU CMV RV primer (table S4), and 20 μ L of packed CAGEs. The reaction was carried out by first incubating the tubes on ice for 5 min, followed by a PCR cycling program: 98°C for 30s; 16 cycles of 98°C for 15s, 70°C for 20s, 72°C for 2 min; a final extension of 72°C for 2min and hold at 25°C. After PCR, CAGEs were washed once with TIS (pH 8.0), three times with TI100 (pH 8.0) and three times with TI10 (pH 8.0). A second PCR reaction was then performed in a 50 μ L volume as described in *CAGE dU PCR* using 2.5 μ L of 10 μ M Read1_Biotin FW primer (table S4), 2.5 μ L of 10 μ M universal dU CMV RV primer, 2.5 μ L of 10 μ M universal dU RV primer, 2.5 μ L of 10 μ M of TSO FW primer with 25 μ L of packed CAGEs. After the DNA amplification, CAGEs were washed as described in *Post dU PCR wash steps* and were subjected to split-and-pool barcoding as described above, except using Read1_Biotin FW and TSO FW primer used in the PCR-3 barcode additions.

Following split-and-pool barcoding, CAGEs were split into multiple 1.5mL tubes and were dissolved as described above. DNA was purified using MinElute PCR purification kit and barcoded cDNA was eluted into 20 μ L of the elution buffer. Barcoded and purified DNA was amplified using PCR as described in *Post-purification barcoded DNA amplification*, using 2.5 μ L of TSO_fw and Read1_Biotin FW primers. After the

reaction, amplified barcoded DNA was purified using 0.8X AMPure XP beads (Beckman Coulter, A63881) and was eluted into a 20 μ L Elution Buffer. Next biotinylated DNA corresponding to gDNA amplicons was purified as described in *Biotinylated DNA pulldown* above. Final amplicon library amplification and indexing PCR was carried out using 20 μ L of the purified DNA as in *Library indexing PCR* with 2.5 μ L of 20 μ M truR1_x_p5 primer that houses an 8bp library index. Following final library indexing, DNA was purified using 0.8x SPRIselect magnetic beads (Beckman Coulter, B23318) and eluted into 15 μ L.

Imaging lysed cells and cDNA libraries in CAGEs

After lysis, CAGEs were resuspended in 3X SSCI buffer with 1X SYBR Safe DNA Gel Stain (Thermo Fisher, S33102). 5 μ L of packed CAGEs were transferred to a hemocytometer or a well of 96-well glass bottom plate and were topped with 3X SSCI buffer with 1X stain and imaged on a wide field or confocal microscope.

CAGE single cell RNA-seq

CAGEs housing single-cells were prepared as described in *CAGE synthesis protocol for cell encapsulation* with a final concentration of 1–1.4M cells/mL. Experiment represented in Fig. 3C and fig. S3B used 1M cells/mL; Figs. 2D, 3, D and F, and figs. S3, A, C to H, and S4F used 1.2M; Fig. 3, E, and G to K, and fig. S4 used 1.4M. The resulting CAGE pellets were resuspended in 1 mL room temperature IFLB and were washed twice with 300 g spins for 1 min. Next, core dextran was removed as described in *Dextran digestion reaction*.

Following this, all steps were performed on ice to lyse the cells. First, CAGEs were washed three times with an ice-cold 3X SSCI buffer with 800 g spins for 1 min. Cells were lysed by washing with ice-cold 7.5pH TEIS with 5 min incubations on ice. After lysis, CAGEs were imaged (check above), washed three times with 3X SSCI and three times with HI10, packed at 800 g for 1 min and stored on ice.

Reverse-transcription reaction was set on ice and performed in PCR strip tubes as described in *CAGE reverse-transcription reaction* followed by *post reverse-transcription wash steps*.

Next, CAGE dU PCR was performed to carry out cDNA amplification as described in *CAGE dU PCR*, using, 5 μ L of 10 μ M TSO FW primer, 5 μ L of 10 μ M labeled dU universal RV primer (table S4), followed by *post dU PCR wash steps*. CAGEs were imaged on a confocal microscope to inspect the cDNA signal (from labeled dU primer). Washed CAGEs were subjected to split-and-pool barcoding as described above with the TSO FW primer used in the PCR-3 barcode additions.

Following split-and-pool barcoding, CAGEs were split into multiple 1.5mL tubes and were dissolved as described above. DNA was purified using MinElute PCR purification kit and

barcoded cDNA was eluted into 20 μ L of the elution buffer. Barcoded and purified DNA was amplified using PCR as described in *Post-purification barcoded DNA amplification*, using 2.5 μ L of TSO FW primer. After the reaction, amplified barcoded cDNA was purified using 0.8X AMPure XP beads (Beckman Coulter, A63881) and was eluted into a 15 μ L Elution Buffer. DNA concentration was measured using QuBit fluorometer and cDNA fragment length profile was assessed using Agilent Tapestation D5000 kits (Agilent, 5067).

Illumina cDNA sequencing libraries were constructed as described in *Fragmentation-based Illumina library construction*

Final sequencing library indexing PCR was carried out as described in *Library index PCR* with 2.5 μ L of 20 μ M truR1_x_p5 primer that houses an 8bp library index.

Following final library indexing, DNA was double-side purified using 0.6x-0.8x SPRIselect (Beckman Coulter, B23318) magnetic beads and eluted into 15 μ L

Tn5 transposase loading with barcoded DNA primers

Unloaded Tn5 transposase (Diagenode, C01070010-20) was loaded with adapter sequences carrying a barcode (table S4, Tn5). First, the oligos were resuspended to 100 μ M in 40mM Tris (pH 8.0) with 50mM NaCl, universal Tn5_rev oligo was mixed with oligo Tn5_B0_x or Tn5_oligoB at 1:1 ratio and annealed by in a thermocycler: 95°C, 5min; cool to 65°C (-0.1°C/s); 65°C, 5min; cool to 4°C (-0.1°C/s). Annealed adapters were mixed at ratio 1:1 and 10 μ L were gently added to 10 μ L of unloaded Tn5. The solution was incubated at 23°C for 30 min in a thermocycler and 10 μ L of 99% glycerol was added to enable the sorting of loaded Tn5 at -20°C.

In CAGE gDNA tagmentation

To perform gDNA tagmentation, CAGEs were washed with DPBS with 0.1% (w/v) F127 and 0.1% L31 (w/v). For a 25 μ L reaction, 9 μ L of package CAGEs were mixed with 12.5 μ L of 2x TD buffer (20 mM Tris 10 mM MgCl₂; 20% (vol/vol) DMF, pH 7.6), 1 μ L of loaded Tn5, 2.5 μ L of 10X Lysis additive (5% Tween20, 0.05% Digitonin in DPBS). The solution was incubated on ice for 5 min, followed by incubation at 37°C for 30 min. After the reaction, CAGEs were washed three times with TEIS (pH 7.5) and then Proteinase K treated in 200 μ L of TEIS (pH 7.5) with 10 μ L of Proteinase K (NEB, P8107S) for 15 min at 55°C. After the reaction, CAGEs were washed three times with TI100 (pH 7.5) and twice with TI (pH 7.5).

CAGE ATAC-seq library amplification and normalization

ATAC-seq libraries purified from CAGEs were amplified using qPCR. First, a PCR reaction was set up using 25 μ L 2x NEBNext® High-Fidelity 2X PCR Master Mix (NEB, M0543L), 2.5 μ L of 20 μ M P7_rv primer, 2.5 μ L of 20 μ M NexteraR1_x_p5

FW primer and 20 μ L of purified barcoded DNA. Reactions were performed using a PCR cycling program: 98°C for 50s; 6 cycles of 98°C for 10s, 68°C for 20s, 72°C for 1min; a final extension of 72°C for 4min and hold at 4°C and the reaction was then transferred on ice. Next, 5 μ L of partially amplified DNA was mixed with 0.25 μ L of 20 μ M of each primer, 0.1 μ L of 100X SYBRgreen, 5 μ L of 2x NEBNext® High-Fidelity 2X PCR Master Mix (NEB, M0543L) and 4.4 μ L of water. The reaction was carried out on a qPCR machine using the same PCR program as for the initial 6 cycles, except imaging the samples after every extension step. RFU vs cycle plot was used to determine the number of cycles needed to reach $\frac{1}{3}$ of maximum intensity. This number of cycles was used to amplify the remaining 45 μ L of partially amplified DNA. After the reaction, DNA was purified using double sided size AMPure XP (Beckman Coulter, A63881) size selection with a bead ratio of 0.5X-1.3X. Final library DNA was eluted to 15 μ L of EB, quantified using Qubit and qPCR before sequencing.

CAGE scATAC-seq

Cell encapsulation

CAGEs housing single-cells were prepared as described in *CAGE synthesis for cell encapsulation in media* at a final concentration of 1.2e6 cells/mL, using either a 1:1 mixture of K562 and L1210 (fig. S3, A, E, and G) or using a 2:3:3 mixture of K562, L929 and L1210 (Fig. 3D and fig. S3, C, D, and F) at a final concentration of 1.2e6 cells. The resulting CAGE pellets in both cases were resuspended at room temperature in 1 mL IFLB, incubated for 1 min and then washed twice with pelleting by centrifugation at 300 g, followed by a *Dextran digestion reaction*. Following this, CAGEs wash steps were performed on ice.

Base protocol on K562 and L1210 cells (fig. S3, A, E, and G)

The CAGEs were distributed into five wells for gDNA tagmentation. Reactions were performed in 80 μ L reactions as described in *In CAGE gDNA tagmentation*, with each well receiving a different Tn5 barcode. Next, *CAGE dU PCR* was performed as described before using 2.5 μ L of 20 μ M Nextera R1 FW primer, 2.5 μ L of 20 μ M dU universal RV primer with 25 μ L of packed CAGEs in 50 μ L reactions, followed by *post dU PCR wash steps* (see above section). The resulting tagmented and amplified DNA was barcoded by using the split-and-pool approach as described above, using Nextera R1 FW primer in the PCR-3 reaction.

Following split-and-pool barcoding, CAGEs were dissolved in 1.5mL tubes as described above. DNA was purified using MinElute PCR purification kit and barcoded DNA was eluted into 20 μ L of the elution buffer.

Next, two indexing and CAGE ATAC-seq library amplification and normalization reactions were performed using two distinct NexteraR1_x_p5 FW primers. 9 additional PCR

cycles were added for both libraries. Final libraries were pooled and quantified using Qubit and qPCR before sequencing.

CAGE ATAC-seq of K562, L929 and L1210 cells

For this experiment (Fig. 3D and fig. S3, C, D, and F), the tagmentation reaction is similar to the base protocol, but the CAGEs were split into six wells and the reactions were performed in half the volume (40 μ L reactions, rather than 80 μ L). Then, after tagmentation, CAGEs were washed six times with an ice-cold 3X SSCI buffer with 800 g spins for 1 min and three times with H110, packed at 800 g for 1 min and stored on ice.

In this second experiment, the CAGEs were also then subjected to a reverse-transcription reaction. To this end, the reaction was prepared on ice and performed in 6 PCR strip tubes as described above in *CAGE reverse-transcription reaction* followed by *post reverse-transcription wash steps*. Next, the CAGEs were split into four wells for a *CAGE dU PCR* reaction, as described before using using 2.5 μ L of 20 μ M Read1_Biotin FW, 2.5 μ L of 20 μ M TSO FW (table S4), 5 μ L of 10 μ M dU universal RV primer with 25 μ L of packed CAGEs in 50 μ L reactions, followed by post dU PCR wash steps. Resulting tagmented and amplified DNA was barcoded by using the split-and-pool approach as described previously with the Read1_Biotin FW and TSO FW primer used in the PCR-3 barcode additions.

Following split-and-pool barcoding, CAGEs were dissolved into 4 1.5mL tubes as described above. DNA was purified using MinElute PCR purification kit and barcoded DNA was eluted into 20 μ L of the elution buffer. Barcoded and purified DNA was amplified using PCR as described in *Post-purification barcoded DNA amplification*, using 1.25 μ L of 20 μ M Read1_Biotin FW FW and 1.25 μ L of 20 μ M TSO FW primer. DNA was purified using AMPure XP beads (Beckman Coulter, A63881) with 1.8X ratio and eluted into 20 μ L. Next, biotin pulldown was performed to purify amplified and barcoded biotinylated gDNA fragments from non-biotinylated cDNA libraries as described in *Biotinylated DNA pulldown*.

Next, 4 indexing and CAGE ATAC-seq library amplification and normalization reactions were performed using 4 distinct NexteraR1_x_p5 FW primers. 7, 9, 9 and 11 additional PCR cycles were added for each library respectively. Final libraries were pooled and quantified using Qubit and qPCR before sequencing.

Fixation of cells in suspension

Cell suspensions were centrifuged at 400 g for 3 min and the pellet was resuspended in 1mL of 4% (w/v) PFA in 1X DPBS. The solution was incubated at room temperature for 15 min, followed by 3 washes with 1X DPBS + 0.1% BSA. The fixed cells were encapsulated into CAGEs by adding the cells to

~1.4e6 cells/mL into the core solution.

Reverse cross-linking cells in CAGEs

Fixed cells were reverse cross-linked in CAGEs by washing the CAGEs twice with 3x SSCI + 0.01U/ μ L RiboLock RNase Inhibitor (Thermo Fisher, EO0381). Next, CAGEs were washed three times with 100mM Tris (pH 7.5) with 10mM EDTA, 0.1% Igepal CA-630 (Millipore Sigma, I8896-50ML), 0.1% SDS. Finally, CAGEs were resuspended in the same solution to a volume of 1 mL and 50 μ L of proteinase K (NEB, P8107S) was added followed by incubation at 56°C for 30 min. After the reverse crosslinking, CAGEs were washed three times with 3X SSCI, three times with H110 and were used in a reverse transcription reaction as described in *CAGE single cell RNA-seq*.

Freezing CAGEs

CAGEs were frozen at -80°C in the presence of 10% DMSO. CAGEs containing cell lysates were frozen in a 3X SSCI buffer with 0.05U/ μ L of RNase inhibitor (NEB, M0314S). CAGEs containing DNA only were frozen in TI (pH 8.0) with 10% DMSO.

Preparation of PBMCs

Human Peripheral Blood Mononuclear Cells were obtained from Lonza (CC-2704). The cryo-preserved cell vials were stored in liquid Nitrogen until use, and then thawed in a 37°C metal bead bath for 7 min. The thawed cell suspension was transferred into a 15 mL falcon tube. 1 mL of 37°C pre-warmed DMEM (Thermo Fisher, 10566016) with 10% FBS (ATCC, 30-2020) and 1x Penicillin/Streptomycin (Thermo Fisher, 10378016) was slowly added into the original cryo-preservation tube. This media was then added one drop at a time into the falcon tube. Additional 8 mL of pre-warmed media were added slowly onto the cell suspension. Cells were resuspended gently and centrifuged at room temperature at 300 g for 4 min. All of the media was aspirated from the cell pellet and cells were suspended in 5 mL of fresh, 37°C appropriate media with 10% FBS and 1x Penicillin/Streptomycin and transferred into 25 cm² culture flask with a filter cap (Corning). Cells were cultured in a 37°C incubator with 5% CO₂. In the case of PBMCs, cells were washed one more time with IMDM + 10% FBS +1x Pen/Strep and were passed through a 40 μ m strainer. Following this, cells were washed with ice cold plain IMDM with 0.5% BSA. PBMC were encapsulated at a final concentration of 1.4e6 cells/mL in the core mix and then processed for scRNA-seq as described above.

Mammalian cell line culture

K562, L1210, L929, Jurkat and HEK293t cell lines were cultured and used in experiments described in this work.

K562 and L1210 cells were cultured in IMDM

(ThermoFisher, 10566016) + 10% FBS (ATCC, 30-2020) + 1x penicillin/streptomycin (ThermoFisher, 10378016) in T25 flasks (Corning).

L929 cells were cultured in EMEM (ATCC, 30-2003) + 10% horse serum (ThermoFisher, A5669502) + 1x penicillin/streptomycin (ThermoFisher, 10378016) in T25 flasks (Corning).

Jurkat and HEK293T cells were cultured in RPMI1640 (Thermo Fisher, 61870036) + 10% FBS (ATCC, 30-2020) + 1x penicillin/streptomycin (Thermo Fisher, 10378016) in T25 flasks (Corning).

Cell lines were cultured until they reached ~80–90% confluency. For passaging, suspension cells (K562, L1210, Jurkat) were resuspended gently and collected into a 15 mL tube. 5 mL of 37°C pre-warmed complete media was used to wash the flask and was added into the same 15 mL tube. Adherent cells (HEK293T, L929) cells were passaged by aspirating the used media and washing 2 times with 2 mL of 37°C pre-warmed 1X DPBS buffer. Cells were detached by adding 1 mL of TrypLE reagent (Thermo Fisher Scientific, 12605010), waiting 10s and quickly aspirating all of the TrypLE. Cells were incubated in 37°C for 2 min and were resuspended in 5 mL of 37°C pre-warmed complete media. Cells were centrifuged at 300 g for 4 min, followed by media aspiration and resuspension in 5 mL of fresh pre-warmed complete media. Cells were counted using an automatic cell counter and 500,000 cells were seeded into 5 mL of 37°C pre-warmed complete media in a 25cm² flask. For encapsulation experiments, cells were washed twice with ice cold DPBS with 0.1% BSA or room temperature IMDM with 0.1% BSA and were encapsulated at a final concentration of 1–1.4e6 cells/mL in core mix and then processed immediately for scDNA-seq/scRNA-seq/scATAC-seq as described previously or grown into colonies as described below.

iPSC culturing

Induced pluripotent stem cells were gifted by the lab of Olivier Pourquie (Harvard). Cells were cultured in a 6-well plate on a layer of Matrigel (Corning) in mTeSR media (STEMCELL Technologies, 100-0276). To passage the culture, cells were detached by removing media, rinsing twice with DPBS, adding 1mL of Accutase (STEMCELL Technologies, 07920) and incubating at 37°C for 7 min. Next, cells were resuspended in 5mL of plain DMEM and centrifuged in a swinging bucket centrifuge for 4 min at 300 g. Supernatant was aspirated and the cell pellet was resuspended in 1mL of DMEM. 2mL of Matrigel in DMEM solution was added to a single 6-well plate well. The plate was incubated in 37°C for 30 min. Any remaining solution was aspirated and 500K of iPSC cells in DMEM were seeded into 2mL of mTeSR (STEMCELL Technologies, 100-0276) with 10 μ g/mL Y-27632 (Dihydrochloride) (STEMCELL Technologies, 72302). After a day of culturing, media was changed with fresh mTeSR without Y-27632.

Media changes were performed daily for 3–4 days, until cells reached 80–90% confluency.

Cell-line culturing in CAGEs

CAGEs housing single cells were prepared as described in *CAGE synthesis Modification 4* with the final cell concentration of 1.4×10^6 cells/mL. The resulting hydrogel CAGE pellet was resuspended in 37°C temperature IFLB followed by dextran core digestion, two washes with IFLB and a wash with complete media without Pluronic. Packed CAGEs were transferred to wells of 24-well plate and cultured in 1mL of complete media without Pluronic as described in mammalian cell line culturing part above.

To change media, CAGEs were resuspended in the growth medium and were transferred to a reversible 37 μ m strainer (STEMCELL technologies, 27215). CAGEs were washed in the strainer with 3mL of complete growth medium. Next, the strainer was inverted, and the CAGEs were eluted back into a well of a 24-well plate using 1mL of complete growth medium. Any manipulation of CAGEs was done in complete growth media with 0.1% (w/v) F127 and 0.1% (w/v) L31.

iPSC culturing in CAGEs

CAGEs housing single iPSCs were prepared as described in *CAGE synthesis Modification 4* with the final concentration of 1.4×10^6 cells/mL. CAGE clean-up and dextran removal were done as described above. After clean-up, CAGEs were washed twice with mTeSR 3D + 0.1% F127, 0.1% L31 + 0.1% BSA. Packed CAGEs were transferred to wells of 24-well plate and cultured in 1mL of mTeSR 3D media (STEMCELL Technologies, #03950) with 1X CloneR™2 cloning supplement (STEMCELL Technologies, 100-0691). To change media, CAGEs were resuspended in the surrounding growth medium and were transferred to a reversible 37 μ m strainer (STEMCELL Technologies, 27215). CAGEs were washed in the strainer with 3mL of IFL wash media. Next, the strainer was inverted and the CAGEs were eluted back into a well of a 24-well plate using 1mL of mTeSR 3D media. Any manipulation of CAGEs was done in complete growth media with 0.1% (w/v) F127 and 0.1% (w/v) L31.

mHPC culturing in CAGEs

LSK mHPCs were gifted by the lab of Fernando Camargo (Harvard). CAGEs housing single mHPCs were generated using the protocol described in *CAGE synthesis Modification 4* with a final concentration of 0.8×10^6 cells/mL in the core mix and no BSA. CAGE clean-up and dextran removal were done as described above, except all washes were performed in plain IFL wash media. After clean-up, CAGEs were washed twice F12 media (Thermo Fisher, 11765054) supplemented with 10mM HEPES, 1X P/S/G, 1 mg/ml PVA, 100 ng/ml TPO (Thermo Fisher, 300-18-10UG), 10 ng/ml SCF (Thermo Fisher,

300-07-10UG), 1x ITSX (Thermo Fisher, 51500056), 0.1% F127, 0.1% L31. Packed CAGEs were transferred to wells of 24-well plate and cultured in 1mL F12 expansion media. Media was exchanged every two days by collecting all the CAGEs into a 1.5mL tube, spinning down at 300 g for 1 min and resuspending in fresh media.

Cell division time measurements

CAGEs housing single K562 or L1210 cells were prepared in two separate 20 min encapsulations as described in *Cell-line culturing in CAGEs* using a final cell concentration of 1.2×10^6 cells/mL.

200 μ L of washed and packed CAGEs were transferred to a 6-well plate housing a free suspension of K562 or L1210 cells at a concentration of 10,000 cells/well in 2 mL of full IMDM as described in *Mammalian cell line culture* without Pluronics. The plate was transferred to a A1R Nikon Confocal laser-scanning microscope with Tokai Hit Live-Cell Microscope Incubation Systems set to 37°C with 5% CO₂. Time lapse microscopy was used to acquire images every 30 min. The resulting time series were manually examined to identify consecutive cell divisions of single cells within CAGEs and in the surrounding cell suspension, and the time interval was recorded. For each single cell clone, up to three divisions (two intervals) were used.

Transcriptomic comparison of cells grown in CAGEs and wells

CAGEs housing iPSCs, K562 and L929 cells were prepared in two separate 20 min encapsulations as described in *Cell-line culturing in CAGEs* and *iPSC culturing in CAGEs* using a final cell concentration of 2×10^6 cells/mL.

100 μ L of washed and packed CAGEs or 20,000 cells were transferred to a 24-well plate and cultured for 3 (K562 and L929) or 4 (iPSCs) days in 1mL of complete media without Pluronics as described in *iPSC culturing* or *mammalian cell line culturing* part above. No media changes were performed for cell lines and, in the case of iPSCs, mTeSR 3D media feed supplement (STEMCELL Technologies, #03950) was added every day.

Cells in wells and CAGEs were analyzed using bulk RNA-seq. First, all samples were washed three times with ice-cold DPBS. 40,000 cells or 40.5 μ L of CAGEs were lysed using 0.45% Igepal CA-630 on ice in a final volume of 63.5 μ L. RT step was carried out as described in the *CAGE reverse-transcription reaction* methods section using CAGEs or cell suspension as material input and R2-polyT as RT primer. After the RT reaction, 2 μ L of 10M NaOH were added to all reactions followed by 20 min incubation at 37°C to dissolve the CAGEs [NaOH was also added to samples with no CAGEs]. cDNA was purified into 20 μ L of water using 0.8X AMPure XP beads (Beckman Coulter, A63881).

Next cDNA was amplified using PCR as described in *Post-purification barcoded DNA amplification*, using 2.5 μ L of 10 μ M TSO FW primer and 2.5 μ L of 10 μ M R2 FW primer. After the reaction, amplified barcoded cDNA was purified using 0.8X AMPure XP beads (Beckman Coulter, A63881) and was eluted into a 15 μ L Elution Buffer. DNA concentration was measured using QuBit fluorometer and cDNA fragment length profile was assessed using Agilent TapeStation D5000 kits (Agilent, 5067).

Illumina sequencing libraries were constructed using 50ng cDNA input as described in *Fragmentation-based Illumina library construction*

Final sequencing library indexing PCR was carried out as described previously with 2.5 μ L of 20 μ M truR1_x_p5 primer that houses an 8bp library index and 2.5 μ L of 20 μ M p7_B3_X to index the other side of the molecules as no barcoding was carried out.

Following final library indexing, DNA was double-side purified using 0.6x-0.8x SPRIselect (Beckman Coulter, B23318) magnetic beads and eluted into 15 μ L. Sequencing libraries were pooled and sequenced on NextSeq 1000 P1 100 cycle kit.

Fixing and staining clones in CAGEs

Colonies in CAGEs were fixed by resuspending packed CAGEs in 4% (w/v) PFA solution in 1X DPBS and incubating at room temperature for 20 min. CAGEs were washed twice and stored in ice-cold DPBS with 0.1% (w/v) BSA, 0.1% (w/v) F127, 0.1% (w/v) L31. Cells were permeabilized by adding Triton X-100 to a final concentration of 0.1% and incubating for 10 min. CAGEs were washed twice and stored in a fridge. For iPSC colonies, immunostaining was performed against SOX2 by adding 2 μ L of Alexa Fluor® 488 anti-SOX2 Antibody (BioLegend, 656109) (0.5mg/mL) to 600 μ L of CAGEs solution in DPBS-BFL buffer (1X DPBS with 0.1% (w/v) BSA, 0.1% (w/v) F127 and 0.1% (w/v) L31). As a control, 2 μ L of 0.5mg/mL FITC anti-mouse CD45.1 Antibody (BioLegend, 110705) was used. CAGEs were stained on ice for 15 min, followed by 3 washes with DPBS-BFL buffer, with 5 min incubation on ice between washes. Stained CAGEs were imaged on a confocal microscope.

Treatment with HDAC and DNMT inhibitors

Survival curves for DNMT inhibitors [Decitabine (Dec) and 5-aza-cytidine (Aza)], and the HDAC inhibitor Vorinostat (Vor) (MedChemExpress, HY-A0004, HY-10586, HY-10221) were generated by seeding 15,000 (L1210) and 20,000 (K562) into 96-well plates and growing the cells in 100 μ L of IMDM + 10% FBS, 1X Pen/Strep in the presence of 0, 0.001, 0.003, 0.01, 0.03, 0.1, 0.3, 1, 3, 10 μ M of each drug. Cells were passaged every two days by transferring 15 μ L of cells into a fresh 85 μ L of media loaded with the same concentration of drugs. Cell viability and number was evaluated every two days using a

confocal microscope by mixing 7.5 μ L of cell culture with 42.5 μ L of DPBS with 1X DAPI stain.

CAGE clonal expansion and inC-RNA-seq

Cells were pre-treated with drugs by seeding 150,000 (L1210) and 200,000 (K562) cells into 6-well plate wells (Corning). Cells were grown in 2mL of IMDM with 10% FBS, 1X Pen/Strep in the presence of 0.05 μ M Dec, 0.7 μ M Aza, 0.4 μ M Vor (MedChemExpress) or no drugs for 48 hours. Next, cells were collected and K562 and L1210 cell-lines were pooled based on the drug condition at a 1:1 ratio. Cell suspensions were encapsulated into CAGEs as described in *Cell-line culturing in CAGEs* using a final cell concentration of 1.4e6 cells/mL. Each drug condition was encapsulated for 2 hours, collecting the droplets into four 30 min fractions. One fraction of each condition was sampled instantly by lysing the cells inside the CAGEs - CAGEs were washed three times with ice-cold 3X SSCI buffer with 800 g spins for 1 min. Cells were lysed by washing twice with ice-cold TEIS (7.5pH) with 5 min incubations on ice. After cell lysis, CAGEs were frozen down at -80°C in 3X SSCI with 10% DMSO and 0.05U/ μ L Murine RNase Inhibitor (NEB, M0314S). Remaining fractions of packed hydrogel CAGEs were seeded into a different 24-well plate well into 1mL of IMDM with 10% FBS and 1x Pen/Strep, 0.1% (w/v) F127 and 0.1% (w/v) L31 with appropriate drug. 10,000 K562 and L1210 cells from appropriate drug conditions were seeded outside the CAGEs for the first two days. Every two days, media was exchanged by collecting the CAGEs using an reversible a 37 μ m strainer (STEMCELL Technologies) and eluting into fresh IMDM with 10% FBS and 1x Pen/Strep, 0.1% (w/v) F127 and 0.1% (w/v) L31 with the appropriate drug as described before in *Cell-line culturing in CAGEs*. Every 2 days, one well of each condition was sampled for inC-RNA-seq by lysing the cells and freezing at -80°C as described above.

After all the time points were collected, CAGEs were thawed, washed once with 3X SSCI, once with TI10 (pH 7.5), once with TI (pH 7.5). Next, CAGEs were treated with Turbo DNase I (Thermo Fisher) by adding 8 μ L of 2U/ μ L enzyme to 300 μ L of CAGEs in a 1X reaction buffer and incubating at 37°C for 10 min. CAGEs were then washed once with 3X SSCI and TEIS (pH 7.5) and resuspended in 500 μ L of TEIS (pH 7.5). Next, 8 μ L of Proteinase K (NEB, P8107S) were added to each tube, followed by 15 min incubation at 37°C. Finally, CAGEs were washed three times with 3X SSCI buffer and three times with HI10. Resulting CAGEs were used in the inC-RNA-seq protocol as described in *CAGE single cell RNA-seq*. Resulting barcoded cDNA libraries were purified, fragmented and indexed with 48 indices as described above.

Sequencing

DNA library fragment size was assessed using Agilent

Bioanalyzer hsDNA or Tapestation D5000 kits (Agilent). Sequencing libraries, housing different P5 indices were pooled, and concentration was assessed using a qPCR using KAPA SYBR® FAST Universal qPCR Kit (KAPA Biosystems) according to manufacturer's guidelines. All libraries were sequenced on a Paired-end Illumina Next-Generation sequencing platform. See table S5 for sequencing kit and parameters

Data analysis

Software packages

Except as stated otherwise below, all data analysis was carried out in Python (3.10) with the packages scanpy (1.10.4), numpy (1.26.4), pytorch (2.2.2), pymc (5.20.0), scikit-learn (1.6.1), SnapATAC2 (2.8).

Doublet rate estimation from inC-seq experiments

For barnyard plots in Figs. 2D, 3, C and D, and fig. S3A, comparison of observed doublet rates (shown on plots) to expected doublet rates (see figure captions) was carried out using the following calculations:

The expected number of doublets was calculated using the standard formula for calculating the multiplet fraction for Poisson-loading experiment is $\approx \lambda_{\text{cells}}/2$ [when $\lambda_{\text{cells}} \ll 1$, see (5); calculation in supplementary materials].

In most of the paper, cells were loaded at $\lambda_{\text{cells}} \approx 0.1$ and we expect a total doublet rate of [doublet%] = 5%.

For the species mixing experiments, the observed doublet rate is a fraction of the total doublet rate: [doublet%]^(obs.) = [doublet%] × $[2p(1-p)]$, where p is the fraction of cells from species A (e.g., human), and $(1-p)$ is the fraction from species B (e.g., mouse).

This formula was used as a basis for comparison in all plots with $\lambda_{\text{cells}} = 0.083$ cells/CAGE for Fig. 3C and $\lambda_{\text{cells}} = 0.1$ cells/CAGE for Figs. 2D, 3D, and fig. S3A. p was set to observed human cell fraction.

Sequencing read demultiplexing and pre-processing

BCL sequencing files received from Illumina BaseSpace portal were converted to raw FASTQ files for each of the 4 (R1, R2, R3, R4) sequencing reads without read trimming or masking by using bcl2fastq software (Illumina, v2.20.0.422). Raw FASTQ files were processed using a custom pipeline adapted from inDrops.py (39). The new code, inCseq_filtering.py is available at (38). In brief, sequencing reads were demultiplexed based on Read1 index sequence, and were filtered based on expected structure (all barcodes matching a barcode whitelist with hamming distance of 1). The demultiplexed reads were then processed using Trimmomatic (40) (with parameters: LEADING 28; SLIDINGWINDOW 4:20; MINLEN 20), followed by Poly-A length trim to 4 nt and

selecting reads that are longer than MINLEN. For surviving reads, the fraction of length composed of same base repeats (only considering > 5) was determined, rejecting reads whose fraction is greater than 0.5. As an output, for every input run (sequencing lane), this produces two filtered FASTQ files containing both the Read1 sequence and barcode sequence for every library index contained in that run. Demultiplexed and filtered FASTQ files were used for mapping and counting as described in the following sections.

Read mapping for amplicon sequencing

Sequencing reads were first processed as described above to generate demultiplexed, filtered and sorted FASTQ files. Sequencing reads were aligned to a GFP or RFP reference using STARSolo (41) counting the number of reads aligning to each reference for each detected barcode.

CAGEs RNA-seq count matrix generation

Sequencing reads were first processed as described above to generate demultiplexed, filtered and sorted FASTQ files. Count matrices were generated using STARSolo pipeline (41) (STAR 2.7.9a) with parameters (`-soloFeatures Gene, Gene-Full, -genomeSAsparseD 3, -outSAMtype BAM SortedByCoordinate, -outSAMattributes CB CR UR UB UY GN GX, -outFilterScoreMinOverLread 0.3, -outFilterMatchNminOverLread 0.3, -soloCBmatchWLtype IMM_multi_Nbase_pseudocounts, -soloUMIfiltering Multi-GeneUMI_CR, -soloUMIidedup IMM_CR, -soloMultiMappers EM, -soloCellFilter EmptyDrops_CR, -quantMode TranscriptomeSAM GeneCounts`) or InDrops pipeline with Bowtie 1.2.2 ($m = 200$; $n = 1$; $l = 30$; $e = 70$) unless stated otherwise. Human GRCh38.101 and mouse GRCm.101 genome assemblies were used.

Cell-line CAGE scRNA-seq data analysis

Count matrices were generated as described above using the STARSolo pipeline. Count matrices were filtered by selecting barcodes with >3000 UMIs and fraction of mitochondrial counts being below 0.2. Counts were quantile-quantile (Q-Q) downsampled across all samples. Gene-expression correlation analysis was performed after transcript per million (TPM) normalization and $\log_{10}(1+TPM)$ scaling.

PBMC CAGE scRNA-seq data analysis

Count matrices were generated using the STARSolo pipeline with intronic reads as described above. Count matrices were filtered by selecting barcodes with >2000 UMIs and fraction of mitochondrial counts being below 0.2. Counts were normalized to CPI0K, and z-scored. Scanpy was used to identify high variable genes (`sc.pp.highly_variable_genes` with `flavor='seurat'`, `min_mean=0.0125`, `max_mean=3`, `min_disp=0.2`) and to perform PCA, UMAP and Leiden

clustering.

Data comparison to 10x Genomics

CAGE scRNA-seq HEK293t, NIH 3T3 and PBMC data was compared to an equivalent sample of 10x Genomics scRNA-seq data downloaded from the companies website in raw FASTQ format. Count matrices for CAGE and 10x Genomics data (downloaded from the companies website) were generated using STARSolo pipeline with the same parameters (`-soloType CB_UMI_Simple,-soloUMIlen 12,-soloFeatures Gene,-genomeSAsparseD 3,-outSAMtype BAM SortedByCoordinate,-outSAMattributes CB CR UR UB UY GN GX,-outFilterScoreMinOverLread 0.3,-outFilterMatchNminOverLread 0.3,-soloCBmatchWLtype IMM_multi_Nbase_pseudocounts,-soloUMIfiltering MultiGeneUMI_CR,-soloUMIdedup IMM_CR,-soloMultiMappers EM,-soloCellFilter EmptyDrops_CR,-quantMode TranscriptomeSAM GeneCounts`). Human GRCh38.101 and mouse GRCm.101 assemblies were used. Count matrices were filtered by selecting barcodes with >3000 UMIs and fraction of mitochondrial counts being below 0.2. Total counts were quantile-quantile (Q-Q) downsampled among all samples. Gene-expression correlation analysis was performed after transcript per million (TPM) normalization and $\log_{10}(1+TPM)$ scaling. To compare gene-expression for different transcript length, *gffutils* was used to extract transcripts length from GTF files used to make the genome reference. Gene expression was compared between platforms at different length cutoffs. To generate UMI saturation curves, random sampling was used to sample barcode-umi-gene combinations from output aligned and sorted bam files to generate a table of number of UMIs and reads for each barcode at different sequencing read depths.

Genes/cells as a function of UMIs/cell analysis

10x Genomics PBMC data were downloaded from the companies website as raw FASTQ files that were processed through the same pipeline used to process inC-seq data. SMARTseq3xpress PBMC data was downloaded from the original EMBL-EBI repository as processed gene expression matrices. For the analysis, the number of genes/cell were compared after binning cells by UMIs/cell into low (10K UMI/cell), medium (25K UMI/cell) and high (50K UMI/cell) depths by selecting all cells with an UMI number greater or equal to the bin and downsampling the counts to the bin threshold (10K, 25K, 50K) using *scanpy.pp.downsample_counts* function. 10x Genomics data downloaded as processed gene expression matrices agreed with the data downloaded as raw FASTQ files and processed through STARSolo pipeline.

CAGE ATAC-seq data processing

CAGE ATAC-seq approach used pair-end sequencing to map out gDNA fragment length. While Read1 sequencing reads contain only the gDNA fragment, Read4 contains CAGE barcodes followed by a gDNA fragment. First, parts of Read4 sequencing reads corresponding to gDNA fragments were extracted using *seqtk trimfq* command with a parameter `-b 50`. Sequencing reads were then processed as described above to generate demultiplexed, filtered and sorted FASTQ files twice, once for each gDNA read (Read1 and Read4). Next, *fastq-pair* was used to pair the two demultiplexed and filtered gDNA fragment FASTQ files, which were then merged using *NGmerge*. Merged FASTQ files were aligned to merged human GRCh38.101 and mouse GRCm.101 index using *bowtie2* (with parameter `-very-sensitive, -p 4 -local -X 2000`) and bam files were sorted using *samtools sort* function. Mitochondrial reads were counted and removed, followed by BAM file deduplication, sorting and indexing using *samtools*. Next a pysam script was used to move barcode sequences from BAM read headers to bam CB tag and a string tag was appended to each barcode to label runs corresponding to different indices. Following this, bam files were merged for all the runs. BAM files were analyzed using the SnapATAC2 python package (42). First, BAM files were converted into fragments files using *snapatac2.pp.make_fragment_file* which were filtered (TSSE >5; fragment number >400; FRiP>0.4) after metric calculation (*snapatac2.metrics.tsse*, *snapatac2.tl.macs3*, *snapatac2.metrics.frip*). UMAPs were generated by selecting 50000 features, removing doublets using *scrublet*. Gene activity matrices were generated using *snapatac2.pp.make_gene_matrix*.

CAGE colony RNA sequencing

Preprocessing: Read mapping and cell filtering based on counts and mitochondrial fraction were carried out as described above. Raw counts were then normalized to 10,000 counts per cell (CP10K). Highly variable genes were selected using Scanpy's implementation of the Seurat v3 approach (*sc.pp.highly_variable_genes* with `flavor='seurat_v3'`; 4,000 genes for K562 cells and 5,000 genes for L1210 cells; using unnormalized counts for variance calculations).

Mock colony generation: Mock-colony size-matched controls were generated using a count-based resampling approach. Let $\{n_1, \dots, n_M\}$ be the total counts for the M cells sampled for a given timepoint and condition in the real data. For each value of n_k , we generated a mock cell with the same total counts by (1) randomly sampling single cells (time point 0 transcriptomes) from the corresponding condition, (2) combining raw unnormalized counts to match or exceed target colony sizes, with a minimum of $2^{t/\text{days}}$ cells from time point 0

combined for mock colonies at time t , and enough cells such that the total counts equal or exceed n_k ; then (3) downsampling the counts to n_k . The mock colonies were then processed as the real colonies, except where stated otherwise below.

Visualizing variable genes in colonies: Gene coefficient of variation (CV) vs mean plots were generated by first calling scanpy's `sc.pp.highly_variable_genes` as described above to obtain means, variance and normalized variances for every time point and replicate in for control (untreated samples), for real data and mock clones. The values for each gene were then weighted averaged over the three control replicate samples for colonies from days 4 and 6, weighted by the number of colonies sampled per library. Gene-gene correlation maps were similarly generated for the real data by first calculating pairwise gene correlations within each sample, and then obtaining the weighted average correlation across the three untreated replicates. The top 150 variable genes were selected after filtering out mitochondrial, ribosomal and other predefined gene sets using exclusion lists. For the heatmap of the mean correlations, the genes were hierarchically clustered using average linkage clustering.

Dimensionality reduction: Non-negative matrix factorization (NMF) was performed on the highly variable genes using scikit-learn's NMF implementation with the following parameters: `init='nndsvd'`, `random_state=0`, `n_iter=400`, `alpha_W=0`, `alpha_H=0`, `l1_ratio=1.0`, and `tol=1e-3`. The solver was set to 'cd' (coordinate descent). NMF was carried out on $\log_{10}(1+CP10k)$ -transformed counts. To determine the optimal number of factors, we compared explained variance ratios between real and permuted data across 1-20 components (5), with 5 permutations and sampling 10% of cells for computational efficiency. This procedure identified the top 15 components as explaining variance above-random and this number of NMF programs was used for both cell lines.

Visualization by UMAP embedding: For visualization only, batch effects from library preparation and cell seeding were corrected using Harmony (43) using the scanpy implementation, with the following non-default parameters: `theta=4.0`, `lambda=10.0`, and `sigma=0.05`. Harmony was applied to the 15 NMF program usages per cell. A nearest neighbor graph was constructed using the Harmony-corrected program usages with `n_neighbors=10`, and scanpy's UMAP was then applied to the batch-corrected embedding with default parameters.

Heatmaps of NMF gene loadings: For heatmaps of gene contributions to NMF programs, the NMF loading matrix ("H" matrix in scikit-learn) was normalized by dividing each

gene's loading by its mean nonzero expression level (calculated as mean expression when expressed / fraction of cells expressing). The genes shown in the final heatmaps were filtered to remove mitochondrial genes and non-coding genes to improve interpretability, and ordered by descending normalized loading values.

NMF usage means, CVs and Fano factors: NMF program usage values were first normalized for each cell such that the sum of usages across all programs equaled 1. The mean, CV, and Fano factor of each program's usage across colonies were then calculated at each timepoint. To calculate mock colony CV and Fano factors, we corrected for systematic differences in means over time by regressing the CVs using the power law relationship observed between mean and CV. For comparisons across conditions, each metric was normalized to its value in control colonies at the same timepoint.

Machine learning program ON/OFF switching rates: A model of stochastic state transitions within cells of an expanding clone was fit to the distribution of usages for each NMF program over time, in each condition. Referring to the supplementary text for details on the model, the parameters are: the on rate (r_{01} / cell cycle), the off rate (r_{10} / cell cycle), the initial state probability (p_0), the NMF program usage background level (ϵ), and the NMF program usage expression scale and noise parameters (a, b, σ, σ_0). These parameters were inferred by Markov Chain Monte Carlo (MCMC) sampling using PyMC, with a prior distribution and likelihood as defined in supplementary text 1. To generate the likelihood, we carried out $1.6 \cdot 10^6$ simulations of stochastic cell state switching with different rates (r_{01}, r_{10}) and then trained a surrogate neural network model (2 hidden layers: 64, 32 nodes) to predict the number of cells in states (0,1) in each clone over time t , enabling rapid likelihood computation during MCMC inference. Parameters were inferred using 8 parallel MCMC chains of 2,000 samples each after 2,000 tuning steps. Convergence was assessed using the Gelman-Rubin statistic. Rates (r_{01}, r_{10}) increasing following drug treatment imply shorter persistence time for a given NMF program, while the opposite implies longer persistence. Code implementing simulations, surrogate model training and MCMC inference are available at (38).

REFERENCES AND NOTES

1. K. Vandereyken, A. Sifrim, B. Thienpont, T. Voet, Methods and applications for single-cell and spatial multi-omics. *Nat. Rev. Genet.* **24**, 494–515 (2023). [doi:10.1038/s41576-023-00580-2](https://doi.org/10.1038/s41576-023-00580-2) [Medline](#)
2. A. Baysoy, Z. Bai, R. Satija, R. Fan, The technological landscape and applications of single-cell multi-omics. *Nat. Rev. Mol. Cell Biol.* **24**, 695–713 (2023). [doi:10.1038/s41580-023-00615-w](https://doi.org/10.1038/s41580-023-00615-w) [Medline](#)
3. M. Hagemann-Jensen, C. Ziegenhain, P. Chen, D. Ramsköld, G.-J. Hendriks, A. J. M. Larsson, O. R. Faridani, R. Sandberg, Single-cell RNA counting at allele and

- isoform resolution using Smart-seq3. *Nat. Biotechnol.* **38**, 708–714 (2020). [doi:10.1038/s41587-020-0497-0](https://doi.org/10.1038/s41587-020-0497-0) [Medline](#)
4. M. Hagemann-Jensen, C. Ziegenhain, R. Sandberg, Scalable single-cell RNA sequencing from full transcripts with Smart-seq3xpress. *Nat. Biotechnol.* **40**, 1452–1457 (2022). [doi:10.1038/s41587-022-01311-4](https://doi.org/10.1038/s41587-022-01311-4) [Medline](#)
 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). [doi:10.1016/j.cell.2015.04.044](https://doi.org/10.1016/j.cell.2015.04.044) [Medline](#)
 6. T. M. Gierahn, M. H. Wadsworth 2nd, T. K. Hughes, B. D. Bryson, A. Butler, R. Satija, S. Fortune, J. C. Love, A. K. Shalek, Seq-Well: Portable, low-cost RNA sequencing of single cells at high throughput. *Nat. Methods* **14**, 395–398 (2017). [doi:10.1038/nmeth.4179](https://doi.org/10.1038/nmeth.4179) [Medline](#)
 7. I. C. Clark, K. M. Fontanez, R. H. Meltzer, Y. Xue, C. Hayford, A. May-Zhang, C. D'Amato, A. Osman, J. Q. Zhang, P. Hettige, J. S. A. Ishibashi, C. L. Delley, D. W. Weisgerber, J. M. Replogle, M. Jost, K. T. Phong, V. E. Kennedy, C. A. C. Peretz, E. A. Kim, S. Song, W. Karlson, J. S. Weissman, C. C. Smith, Z. J. Gartner, A. R. Abate, Microfluidics-free single-cell genomics with templated emulsification. *Nat. Biotechnol.* **41**, 1557–1566 (2023). [doi:10.1038/s41587-023-01685-z](https://doi.org/10.1038/s41587-023-01685-z) [Medline](#)
 8. J. Cao, D. A. Cusanovich, V. Ramani, D. Aghamirzaie, H. A. Pliner, A. J. Hill, R. M. Daza, J. L. McFaline-Figueroa, J. S. Packer, L. Christiansen, F. J. Steemers, A. C. Adey, C. Trapnell, J. Shendure, Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**, 1380–1385 (2018). [doi:10.1126/science.aau0730](https://doi.org/10.1126/science.aau0730) [Medline](#)
 9. A. B. Rosenberg, C. M. Roco, R. A. Muscat, A. Kuchina, P. Sample, Z. Yao, L. T. Graybuck, D. J. Peeler, S. Mukherjee, W. Chen, S. H. Pun, D. L. Sellers, B. Tasic, G. Seelig, Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176–182 (2018). [doi:10.1126/science.aam8999](https://doi.org/10.1126/science.aam8999) [Medline](#)
 10. J. de Rutte, R. Dimatteo, M. M. Archang, M. van Zee, D. Koo, S. Lee, A. C. Sharrow, P. J. Krohl, M. Mellody, S. Zhu, J. V. Eichenbaum, M. Kizerwetter, S. Udani, K. Ha, R. C. Willson, A. L. Bertozzi, J. B. Spangler, R. Damoiseaux, D. Di Carlo, Suspendable Hydrogel Nanovials for Massively Parallel Single-Cell Functional Analysis and Sorting. *ACS Nano* **16**, 7242–7257 (2022). [doi:10.1021/acsnano.1c11420](https://doi.org/10.1021/acsnano.1c11420) [Medline](#)
 11. S. Udani, J. Langerman, D. Koo, S. Baghdasarian, B. Cheng, S. Kang, C. Soemardy, J. de Rutte, K. Plath, D. Di Carlo, Associating growth factor secretions and transcriptomes of single cells in nanovials using SEC-seq. *Nat. Nanotechnol.* **19**, 354–363 (2024). [doi:10.1038/s41565-023-01560-7](https://doi.org/10.1038/s41565-023-01560-7) [Medline](#)
 12. J. L. McFaline-Figueroa, S. Srivatsan, A. J. Hill, M. Gasperini, D. L. Jackson, L. Saunders, S. Domcke, S. G. Regalado, P. Lazarchuck, S. Alvarez, R. J. Monnat Jr., J. Shendure, C. Trapnell, Multiplex single-cell chemical genomics reveals the kinase dependence of the response to targeted therapy. *Cell Genom.* **4**, 100487 (2024). [doi:10.1016/j.xgen.2023.100487](https://doi.org/10.1016/j.xgen.2023.100487) [Medline](#)
 13. G. H. T. Yeo, L. Lin, C. Y. Qi, M. Cha, D. K. Gifford, R. I. Sherwood, A Multiplexed Barcodelet Single-Cell RNA-Seq Approach Elucidates Combinatorial Signaling Pathways that Drive ESC Differentiation. *Cell Stem Cell* **26**, 938–950.e6 (2020). [doi:10.1016/j.stem.2020.04.020](https://doi.org/10.1016/j.stem.2020.04.020) [Medline](#)
 14. G. Leonaviciene, L. Mazutis, RNA cytometry of single-cells using semi-permeable microcapsules. *Nucleic Acids Res.* **51**, e2 (2023). [doi:10.1093/nar/gkac918](https://doi.org/10.1093/nar/gkac918) [Medline](#)
 15. G. Leonaviciene, K. Leonavicius, R. Meskys, L. Mazutis, Multi-step processing of single cells using semi-permeable capsules. *Lab Chip* **20**, 4052–4062 (2020). [doi:10.1039/D0LC00660B](https://doi.org/10.1039/D0LC00660B) [Medline](#)
 16. D. Baronas, J. Zvirblyte, S. Norvaisis, G. Leonaviciene, K. Goda, V. Mikulenaite, V. Kasetka, K. Sablauskas, L. Griskevicius, S. Juzenas, L. Mazutis, High-throughput single cell -omics using semi-permeable capsules. *Science* **2025**, eady7227 (2025). [doi:10.1126/science.ady7227](https://doi.org/10.1126/science.ady7227) [Medline](#)
 17. T. Y. Lee, T. M. Choi, T. S. Shim, R. A. M. Frijns, S.-H. Kim, Microfluidic production of multiple emulsions and functional microcapsules. *Lab Chip* **16**, 3415–3440 (2016). [doi:10.1039/C6LC00809G](https://doi.org/10.1039/C6LC00809G) [Medline](#)
 18. S. Ma, J. Thiele, X. Liu, Y. Bai, C. Abell, W. T. S. Huck, Fabrication of microgel particles with complex shape via selective polymerization of aqueous two-phase systems. *Small* **8**, 2356–2360 (2012). [doi:10.1002/sml.201102715](https://doi.org/10.1002/sml.201102715) [Medline](#)
 19. M. Yanagisawa, S. Nigorikawa, T. Sakaue, K. Fujiwara, M. Tokita, Multiple patterns of polymer gels in microspheres due to the interplay among phase separation, wetting, and gelation. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15894–15899 (2014). [doi:10.1073/pnas.1416592111](https://doi.org/10.1073/pnas.1416592111) [Medline](#)
 20. D. Attwood, J. Collett, C. Tait, The micellar properties of the poly(oxyethylene) - poly(oxypropylene) copolymer Pluronic F127 in water and electrolyte solution. *Int. J. Pharm.* **26**, 25–33 (1985). [doi:10.1016/0378-5173\(85\)90197-8](https://doi.org/10.1016/0378-5173(85)90197-8)
 21. M. Di Biase, P. de Leonardis, V. Castelletto, I. W. Hamley, B. Derby, N. Tirelli, Photopolymerization of Pluronic F127 diacrylate: A colloid-templated polymerization. *Soft Matter* **7**, 4928 (2011). [doi:10.1039/c1sm05095h](https://doi.org/10.1039/c1sm05095h)
 22. A. Y. L. Sim, J. Lipfert, D. Herschlag, S. Doniach, Salt dependence of the radius of gyration and flexibility of single-stranded DNA in solution probed by small-angle x-ray scattering. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **86**, 021901 (2012). [doi:10.1103/PhysRevE.86.021901](https://doi.org/10.1103/PhysRevE.86.021901) [Medline](#)
 23. Y. Lu, B. Weers, N. C. Stellwagen, DNA persistence length revisited. *Biopolymers* **61**, 261–275 (2001-2002). [doi:10.1002/bip.10151](https://doi.org/10.1002/bip.10151) [Medline](#)
 24. J. D. Buenrostro, B. Wu, H. Y. Chang, W. J. Greenleaf, ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 29.1.9 (2015). [doi:10.1002/0471142727.mb2129s109](https://doi.org/10.1002/0471142727.mb2129s109) [Medline](#)
 25. T. Scully, A. Klein, A mannitol-based buffer improves single-cell RNA sequencing of high-salt marine cells, *bioRxiv* 2023.04.26.538465 [Preprint] (2023). <https://doi.org/10.1101/2023.04.26.538465>
 26. M. De Simone, J. Hoover, J. Lau, H. Bennet, B. Wu, C. Chen, H. Menon, A. Au-Yeung, S. Lear, S. Vaidya, M. Shi, J. M. Lund, A. Xavier-Magalhaes, Y. Liang, A. Kurdoglu, W. E. O'Gorman, Z. Modrusan, D. Le, S. Darmanis, Comparative Analysis of Commercial Single-Cell RNA Sequencing Technologies. *bioRxiv* 2024.06.18.599579 [Preprint] (2024). <https://doi.org/10.1101/2024.06.18.599579>
 27. B. Xia, K. Krutkramelis, J. Oakey, Oxygen-purged microfluidic device to enhance cell viability in photopolymerized PEG hydrogel microparticles. *Biomacromolecules* **17**, 2459–2465 (2016). [doi:10.1021/acs.biomac.6b00597](https://doi.org/10.1021/acs.biomac.6b00597) [Medline](#)
 28. S. L. Spencer, S. Gaudet, J. G. Albeck, J. M. Burke, P. K. Sorger, Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* **459**, 428–432 (2009). [doi:10.1038/nature08012](https://doi.org/10.1038/nature08012) [Medline](#)
 29. S. M. Shaffer, M. C. Dunagin, S. R. Torborg, E. A. Torre, B. Emert, C. Krepler, M. Beqiri, K. Sproesser, P. A. Brafford, M. Xiao, E. Eggen, I. N. Anastopoulos, C. A. Vargas-Garcia, A. Singh, K. L. Nathanson, M. Herlyn, A. Raj, Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* **546**, 431–435 (2017). [doi:10.1038/nature22794](https://doi.org/10.1038/nature22794) [Medline](#)
 30. A. Gnyoszka, Z. Jastrzebski, S. Flis, DNA methyltransferase inhibitors and their emerging role in epigenetic therapy of cancer. *Anticancer Res.* **33**, 2989–2996 (2013). [Medline](#)
 31. J. E. Bolden, M. J. Peart, R. W. Johnstone, Anticancer activities of histone deacetylase inhibitors. *Nat. Rev. Drug Discov.* **5**, 769–784 (2006). [doi:10.1038/nrd2133](https://doi.org/10.1038/nrd2133) [Medline](#)
 32. J. Strauss, W. D. Figg, Using epigenetic therapy to overcome chemotherapy resistance. *Anticancer Res.* **36**, 1–4 (2016). [Medline](#)
 33. B. T. Oronsky, A. L. Oronsky, M. Lybeck, N. C. Oronsky, J. J. Scicinski, C. Carter, R. M. Day, J. F. Rodriguez Orengo, M. Rodriguez-Torres, G. F. Fanger, T. R. Reid, Episenitization: Defying Time's Arrow. *Front. Oncol.* **5**, 134 (2015). [doi:10.3389/fonc.2015.00134](https://doi.org/10.3389/fonc.2015.00134) [Medline](#)
 34. M. Szyf, Epigenetics, DNA methylation, and chromatin modifying drugs. *Annu. Rev. Pharmacol. Toxicol.* **49**, 243–263 (2009). [doi:10.1146/annurev-pharmtox.061008.103102](https://doi.org/10.1146/annurev-pharmtox.061008.103102) [Medline](#)
 35. S. M. Shaffer, B. L. Emert, R. A. Reyes Hueros, C. Cote, G. Harmange, D. L. Schaff, A. E. Sizemore, R. Gupte, E. Torre, A. Singh, D. S. Bassett, A. Raj, Memory sequencing reveals heritable single-cell gene expression programs associated with distinct cellular behaviors. *Cell* **182**, 947–959.e17 (2020). [doi:10.1016/j.cell.2020.07.003](https://doi.org/10.1016/j.cell.2020.07.003) [Medline](#)
 36. Z. Meir, Z. Mukamel, E. Chomsky, A. Lifshitz, A. Tanay, Single-cell analysis of clonal maintenance of transcriptional and epigenetic states in cancer cells. *Nat. Genet.* **52**, 709–718 (2020). [doi:10.1038/s41588-020-0645-y](https://doi.org/10.1038/s41588-020-0645-y) [Medline](#)
 37. K. Nepali, J.-P. Liou, Recent developments in epigenetic cancer therapeutics: Clinical advancement and emerging trends. *J. Biomed. Sci.* **28**, 27 (2021). [doi:10.1186/s12929-021-00721-x](https://doi.org/10.1186/s12929-021-00721-x) [Medline](#)

38. I. Mazelis, A. M. Klein, AllonKleinLab/Mazelis2025_CAGEs: Paper release, version v0.91, Zenodo (2025); <https://doi.org/10.5281/zenodo.17373720>.
39. R. Zilionis, J. Nainys, A. Veres, V. Savova, D. Zemmour, A. M. Klein, L. Mazutis, Single-cell barcoding and sequencing using droplet microfluidics. *Nat. Protoc.* **12**, 44–73 (2017). [doi:10.1038/nprot.2016.154](https://doi.org/10.1038/nprot.2016.154) [Medline](#)
40. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014). [doi:10.1093/bioinformatics/btu170](https://doi.org/10.1093/bioinformatics/btu170) [Medline](#)
41. B. Kaminow, D. Yunusov, A. Dobin, STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. *bioRxiv* 2021.05.05.442755 [Preprint] (2021). <https://doi.org/10.1101/2021.05.05.442755>.
42. K. Zhang, N. R. Zemke, E. J. Armand, B. Ren, A fast, scalable and versatile tool for analysis of single-cell omics data. *Nat. Methods* **21**, 217–227 (2024). [doi:10.1038/s41592-023-02139-9](https://doi.org/10.1038/s41592-023-02139-9) [Medline](#)
43. I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, F. Zhang, K. Wei, Y. Baglaenko, M. Brenner, P.-R. Loh, S. Raychaudhuri, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019). [doi:10.1038/s41592-019-0619-0](https://doi.org/10.1038/s41592-019-0619-0) [Medline](#)
44. J. Peccoud, B. Ycart, Markovian modeling of gene-product synthesis. *Theor. Popul. Biol.* **48**, 222–234 (1995). [doi:10.1006/tpbi.1995.1027](https://doi.org/10.1006/tpbi.1995.1027)
45. A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, S. Tyagi, Stochastic mRNA synthesis in mammalian cells. *PLOS Biol.* **4**, e309 (2006). [doi:10.1371/journal.pbio.0040309](https://doi.org/10.1371/journal.pbio.0040309) [Medline](#)
46. A. M. Klein, D. P. Dou  , P. H. Jones, B. D. Simons, Kinetics of cell division in epidermal maintenance. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **76**, 021910 (2007). [doi:10.1103/PhysRevE.76.021910](https://doi.org/10.1103/PhysRevE.76.021910) [Medline](#)

ACKNOWLEDGMENTS

We thank Linas Mazutis, Denis Baronas and Tim Mitchison for ongoing discussions and feedback during this project, John Oakey for guidance of ROS generation during photopolymerization, the labs of Olivier Pourqui   and Fernando Camargo for providing iPSC and HSC cells; Sean McGeary for providing K562 and L929 cells, Yuyang Chen for providing L1210 cells, the lab of Michael Greenberg for access to their Illumina Sequencer; Jeffery A. Nelson at the Bauer Core Facility at Harvard University for FACS work coordination; and the Single Cell Core at HMS for supporting work by TT and AK. The cryo-SEM work was performed in part at the Harvard University Center for Nanoscale Systems (CNS); a member of the National Nanotechnology Coordinated Infrastructure Network (NNCI), which is supported by the National Science Foundation under NSF award no. ECCS-2025158. We extend our heartfelt gratitude to Adam Graham at the Harvard CNS core, whose expertise was instrumental in bringing this cryo-SEM work to fruition. We thank all members of the Klein lab for critical reading and feedback on the manuscript. **Funding:** This work was supported by NIH grants R21HG012771 and R33CA278392, and by an Edward Mallinckrodt Jr Scholar Award to AMK. Pilot work on the genomic assays was supported by an HMS Q-FASTR Pilot Grant and pilot work solving viability in live-cell assays by an HMS Blavatnik Biomedical Accelerator Pilot Grant. The Harvard University Center for Nanoscale Systems (CNS) is a member of the National Nanotechnology Coordinated Infrastructure Network (NNCI), which is supported by the National Science Foundation under NSF award no. ECCS-2025158. **Author contributions:** IM developed the capsules and carried out all experiments except FACS and Cryo-SEM. HS supported diffusion assays and optimizing cell culture conditions in capsules. TT carried out flow-sorting in capsules. AK carried out SEM imaging with support of the CNS core facility. IM and AMK conceived and designed experiments and carried out data analysis. AMK supervised the work. IM and AMK wrote the manuscript. All authors provided feedback on the manuscript. **Competing interests:** AMK is a co-founder of Somite Therapeutics, Ltd. IM and AMK are inventors on patent application PCT/US2023/029364 filed by Harvard University. **Data, code, and materials availability:** Sequencing data available at Gene Expression Omnibus (GSE306708) and Sequence Read Archive (PRJNA1311090). Python code for capsule segmentation and pipeline for scRNA-seq read pre-processing are archived at Zenodo (38). **License information:** Copyright    2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US

government works. <https://www.science.org/about/science-licenses-journal-article-reuse>

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ady7209

Supplementary Text

Figures S1 to S7

Tables S1 to S5

References (44–46)

Movie S1

MDAR Reproducibility Checklist

Submitted 2 May 2025; accepted 5 December 2025

Published online 18 December 2025

10.1126/science.ady7209

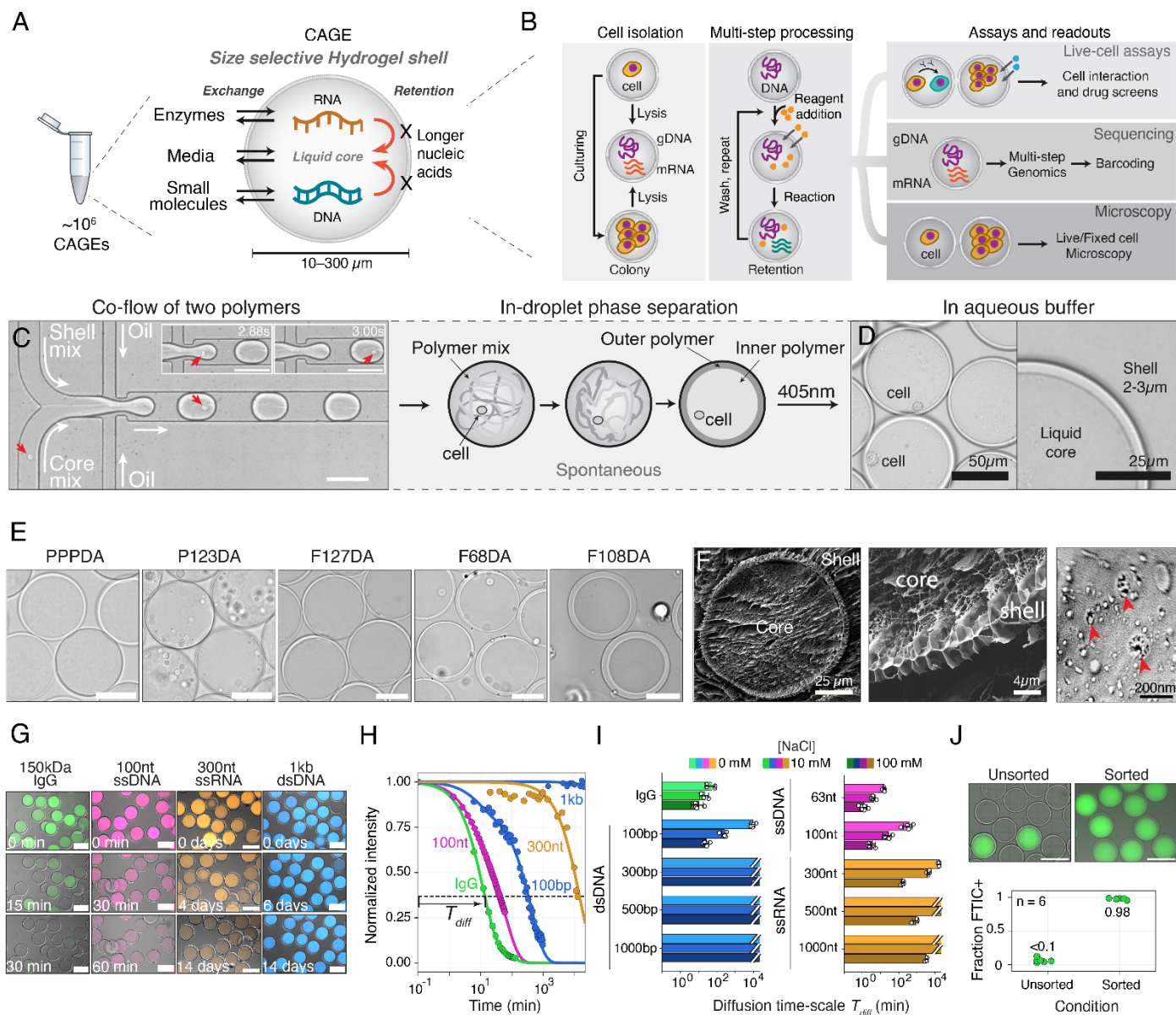


Fig. 1. Concept, development, and characterization of Pluronic diacrylate capsules. (A) Schematic of semi-permeable capsules and their relevant properties for high-throughput, single-cell profiling. (B) Schematic of potential massively-parallel, multistep analyses in capsules including processing of single or cultured cells through iterative addition and removal of reagents, imaging, sorting and sequencing. (C) Bright-field micrograph (left) and schematic (right) showing generation of CAGEs (compartments with amphiphilic gel envelopes) using a droplet generator co-flowing an amphiphilic functionalized outer (shell) polymer and a hydrophilic core polymer that undergo phase-separation upon droplet formation. These are subsequently converted to CAGEs by 405nm-dependent cross-linking of the shell polymer. Red arrows indicate single cells. Inset displays two snapshots spaced by 0.12s. Scale bar = 100 μm . (D) Bright-field micrographs of 8% (w/v):1% (w/v) F127DA:PPPDA shell, 13% (w/v) dextran core CAGEs in DPBS containing individual cells. (E) Bright-field micrographs demonstrating that multiple amphiphilic PEG copolymers form capsules. Scale bar = 50 μm . (F) Cryo-SEM images of freeze-fractured capsules with magnification of pores in the Pluronic diacrylate shell membrane. Red arrows point to surface pores. (G) Illustrative composite micrographs of confocal and bright-field imaging of capsules from analyte diffusion time-series. (H) Intensity time-series quantifying diffusion half-times through the CAGE shell [composition as in (D)] for analytes of different sizes and in the presence of different salt concentrations. (I) Diffusion half-times showing a sharp size-dependent cut-off in CAGE permeability. Plot represents mean data from ≥ 3 independent experiments with SD as error bars; full data in table S2. (J) Flow cytometry-based enrichment of CAGEs. Composite fluorescence and bright-field micrographs of a sample initially containing <10% labeled capsules (top-left) and sample after sorting (top-right), yielding ~98% FITC-positive CAGEs (bottom). Scale bar = 100 μm .

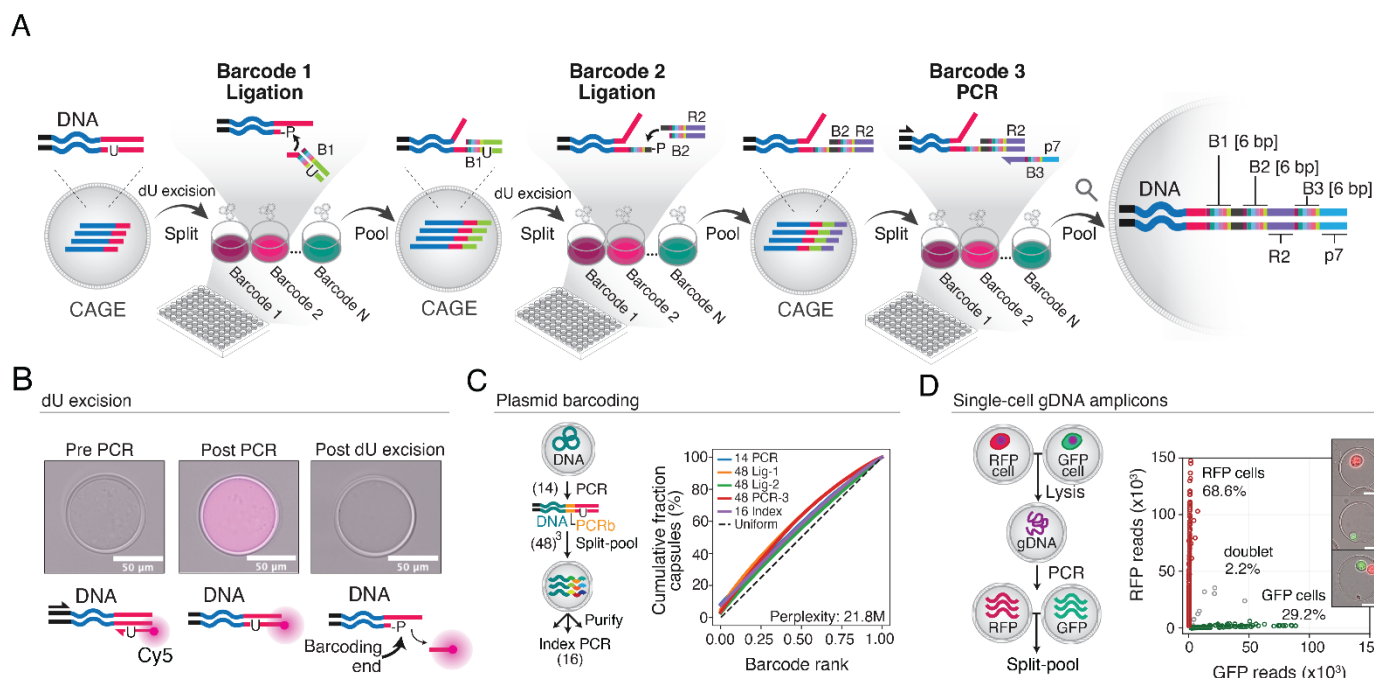


Fig. 2. Split-and-pool-based barcoding of CAGEs for high-throughput sequencing (inC-seq) (A) Schematic for split-and-pool barcoding of DNA libraries in capsules, showing an example of three rounds of barcoding through ligation and PCR used in this work. Initial libraries are generated with a uracil residue in the common library end-sequence (red), which is cleaved by a uracil-specific excision reagent to generate a sticky end with a 5'-phosphate. The sticky ends are ligated to an array of barcoding dsDNA oligos that contain a well-specific barcode. The process is repeated to introduce a second barcode, and a third barcode is introduced through a PCR primer. (B) Demonstration of efficient sequential enzymatic processing of DNA in CAGEs, realizing step (1) from (A). An initial unlabeled library (left panel) is amplified by incubation of capsules in a PCR reaction mix, with a PCR primer containing a uracil and a Cy5 label at its 3' terminus. Following PCR and washing (middle panel), the CAGEs show retention of Cy5 fluorescence that now is attached to the DNA in the capsule. Uracil excision and washing cleaves the labeled primer generating an overhang and losing Cy5 fluorescence (right panel). (C) Representation of barcodes in the DNA amplicon sequencing data after a 5-step barcoding approach: 14 PCR barcodes (PCRb) were added during amplification, followed by 3-steps of 48 split-pool barcode addition and 16 indices following capsule solubilization. (D) Effective library partitioning by capsules with low cross-contamination seen by inC-seq analysis of single-cell gDNA transgenes loci encoding GFP or RFP sequences. The plot shows the number of reads from mapping to GFP or RFP, with each point corresponding to a unique capsule barcode. Observed doublet rate is in line with expectations (2.1%) for CAGEs generated with a mean cell loading of 0.1 cells/CAGE and an observed cell loading ratio of 70:30 RFP:GFP. Inset displays examples of CAGEs with cells prior to lysis. Scale bar = 50 μ m.

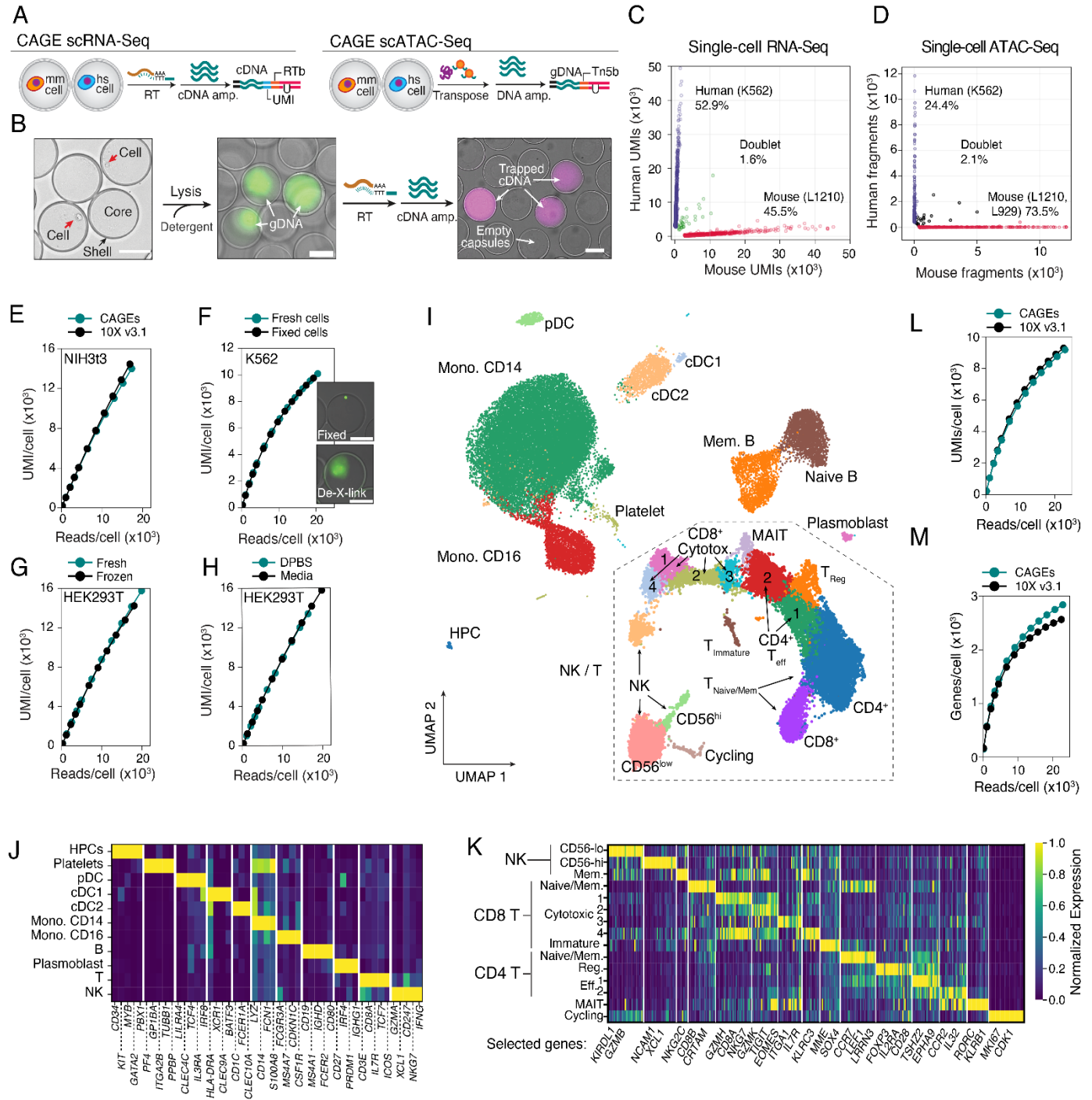


Fig. 3. Capsules enable single-cell genomic assays with multistep processing. (A) Schematic of inC-RNA-seq and inC-ATAC-seq library preparation modalities. Additional barcodes and sequences are added during transcript (RTb and UMI) and accessible gDNA region (Tn5b) capture via RT and Tagmentation steps. Captured and amplified dU bearing molecules are subjected to split-pool barcoding and sequencing. (B) Micrographs showing cells at progressive stages of single-cell RNA-seq library preparation in CAGEs, from initial cell encapsulation through lysis, to generation of amplified complementary DNA (cDNA) libraries prior to in-CAGE barcoding. gDNA is stained with SYBR Safe dye; cDNA stained with Cy5 fluorophores conjugated to the PCR amplification primers. (C and D) Plots showing the number of UMIs or DNA fragments corresponding to each CAGE mapping to human (y-axis) and mouse (x-axis) RNA-seq (left) or ATAC-Seq (right) inC-seq libraries. The observed mixed species doublet rates are in line with the expectations: 2.1% for (C) with (loading of 0.083 cells/CAGE) and 1.9% for (D) (loading of 0.1 cells/CAGE). (E) UMI/read plots following down-sampling reads from CAGE scRNA-seq libraries in comparison to publicly available 10x Genomics NextGem v3.1 data. (F to H) UMI/reads plots as in (E) for libraries generated from fresh or PFA fixed K562 cells (F), fresh or frozen HEK293T cells (G) or capsules made using DPBS or media (H). (I) A two-dimensional UMAP embedding of inC-RNA-Seq data from peripheral blood mononuclear cells (PBMCs), (J) heatmaps of gene expression showing coverage of low and high-abundance cell type-specific genes, and (K) high population structure of partitioning subsets of T and NK cells. (L) UMI/reads plots as in (F), now generated for PBMC data and comparing inC-seq to data from 10x Genomics GEMCode. (M) Comparison of number of genes detected per cell as a function of read depth between inC-seq and 10x Genomics NextGem v3.1.

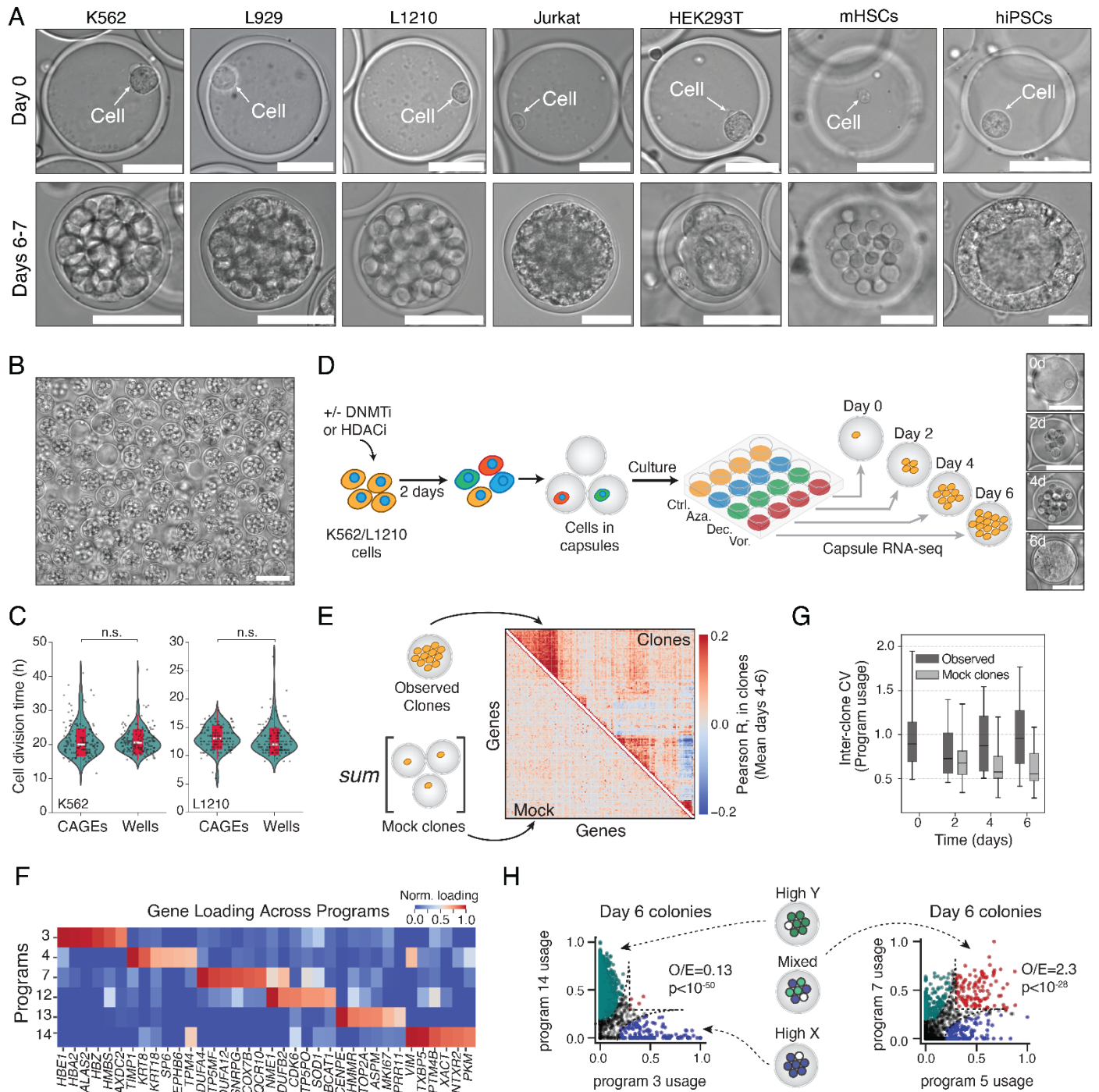


Fig. 4. Live-cell clonal expansion in CAGEs reveals persistent gene expression programs. (A) Micrographs showing growth in CAGEs of clones from cultured cell lines, primary mouse bone marrow hematopoietic stem cells (mHSCs) and human induced pluripotent stem cells (hiPSCs). Scale bars = 50 μm . (B) Micrograph of encapsulated colonies concurrently cultured. Scale bar = 100 μm . The image shows high-density K562 (poly-clonal) colonies expanded in CAGEs. (C) Division times of cells cultured in wells or CAGEs. Each dot corresponds to a single cell-division time evaluated by time-lapse microscopy (K562: $n=156$; L1210: $n=127$). t test $p\text{-value} > 0.8$. (D) Experimental schema for identifying persistent gene expression programs in cancer cell lines upon treatment with vehicle (Ctrl), DNMT inhibitors (5-azacytidine (Aza); decitabine (Dec)), or an HDAC inhibitor (vorinostat (Vor)). Micrographs show example capsules at different timepoints. (E to H) Evidence of persistent gene expression programs in K562 cells; see fig. S6 for L1210 cells. (E) Clustered inter-clonal gene-expression correlations in K562 cell control samples at days 4-6 show evidence of structured programs, which are absent when randomly combining cells sampled at day 0 into 'mock' clones. (F) Top genes contributing to gene expression programs identified by nonnegative matrix factorization (NMF), identifying erythroid, cell cycle, vimentin- and keratin-associated programs. Full program loadings in table S3. (G) Box plots of variation in 15 NMF-derived programs between control clones, showing persistent heterogeneity over time compared to mock clones. (H) Some gene expression programs remain mutually exclusive after clonal expansion, as seen from low numbers of clones observed mixed NMF programs as compared to expectation ($\text{Observed/Expected} = f_{XY}/f_X f_Y$; $p\text{-value}$ from Fisher's exact test).

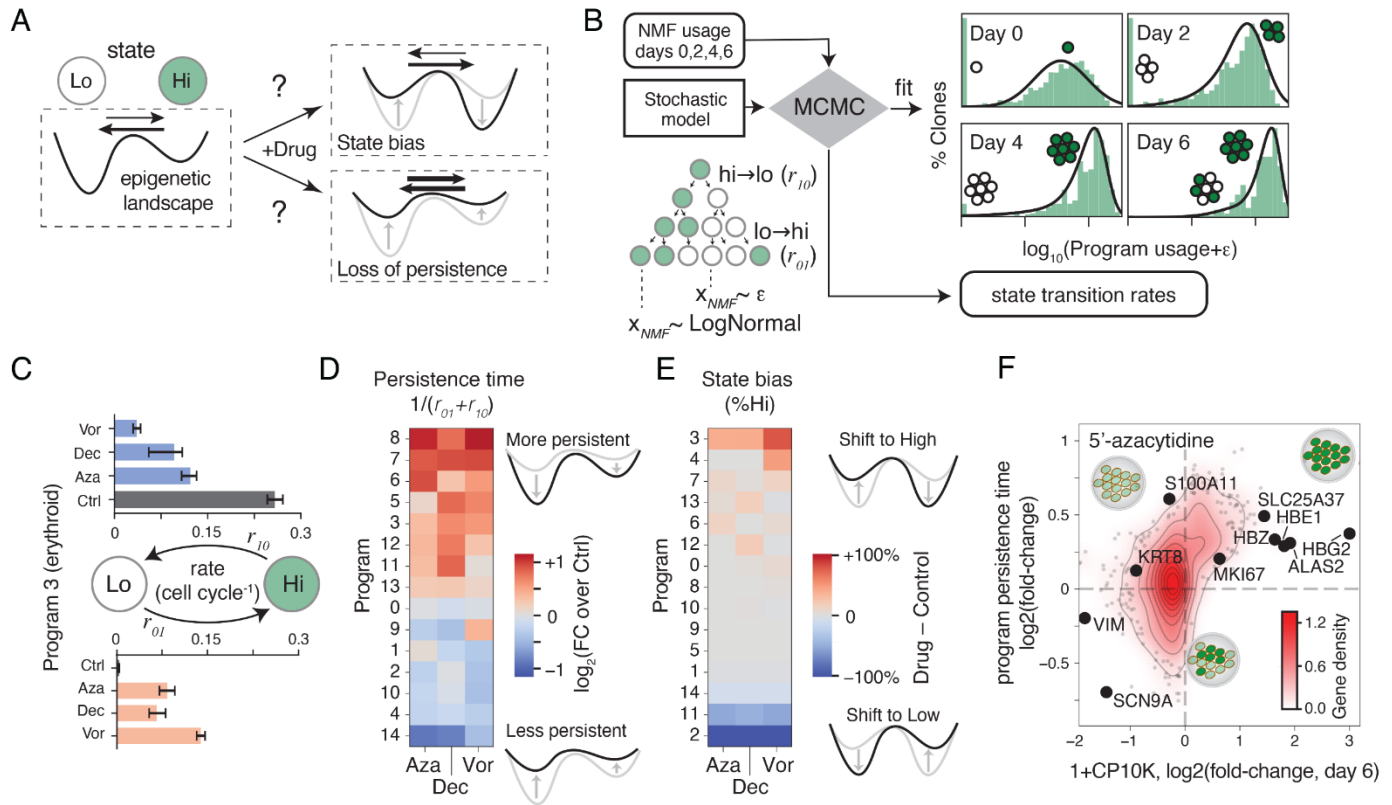


Fig. 5. Altered clonal memory upon treatment with DNMT or HDAC inhibitors. (A) A schematic of possible effects of drug treatment on epigenetic memory. In untreated controls, cells occupy high and low gene expression states for each observed program. Inhibitors of epigenetic modifying enzymes may bias the relative stability of states (top scenario), and may also reduce persistence (shallower wells, bottom scenario). From inC-RNA-seq on growing clones one can infer state transition rates (arrows). (B) Schema for statistical inference of transition rates from clonal inC-RNA-seq data by fitting stochastic models of cell state-transitions to observed NMF program usage over time. The model was used to infer rates of program induction (r_{01}) and loss (r_{10}) for each program in each treatment condition (see supplementary text 1). (C) Fitted transition rates exemplified for one program (program 3), with all drugs increasing persistence of the 'high' state, while destabilizing the 'low' state. (D) Changes in K562 gene expression program persistence times from the fitted dynamic transition rates, showing reduced clonal memory for some programs (blue) but not others (red). (E) Corresponding changes in steady-state bias (i.e., the fraction of cells in the active state), showing distinct changes from program persistence. (F) Changes in persistence are distinct from changes in mean expression for single genes, shown for one drug (Aza) in K562 cells.