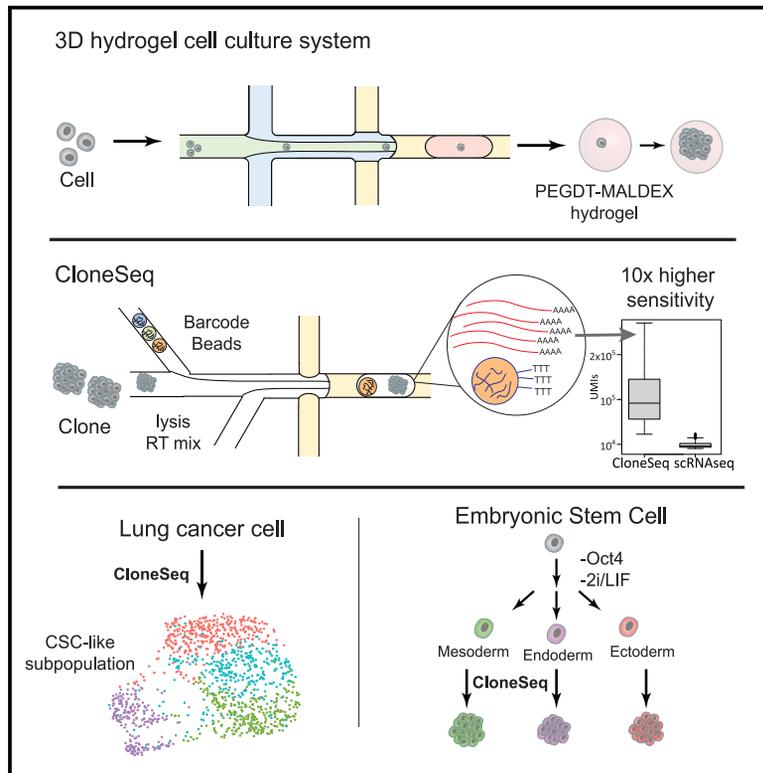


Developmental Cell

CloneSeq: A highly sensitive analysis platform for the characterization of 3D-cultured single-cell-derived clones

Graphical abstract



Authors

Danny Bavli, Xue Sun, Chen Kozulin, ..., Eran Meshorer, Amnon Buxboim, Oren Ram

Correspondence

amnon.buxboim@mail.huji.ac.il (A.B.), oren.ram@mail.huji.ac.il (O.R.)

In brief

Bavli et al. establish a microfluidic-based analysis platform, which they name CloneSeq, for 3D-cultured clones derived from single cells. CloneSeq has an enhanced sensitivity over scRNA-seq, and the 3D culture system itself supports cell stemness. They use this platform to analyze cancer cell subpopulations and embryonic stem cell differentiation.

Highlights

- Integrative mRNA sequencing of single clones shows enhanced sensitivity
- PEGDT-MALDEX hydrogel supports stemness in cancer cells and ESCs
- CloneSeq identifies cancer stem-like subpopulation
- Differentiation decisions of ESCs are maintained during clonal expansion



Technology

CloneSeq: A highly sensitive analysis platform for the characterization of 3D-cultured single-cell-derived clones

Danny Bavli,^{1,3,5} Xue Sun,^{1,5} Chen Kozulin,^{1,5} Dena Ennis,¹ Alex Motzik,¹ Alva Biran,^{1,2} Shlomi Brielle,^{3,4} Adi Alajem,¹ Eran Meshorer,² Amnon Buxboim,^{3,4,6,*} and Oren Ram^{1,6,*}

¹Department of Biological Chemistry, Alexander Silberman Institute of Life Sciences, the Hebrew University of Jerusalem, Jerusalem, Israel

²Department of Genetics, Alexander Silberman Institute of Life Sciences and the Edmond and Lily Safra Center for Brain Sciences (ELSC), the Hebrew University of Jerusalem, Jerusalem, Israel

³Alexander Grass Center for Bioengineering, The Hebrew University of Jerusalem, Jerusalem, Israel

⁴Department of Cell and Developmental Biology, Hebrew University of Jerusalem, Givat Ram, Jerusalem, Israel

⁵These authors contributed equally

⁶Lead contact

*Correspondence: amnon.buxboim@mail.huji.ac.il (A.B.), oren.ram@mail.huji.ac.il (O.R.)

<https://doi.org/10.1016/j.devcel.2021.04.026>

SUMMARY

Single-cell assays have revealed the importance of heterogeneity in many biological systems. However, limited sensitivity is a major hurdle for uncovering cellular variation. To overcome it, we developed CloneSeq, combining clonal expansion inside 3D hydrogel spheres and droplet-based RNA sequencing (RNA-seq). We show that clonal cells maintain similar transcriptional profiles and cell states. CloneSeq of lung cancer cells revealed cancer-specific subpopulations, including cancer stem-like cells, that were not revealed by scRNA-seq. Clonal expansion within 3D soft microenvironments supported cellular stemness of embryonic stem cells (ESCs) even without pluripotent media, and it improved epigenetic reprogramming efficiency of mouse embryonic fibroblasts. CloneSeq of ESCs revealed that the differentiation decision is made early during *Oct4* downregulation and is maintained during early clonal expansion. Together, we show CloneSeq can be adapted to different biological systems to discover rare subpopulations by leveraging the enhanced sensitivity within clones.

INTRODUCTION

Single-cell studies have revealed considerable cell-to-cell variation within tumors of different cancer types (Dalerba et al., 2011; Kim et al., 2015; Tirosh et al., 2016; Li et al., 2017) and during embryonic stem cell (ESC) differentiation (Yan et al., 2013; Klein et al., 2015). For example, single cells derived from glioblastomas have inherent variation in their transcriptional expression (Patel et al., 2014), and lung adenocarcinoma cells have heterogeneity of immune-response-related gene expression (Ma et al., 2019). Other cancer-related single-cell studies have characterized infiltrating immune cells such as T cells, which have furthered our understanding of their heterogeneous organization, clonal expansion, migration, and functional-state transitions (Chung et al., 2017). As cells from cancerous tumors have an assortment of cellular mutations that give rise to cells with the ability to de-differentiate, tumors are highly heterogeneous, making single-cell-based profiling a powerful tool to dissect the underlying cellular structures. Similarly, investigations of early development and differentiation of stem cells greatly benefit from single-cell resolution approaches. ESCs grown *in vitro* perpetuate the broad developmental potential of naive founder cells in the pre-implantation embryo (Martello and Smith, 2014). ESCs are composed

of cells in different states and differentiation potentials (Evans and Kaufman, 1981); however, limitations in the sensitivity of single-cell technologies hinder our ability to understand the fine-tuning of cellular hierarchies. Technological advances such as high-resolution cell imaging (Raj and van Oudenaarden, 2009) and single-cell profiling of epigenomic and genomic sequences (Litzenburger et al., 2017; Chen et al., 2019) suggest that cellular heterogeneity results from more than just a mixture of different cell types. It appears that a given cell type can be composed of a subtle assortment of cells with different states. These different states allow cells to adjust to changing conditions by committing to a certain differentiation trajectory or simply minimizing metabolism. This, in turn, appears to enhance drug resistance development, as recently suggested for acute myeloid leukemia cells (Pan et al., 2014).

Stochasticity, or randomness, is a strong component in the accumulation of cellular variation (Losick and Desplan, 2008). Stochastic effects are very difficult to study, as in many cases they describe chaotic processes that we do not understand and that are difficult to separate from biological and technical noise. Stochastic effects and noise often confound single-cell measurements. For example, the nonlinearity of transcription can lead to errors in clustering of data (Dar et al., 2012).



Cell-cycle regulation is an important biological feature, but if cells are in different stages of the cell cycle, alterations in cell-cycle regulators and cycling genes can mask subtle differences that determine distinct cellular states (Nitzan et al., 2019). Uneven culture conditions, in terms of the distribution of reagents and oxygen as well as differences in surface tension and elasticity, can affect cellular outcomes. Moreover, the presence of inevitable technical variability introduced during sample processing steps also causes batch effects (Tung et al., 2017). Finally, single-cell profiling techniques inherently suffer from low sensitivity that can lead to false-negative and false-positive results (Ziegenhain et al., 2017). These confounders can strongly influence the level of randomness attributed to measurements, and thus, single-cell experimental data are highly noisy and difficult to interpret, especially in the context of cellular states.

To overcome these hurdles, we developed CloneSeq, a 3D clone-based RNA-seq approach. Our hypothesis was that clones are composed of cells more similar to each other than cells picked at random, and that analysis of clonal cells would have improved sensitivity and coverage relative to single-cell RNA-seq (scRNA-seq). Our results support this hypothesis, as cells originating from a given clone had more similar transcriptional profiles than cells across clones. The small clones in our 3D system also had detectably different phenotypes. We leveraged this observation to perform an in-depth dissection of cellular heterogeneity in lung adenocarcinoma PC9 cells. We were able to characterize different cellular states including cancer stem-like cells (CSCs), high- and low-replicative cancer cellular states, and different levels of invasiveness. Such features cannot be detected using scRNA-seq due to its low mapping resolution. As our 3D culturing method supports cancer cell growth and nourishes cellular stemness, it could be optimized for primary tumor cell expansion. Finally, we show that this 3D system induces ESC formation without standard supplements, such as leukemia inhibitory factor (LIF) and MEK and GSK3 inhibitors (2i) and improves efficiency of induced pluripotent stem cell (iPSC) production, making it superior to standard ESC culturing methods.

Design

Low sensitivity and stochasticity limited the power of scRNA-seq technology. The low sensitivity creates confounder effects that might lead to misinterpretation or masking of important information. To answer these inherent defaults, we have developed CloneSeq. With this approach, we produced small clones from single cells in 3D hydrogels and profiled their clonal transcriptome. We found CloneSeq provided better sensitivity than scRNA-seq and showed its application to cancer cells and ESCs differentiation.

RESULTS

Developing a 3D hydrogel cell culture system

To support the expansion of single cells into small clones in a confined, controlled, and robust setting, we developed CloneSeq, a method that supports spherical 3D tissue culturing and sequencing. The method includes three steps: (1) capture of single cells inside soft, biocompatible, and biodegradable hydrogel spheres using a microfluidic device; (2) clonal expansion of sin-

gle cells within the hydrogel spheres up to 60 cells, depending on cell type and size; and (3) single-clone RNA-seq by uniquely barcoding each clone inside nanoliter droplets. Since single cells are allowed to expand within the hydrogel spheres while maintaining inter-clonal cell states, single-clone transcriptional profiling facilitates the detection of a large number of genes (increased coverage), including genes that are lowly expressed (increased sensitivity), without averaging out transcriptomes as in bulk RNA-seq.

We optimized a microfluidic architecture for capturing single cells within 3D hydrogel spheres (Figure 1A). The inflow of cells and maleimide-dextran (MALDEX) precursor are mixed with polyethylene glycol dithiol (PEGDT) at the first junction (Figure 1A1). Small, nanoliter aqueous MALDEX-PEGDT droplets are formed via oil-phase “pinching” (Table S1). MALDEX polymerization and PEGDT crosslinking occurs spontaneously via thiol-maleimide click chemistry without reacting with the cells (Nuttelman et al., 2001). Gelation is completed within 10 s (Li et al., 2014). The cured gels are then immersed in culture medium (Figure 1A2), permitting the diffusion of growth factors and supporting the proliferation of encapsulated cells.

To verify structural homogeneity and uniformity, we modified the MALDEX backbone with thiolated biotin and stained the hydrogel spheres with rhodamine-conjugated streptavidin (streptavidin-RU). The high affinity of the biotin-streptavidin interaction ensures that once formed, the specific staining will not be affected by changes in pH or rinsing (Diamandis and Christopoulos, 1991). Indeed, confocal microscopy cross-sectional imaging revealed a uniform distribution of streptavidin-RU within the PEGDT-MALDEX hydrogel spheres (Figure 1B). Next, we characterized the mechanical properties of the PEGDT-MALDX hydrogel spheres using micropipette aspiration (Hochmuth, 2000). The spheres deformed elastically in response to applied stresses, reaching a steady-state aspiration length that correlated linearly with the applied intra-pipette pressure (Figure 1D). Using the homogeneous half-space model approximation (Theret et al., 1988), we calculated the stiffness of the hydrogel spheres. The elasticity of the spheres was 3 kPa, which is consistent with the microelasticity of soft tissues such as fat and kidney (Bainer and Weaver, 2013).

Owing to their dextran backbone, the PEGDT-MALDEX hydrogel spheres can be remodeled by the encapsulated cells. The spheres supported the viability and proliferation of several cancer cell lines, including PC9, without extracellular adhesion ligands (Figure 1E). PC9 clones consisted of 12 and 15 cells on an average after 7 and 8 days in culture, respectively (Figures 1E and S1A), and maintained the viability of 87% of the cells within the hydrogel (Figure 1F). The expansion of PC9 clones increased the effective stiffness of the hydrogel spheres from 3 to 6 kPa, which is comparable with lung tissue mechanics, consistent with PC9 tissue origin (Figure 1D) (Bainer and Weaver, 2013). The increase in the effective stiffness of the spheres is attributed to the internal pre-stress that is generated by the contractile cells of the encapsulated clone.

Unlike PC9 cells, supporting the viability and proliferation of ESCs within the hydrogel spheres required the insertion of cell adhesion signals and cleavable properties by cell-secreted metalloproteinase (Burdick and Murphy, 2012; Almalki and Agrawal, 2016). Hence, we encapsulated ESCs within hydrogel spheres

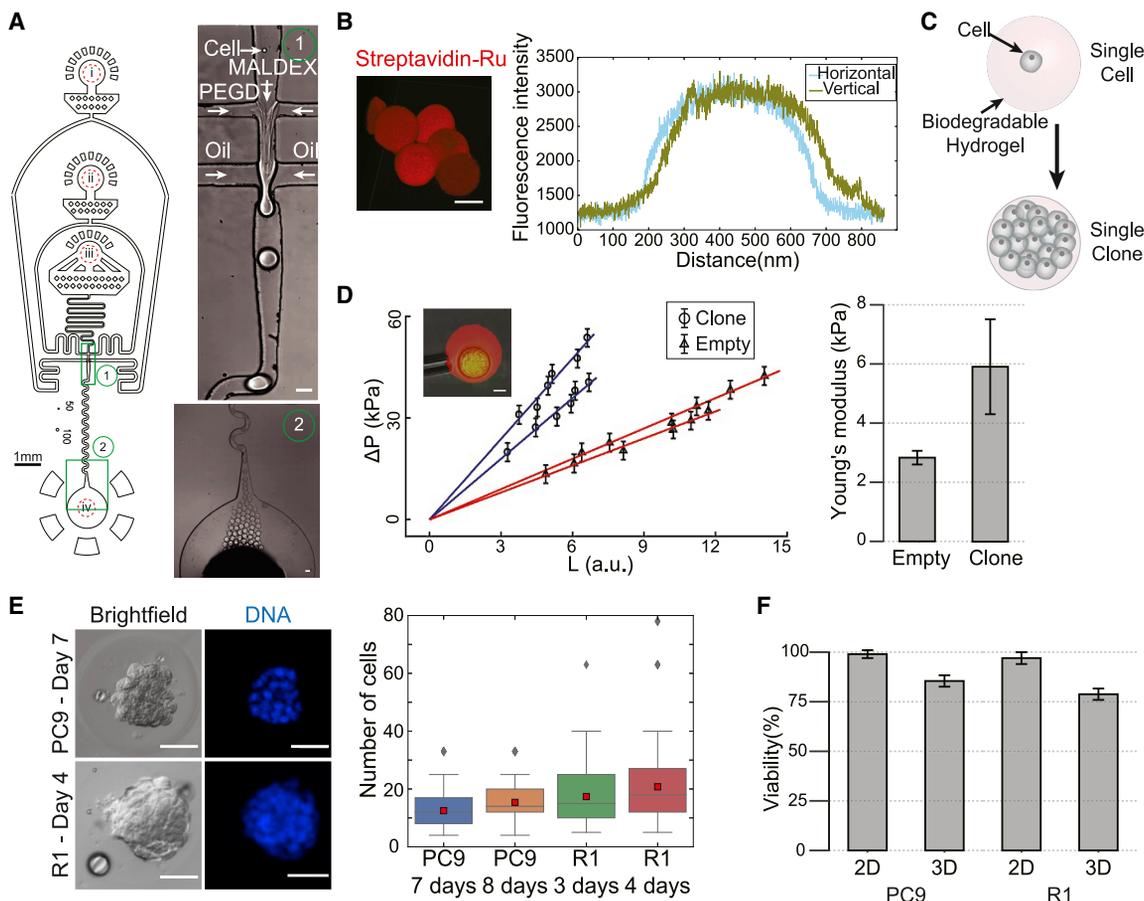


Figure 1. Cell encapsulation, clone formation, and characterization

(A) Clone encapsulation. Single cells were encapsulated within biodegradable hydrogel spheres using a microfluidic device (left). The microfluidic device consists of (i) a carrier-oil inlet, (ii) a PEGDT inlet, (iii) a cell-MALDEX precursor mix inlet, and (iv) the droplet collection outlet. Zoom-in of (1) encapsulation and (2) outlet regions are shown on the right (scale bar, 50 μm).

(B) Confocal imaging (left, scale bar, 50 μm) and fluorescence cross-sectional profiles (right) of the streptavidin-RU staining show round shapes with equal density across the hydrogel spheres.

(C) PEGDT-MALDEX hydrogels support the proliferation of encapsulated cells and clone formation.

(D) Left: the mechanical properties of empty hydrogel spheres and clones were evaluated using micropipette aspiration. Scale bars, 10 μm . Left: the aspirated length (L) of empty hydrogel spheres ($n = 2$) or a clone ($n = 2$) increases linearly with applied suction pressure (ΔP) indicative of a purely elastic response. Right: elasticities (as Young's modulus) calculated based on pipet aspiration test ($n = 2$). Hydrogel spheres encapsulating clones of PC9 cells are stiffer than empty hydrogel spheres.

(E) Left: bright-field and fluorescence microscopy images show the formation of 3D clones of PC9 cells and R1 ESCs. Scale bars: 30 μm . Right: number of cells per clone of PC9 and R1.

(F) Viability comparison of PC9 and R1 ES cells cultured on gelatin-coated plates (2D) and inside 3D hydrogel spheres (3D). $n = 2$; all bar plots are represented as mean \pm SEM. See also [Figure S1](#) and [Videos S1](#) and [S2](#).

supplemented with thiolated RGD peptides that mediate cell adhesion to the gel via integrin transmembrane receptors ([Burdick and Murphy, 2012](#)). Additionally, the PEGDT crosslinker was replaced by a dithiolated PEG-peptide conjugate (MPEGDT) crosslinker, using an amino acid sequence motif (PLGLWA) that serves as a cleavable site for matrix metalloproteinase (MMP; [Burdick and Murphy, 2012](#); [Almalki and Agrawal, 2016](#)). Indeed, the MPEGDT-MALDEX hydrogel spheres supported ESC proliferation and the generation of ESC clones consisting of 15 and 20 cells on an average after 3 and 4 days in culture, respectively ([Figure 1E](#)). Specifically, after 4 days ESCs reached clone size of approximately 50% of the hydrogel diameter, while maintaining the viability of 77% of the cells ([Figure 1F](#)).

CloneSeq: Profiling of clones using modified inDrops protocol

For mRNA profiling of clones, we designed a microfluidic device to capture clones in drops and barcode their mRNAs using a modified inDrops protocol ([Zilionis et al., 2017](#)) ([Figure 2A](#)). The device consists of two junctions: one that combines the lysis, barcodes, and clone suspension, and another junction for encapsulating the aqueous inputs in the oil phase. To extract mRNA from the clones, hydrogel spheres were dissolved using dextranase. During the development of the method, we found that the droplet-based reverse transcription (RT) reaction used in the standard inDrops ([Klein et al., 2015](#)) protocol was inhibited in the presence of dextranase ([Figure S2](#)). However, an attempt

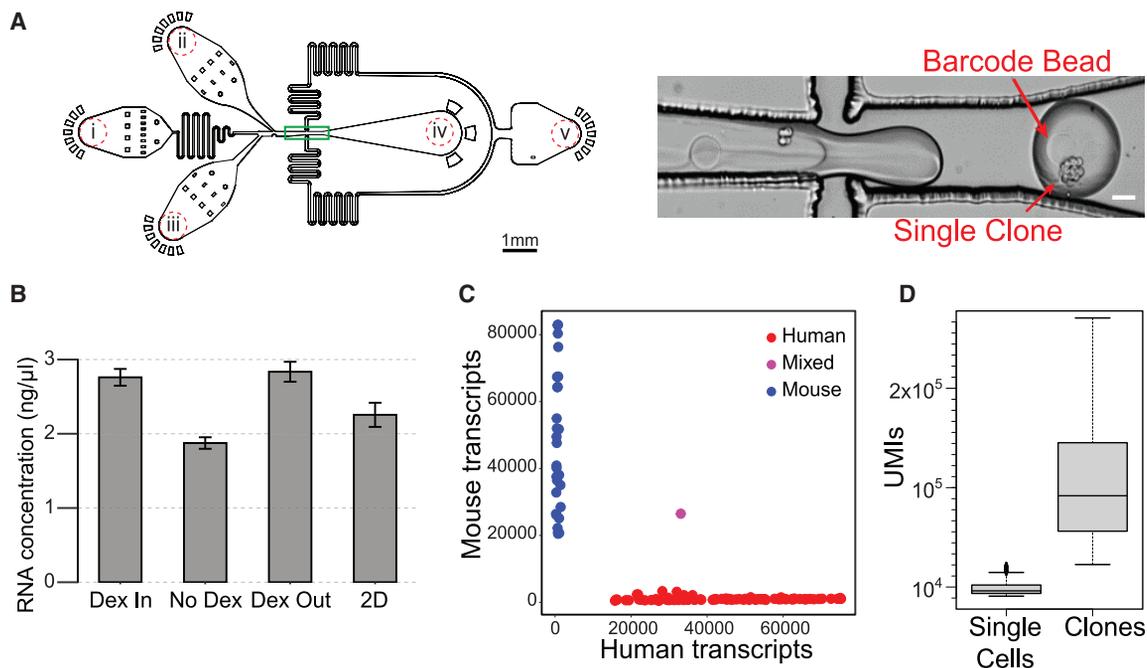


Figure 2. CloneSeq increases sensitivity and coverage of transcriptional profiling

(A) Total RNA barcoding of encapsulated clones was performed within a modified inDrop-based microfluidic configuration. Zoom-in of the encapsulation junction (green frame) shows a single clone co-encapsulated with a barcoded bead immersed in lysis buffer.

(B) Optimization of mRNA extraction from PC9 cells in bulk. mRNA was extracted from cells that were embedded inside a PEGDT-MALDEX hydrogel using lysis buffer supplemented with dextranase (*Dex In*), using lysis buffer only (*No Dex*), and following a two-step approach in which dextranase was first used for degrading the gels and releasing the cells followed by lysis buffer (*Dex Out*). As a control, mRNA was also extracted from a standard 2D culture of PC9 cells (*2D*). Cell number was maintained equal for all conditions, $n = 2$.

(C) CloneSeq of human and mouse clones. The scatter plot shows the number of transcripts associated to each barcode. Blue and red dots indicate human- and mouse-specific transcripts, while the single purple dot indicates a mixed association.

(D) The number of non-redundant transcripts with UMIs is 10-fold greater compared with standard inDrop single-cell RNA-seq. Data are represented as mean \pm SEM. See also [Figure S2](#); [Video S3](#).

to lyse the cells within the hydrogel spheres without dextranase, which would have allowed mRNA to diffuse out of the hydrogel spheres through pores and bind the barcodes inside drops, significantly reduced the number of RNA molecules captured ([Figure 2B](#)).

Next, we tested the Drop-seq protocol ([Macosko et al., 2015](#)), in which cell lysis and mRNA-capture occurs inside drops and the RT reaction occurs after droplet breakup. In this protocol, the dextranase is washed out during the breakup of the droplet before RT takes place. However, we observed that Drop-seq of clones did not produce significantly higher numbers of transcripts compared with Drop-seq of single cells. We suspect that this limitation is due to the fact that each commercially available barcoded bead contains a fixed number of barcoded primers ($\sim 10^6$) ([Macosko et al., 2015](#)).

To overcome the RT inhibition in the inDrops protocol, we took advantage of the observation that when cells grow inside hydrogel spheres, they tend to form highly adhesive and stable structures and remain as spheroids even after the surrounding hydrogel spheres have been dissolved. Therefore, before the clones were introduced into the microfluidics apparatus, the hydrogel spheres were dissolved with dextranase and hydrogel remnants were further removed by a filter ([Figure 2A](#)). This inDrops-based protocol also allowed us to increase the number

of barcoded primers by increasing the diameter of the polyacrylamide beads and/or increasing primer concentration as previously described ([Macosko et al., 2015](#)). The number of barcoded primers per bead was increased to 10^9 for sequencing of clone transcriptomes.

To ensure the purity of single-clone encapsulation by CloneSeq, we mixed clones from human and mouse cell lines (human PC9 cells and mouse R1 ESCs grown for 7 and 4 days, respectively, in hydrogel spheres) at a concentration of 20,000 and 40,000 clones per ml with flow rates designed to capture one clone per 2 s ([Figure S3](#); [Table S1](#)). Both microscopy and a Barnyard (human/mouse) mixing plot of sequencing data showed that at a concentration of 20,000 clones/ml, the encapsulation resulted in excellent separation, with only $\sim 3\%$ of barcodes containing mixed reads, whereas at a higher concentration ($>40,000$ clones/ml), 10% or more of the reads were mixed ([Figures 2C](#) and [S4](#)). These results align with a previous analysis of single-cell inDrops, which showed that reducing cell concentration decreased the likelihood of collecting two cells in one drop and decreased the level of contamination of drops with free mRNA ([Zilionis et al., 2017](#)).

Next, we performed scRNA-seq to an average of $\sim 100,000$ reads per cell and compared the data to the equivalent sequencing coverage applied for the clones. After reducing

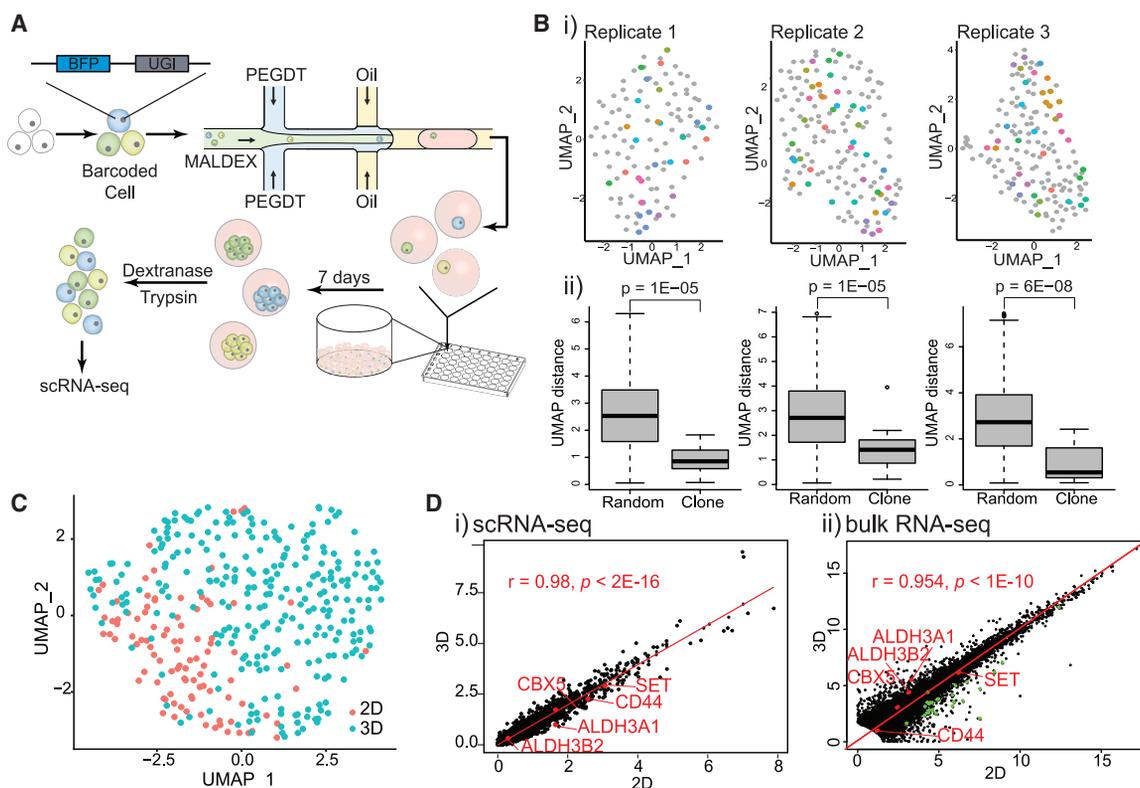


Figure 3. 3D culturing maintains intracolonial similarities and supports cell “stemness”

(A) PC9 cells were genomically barcoded using lentiviral transfection of 10 basepairs unique guide indices (UGI). Barcoded cells were encapsulated inside the hydrogel spheres and the expanded clones were dissociated, single cells were picked at random, and subjected to scRNA-seq.

(B) (i) UMAP projection of single-cell transcriptomes of 113, 154, and 152 PC9 cells, with identified 30, 41, and 37 cells originating from 9, 17, and 16 distinct clones in the three replicates, respectively. (ii) Cell-to-cell UMAP Euclidean distances between cells sharing clonal origin and random pairs. Data are represented as mean \pm SEM.

(C) Single-cell transcriptomes of 272 cells from 3D hydrogel spheres (3D) and 97 cells from a standard 2D culture (2D) were mapped into one cluster on a UMAP space.

(D) A correlation between the gene expression levels of cells cultured in standard 2D and in 3D configurations is maintained both via (i) scRNA-seq and (ii) bulk RNA-seq. Bulk RNA-seq sensitivity reveals stemness signature that are supported by 3D cultures (red dots) and cell-cycle signatures in 2D (green dots). R, Pearson coefficient of correlation. See also [Table S3](#).

PCR duplicates using unique molecular identifier (UMI) counts, we observed that the number of UMIs retrieved from each clone was significantly higher than the numbers retrieved from single cells, with an average number of UMIs of $\sim 100,000$ for clones and $\sim 10,000$ for single cells (Figure 2D). The values for the clones were lower than expected by simple clone-cell-number extrapolations. This is explained by both insufficient primer numbers and reaction inhibition due to the high concentration of cell debris in each drop that originates from the large number of cells composing each clone.

Impact of 3D hydrogel culture on PC9 cells

To evaluate the impact of our 3D culture method on the cells, we first assessed clone homogeneity. We wished to determine whether the cells within the clone are similar enough to each other to allow us to consider the clone an entity representing the original mother cell. We then compared the expression profile of cells cultured inside the 3D hydrogel spheres to cells grown in 2D to determine whether the 3D culture itself altered the cell state. We produced PC9 cells carrying genomic barcodes

located 100 bp upstream to the *BFP* polyA signal (Jaitin et al., 2016). A sequencing library from approximately 50,000 colonies showed that our plasmid pool contained approximately 20,000 unique barcodes. Barcodes exceeding normal distribution were registered as outcasts for later computational analysis. In these experiments, lentiviruses were transduced into cells with a multiplicity of infection (MOI) of 0.05 so that the majority of cells will only retrieve one virus, hence one barcode. We then FACS-sorted BFP-positive cells to collect only virus-containing cells. Next, we encapsulated barcoded PC9 cells into 3D hydrogel spheres and grew them until each clone contained 10–30 cells. We then randomly selected about 2,000 clones, dissolved the hydrogel spheres, dissociated the clones to obtain single cells, and performed scRNA-seq (Figure 3A). The limitation of 2,000 clones was due to the statistical restraints of the 20,000 clonal barcodes. Cells of the same clone were identified by their identical genomic barcodes. We performed the experiment on three groups of randomly chosen clones, each group contained ~ 150 cells. After filtering out cells with unreliable barcodes, for the three replicates, we extracted 30, 41, and 37 cells originating

from 9, 17, and 16 distinct clones, respectively. The similarity between cell expression profiles was analyzed using UMAP and quantified by calculating the Euclidean distance between points on the UMAP plots (Figure 3B). The Euclidean distance between random cells was determined by evaluating all possible cell pairs, and clone distance was calculated as the distance between cells from the same clone. The distance between cells from the same clone was significantly smaller than the distance between randomly selected cells in the reduction projection space, with a *p* value of 1E-05, 1E-05, and 6E-08, respectively. This suggests that cells of the same clone are indeed significantly more similar to each other than to cells selected at random.

To further evaluate the impact of 3D culturing, we compared PC9 cells growing in 2D versus 3D conditions, using both scRNA-seq and bulk RNA-seq. Single-cell projections captured small differences between the 2D and the 3D conditions (Figure 3C). We identified 17 downregulated genes and 106 upregulated genes in the 3D condition (FDR < 0.05, Student's *t* test; Table S3). Gene set enrichment analysis of the downregulated genes in the 2D conditions showed no significant GO terms, whereas the upregulated genes enriched in 3D conditions revealed pathways related to RNA metabolism and cell cycle and growth, which aligned with the observation that 3D culturing enhances cellular proliferation.

In support, gene expression comparison of scRNA-seq showed a highly significant correlation coefficient (0.98; *p* < 1e-16), similar to the overall correlation observed for bulk RNA-seq (0.95; *p* < 1e-10), but the latter also revealed the upregulation of cell-cycle-related genes in 2D (green dots, Figure 3D). In contrast, bulk RNA-seq revealed that 3D culture led to the induction of genes associated with adhesiveness and cancer stem-cell-like signature (Ma and Allan, 2011) (e.g., *Aldh3A1* and *Aldh3A2*). Other genes known to be associated with CSC signature (*SET* [Edupuganti et al., 2017]), *CBX5* [Yu et al., 2012], and *CD44* [Leung et al., 2010]) were found equally expressed in both 2D and 3D culture conditions. As these genes are not highly expressed in PC9 cells, they could not be detected by scRNA-seq. Taken together, these data suggest that whereas cells grown in 3D culture in PEGDT-MALDEX hydrogel are overall similar to those grown in 2D conditions, there are differences in expression of some genes associated with cancer stemness in PC9 cells.

Clone-to-clone variation identifies cancer stem-like subpopulations

To study the effect of clonal expansion within the hydrogel spheres on cell state homogeneity within the clones, we compared the inter-clone correlations of small (*n* < 15 cells) and large (*n* ≥ 15 cells) clones and of pseudo-clones that were formed *in silico* by averaging randomly sampled single-cell transcriptomes. The transcriptional correlation between pseudo-clones that were aggregated *in silico* increases due to averaging out of pseudo-clone-to-pseudo-clone variation with increasing pseudo-clone size, plateauing at *n* = 8 cells, while the correlation among real clones remains relatively low and clone-to-clone variation is not averaged out both for small clones (*n* ≤ 15 cells) and for big clones (*n* > 15 cells). Though the correlation of small clones was lower than for large clones, they were both comparable with

2-cell pseudo-clones and significantly lower than that of pseudo-clones of similar size. This indicates that clonal expansion within the hydrogel spheres maintains cell state and hinders cell-to-cell variation compared with the transcriptomes of non-clonal cells. Importantly, this variation is not the outcome of cell-cycle states or transcriptional bursting, as those are averaged in clones. In this manner, CloneSeq amplifies sequencing sensitivity and coverage by reading the transcripts that were pooled from multiple cells while excluding averaging out distinct cellular states.

To evaluate the association between single-clone and single-cell transcriptomes, we cultured PC9 clones for 7 days and performed CloneSeq for 1,702 clones and scRNA-seq for 1,328 single cells that were dissociated from clones. To define possible improvement in CloneSeq sensitivity for capturing gene-specific information, we applied different thresholds and calculated the percentage of genes, which passed each filter. Next, we compared the percentage of genes that has at least one cell or clone that supports their expression. We found that 55% and 66% of the genes are expressed in at least one cell and clone, respectively. As we raised the threshold, clones kept showing higher coverage compared with single cells, with genes that are expressed in at least 100 cells covered only 14% of the genes, whereas 36% of the genes were captured in clones (Figure 4B).

For CloneSeq, four clusters were identified and projected onto a dimensionally reduced UMAP space, while only two clusters were identified for scRNA-seq (Seurat package 3.0; Figure 4C; Butler et al., 2018; Stuart et al., 2019). Cluster identities were characterized by identifying differentially expressed genes for each cluster using a Wilcoxon rank sum test, then performing gene enrichment analysis using MSigDB (Subramanian et al., 2005; Figure 4D, full list of genes is in Table S4). Cluster "0" is enriched for responses to endoplasmic reticulum (ER) stress and unfolded proteins, which is known to promote tumorigenesis (Chen and Cubillos-Ruiz, 2021). Notably, lncRNA *MALAT1*, which plays a key role in lung cancer metastasis (Gutschner et al., 2013), is highly expressed in cluster 0. *MALAT1*, also reportedly regulated by unfolded protein responses (URP) during ER stress (Bhattacharyya and Vrtati, 2015), aligned with the enriched GO term identified for cluster 0. Cluster "1" is enriched for lipid and sterol biosynthesis and metabolism, previously recognized as hallmarks for identification of CSC (Yi et al., 2018). In addition, key genes for fatty acid synthesis, *FASN* (Yasumoto et al., 2016), *ACLY* (Migita et al., 2008), and *SCD* (Noto et al., 2017) and cholesterol metabolism, *HMGCR*, *FDFT1*, *SQLE*, *MSMO1*, *DHCR7*, and *DHCR24* (Yang et al., 2020) are identified as marker genes for cluster 1 (Figure 4E).

More general functions of tumors such as proliferation, cell cycle, and cell mobility functions are enriched for clusters "2" and "3" (Figure 4E). Intriguingly, many CSC markers were found in cluster 3, with a significant signature for epithelial-to-mesenchymal transition (EMT; Figure 4E). EMT is associated with dedifferentiation of cancer cells and contributes to the acquisition and maintenance of stem-cell-like characteristics (Liu and Fan, 2015; Wang and Unternaehrer, 2019). *TOP2A*, *AURKB*, and *CDK1* were identified as specific markers for stem-like populations in NSCLC (Perumal et al., 2012). *PTTG1* oncogene is reported to promote tumor malignancy via EMT in breast cancer (Li et al., 2013) and to promote migration and invasion of NSCLC (Li et al., 2013). *MK167(Ki-67)* expression in NSCLC indicates a

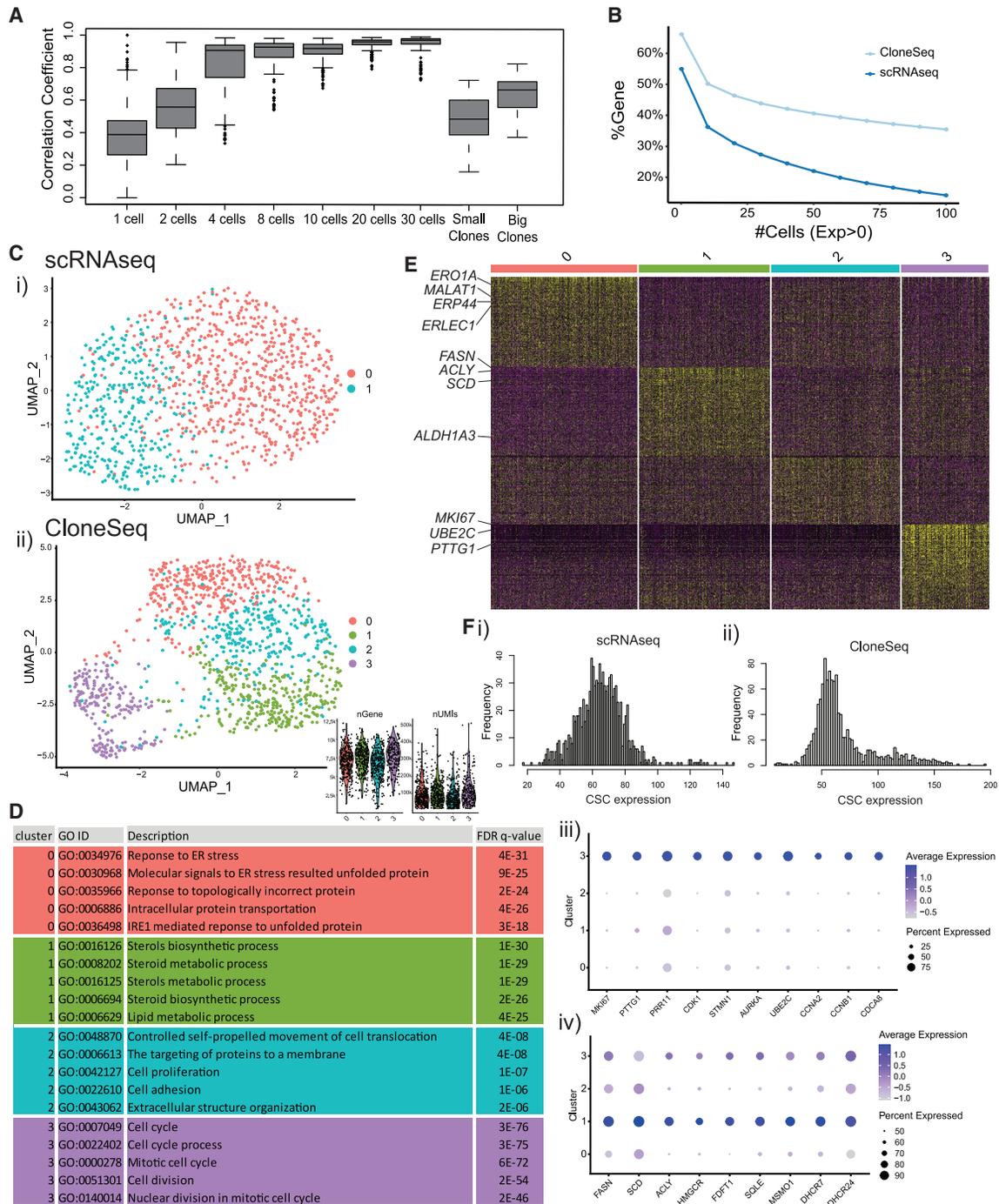


Figure 4. Clone-to-clone variation analysis for PC9 cells

(A) Correlation coefficients between cells or clones, including *in silico* pseudo-clones of different sizes (2–30 cells), small ($n < 15$ cells), and large ($n \geq 15$ cells) clones. Data are represented as mean \pm SEM.

(B) The percentage of genes passing filters based on the number of cells with positive expression for CloneSeq and scRNA-seq.

(C) Clones at day-7 after encapsulation were transcriptionally profiled either after dissociation into (i) single cells ($n = 1,328$, UMI $> 5,000$) or as (ii) clones ($n = 1,702$, UMI $> 10,000$) and mapped onto UMAP space.

(D) Underlying distinctive MSigDB GO term and cell function annotations with FDR q-value for each enrichment.

(E) Differentially expressed genes are identified via KNN unsupervised clustering of CloneSeq transcriptomes.

(F) Distribution of overall expression levels of CSC-related genes per cell in (i) scRNA-seq is normalized and in (ii) CloneSeq is tailed. Relative expression levels of a set of CSC genes in CloneSeq data (iii) EMT process and (iv) lipid and sterol metabolisms showed enrichment in clusters 3 and 1, respectively. The size of the dot reflects the percentage of cells expressing the markers, while the color encodes average expression levels across all cells within the group. See also Table S4.

bad survival prognosis (Martin et al., 2004) and it is required for maintenance of CSC niche in breast and colon cancers (Cidado et al., 2016). *PRR11* is reported to hold a critical role in tumorigenesis of human lung cancer (Ji et al., 2013) and reportedly maintained gastric CSCs self-renewal (Hu et al., 2018). *CDK1* has also been shown to mediate cancer stemness through the *CDK1-SOX2* axis (Boumahdi et al., 2014; Menon and Fujita, 2019). *STMN1* was found to be related to NSCLC progression (Nie et al., 2015) and its expression was found to be associated with CSC marker-gene expression in breast cancer (Obayashi et al., 2017) and hepatocellular carcinoma (Zhang et al., 2020; Wu et al., 2016; Yang et al., 2016; Chen et al., 2017). Overall, we identified two subpopulations of PC9-related CSCs with different functions, one enriched for lipid metabolism (cluster 1) and the second for EMT (cluster 3; Figure 4F). Finally, to validate that CSCs cannot be properly detected using single-cell data, we produced a histogram showing a normal distribution in single cells and a bimodal distribution in clones (Figure 4F). Overall, CloneSeq provides means for elucidating the existence of distinct cell states within a population of cellular states, which is not accommodated by single-cell RNA-seq methods.

3D hydrogel spheres support pluripotency and differentiation

As our 3D culture method induced proliferation of CSCs, we hypothesized that the 3D conditions within the gels may promote stemness. To test this, we explored the effect of 3D culture conditions on ESCs. Ground-state pluripotency can be maintained *in vitro* by culturing the cells with LIF together with GSK3/MEK inhibitors (“2i”); Ying et al., 2008; Schlesinger and Meshorer, 2019). The microenvironment in which the cells grow influences the cell state and can activate or repress differentiation pathways (Ying and Smith, 2017). ESCs grown without 2i/LIF do not retain pluripotency and differentiate spontaneously along the different lineages (ectoderm, endoderm, mesoderm, and extra-embryonic endoderm) (Ying et al., 2008).

To validate the effect of the 3D environment on stemness, we used BYKE ESCs from the Buganim laboratory, which express three pluripotency reporters: GFP-NANOG and BFP-ESRRB as markers for ground-state pluripotency and RFP-UTF1 for the primed state (Bencherit et al., 2019). As expected, ESCs grown in both our 3D system and the 2D gold-standard (0.1% gelatin) setup in the presence of 2i/LIF expressed all three pluripotent markers (Figure 5A). However, when the cells were grown without 2i and LIF for 6 days, cells in the 2D system quickly lost their pluripotency markers, whereas cells growing in 3D continued to express these markers, with NANOG and ESRRB slightly downregulated and UTF1 slightly upregulated. This suggests that our 3D hydrogel spheres support pluripotency and significantly delay ESC spontaneous differentiation in the absence of 2i/LIF, possibly explaining the highly homogeneous states observed for the ESC clones grown in 3D. To further validate the pluripotency of the cells after 8 days in the absence of 2i/LIF, we removed the hydrogel spheres and reseeded the cells on mouse embryonic fibroblasts (MEFs) for 4 days. In cells originating from the hydrogel spheres, about 70% of the cells formed NANOG-positive colonies, compared with less than 8% of the cells originating from the 2D condition (Figure 5B). To corroborate these findings, we performed RNA-seq on cells grown in 2D and 3D conditions for

4 days without 2i/LIF. The results confirm higher expression levels of pluripotent transcription factors in cells grown in hydrogel spheres compared with cells grown on gelatin (Figure 5C). This demonstrates that the 3D hydrogel spheres support the maintenance of the ESC pluripotent state.

We then tested the differentiation potential of ESCs cultured in 3D hydrogels. We encapsulated single ESCs in hydrogel spheres and cultured these cells for 4 days with and without 2i/LIF. Next, we dissolved the gel and dissociated the clones back to a 2D tissue culture plate. We differentiated the ESCs for 4 days with retinoic acid (RA) in three biological replicates (Figure 5D). The bulk mRNA sequencing results revealed that with or without 2i/LIF, ESCs could efficiently exit the pluripotent state (downregulating *Nanog*, *Esrrb*, and *Klf4*). Interestingly, the differentiation outcome was not identical. The ESCs that grew without 2i/LIF differentiated primarily toward the endodermal lineage (GO:0035987, p adjust = $5E-03$), while the group of ESCs cultured with 2i/LIF showed a strong propensity toward the ectodermal lineage, upregulating genes involved in neuronal development (GO:0022008, p adjust = $5E-08$). The difference in the differentiation potential may suggest that ESCs cultured in 3D hydrogels without supporting conditions are in a primed state, whereas ESCs growing in the presence of 2i/LIF are in a ground state, supporting our previous observations (Figures 5A and 5B). Therefore, ESCs cultured without 2i/LIF present a faster differentiation timeline, activating the endodermal differentiation program earlier than ESCs grown with 2i/LIF. Importantly, when we performed bulk RNA-seq, the differences between the two culturing conditions were merely a reflection of the proportion of cells in each differentiation state. Overall, PEG-DEX hydrogels support ESC pluripotency in the absence of pluripotency supporting factors. This method affects the differentiation potential of ESCs *in vitro* and can be further tested on other differentiation assays.

Finally, we tested whether the 3D microenvironment enhances the reprogramming efficiency of MEFs into induced pluripotent stem cells (iPSCs). We compared MEFs containing the OSKM (*Oct4*, *Sox2*, *Klf4*, and *Myc*) cassette under the control of the TET-on promoter (Carey et al., 2010) grown in standard iPSC conditions in 2D, to the same MEFs grown inside our 3D hydrogel system (Figure 5S). After 10 days, during which the cells were supplemented with 4.5 μ M tetracycline, we seeded the cells on MEFs and quantified NANOG-positive colonies. Reassuringly, the 3D culture showed \sim 2.4-fold higher frequency of NANOG-positive cells compared with the 2D cultures (Figure 5E). One clear disadvantage of reprogramming in the 3D setup is that reprogramming is accompanied by a high number of dead cells that failed to complete their cellular transformation (Cheung et al., 2012). These apoptotic cells accumulate inside the hydrogel spheres and cannot be washed away easily. As apoptotic cells secrete signals that interfere with iPSC formation, we predict that iPSC production efficiency could be significantly improved by adding cycles of breaking down the hydrogel spheres, removing dead cells, and re-encapsulating MEFs.

Differentiation decisions following *Oct4* downregulation are maintained during clonal expansion

Each single ESC has the potential to differentiate to all cellular lineages, however, it is not entirely known how a differentiation decision of a single ESC will affect the derived clonal output. In

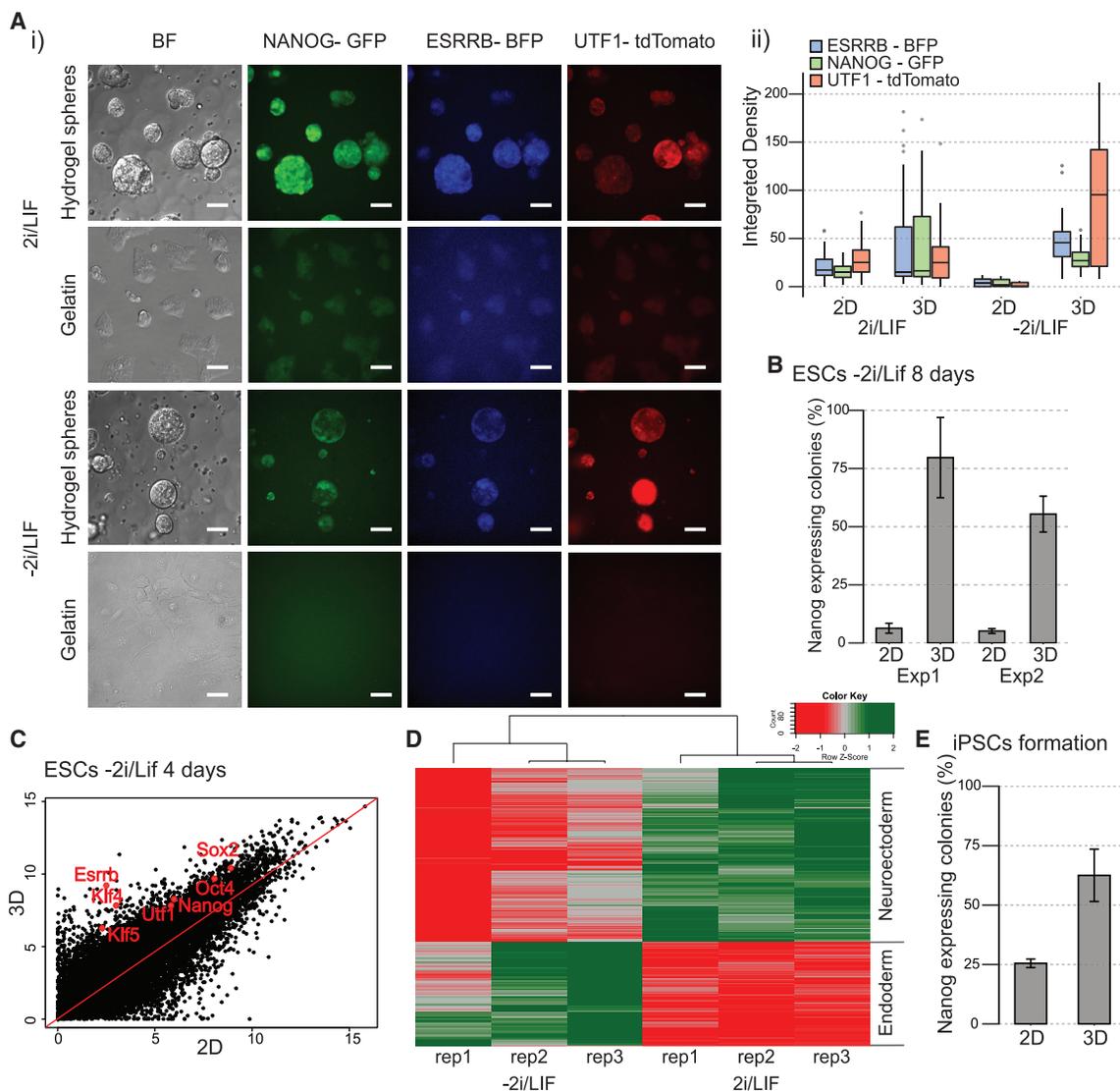


Figure 5. 3D soft hydrogels support a pluripotent state and differentiation

(A) (i) Representative NANOG, ESRRB, and UTF1 confocal images of 3D BYKE1 ESC clones encapsulated inside hydrogel spheres and 2D ESC colonies cultured on gelatin-coated plates. Cells were cultured for six days with or without 2i/LIF supporting factors. Scale bars: 50 μ m. (ii) Averaged fluorescence densities of each image are presented in a box plot.

(B) The percentage of NANOG-positive colonies compared between ESCs that were cultured inside 3D soft hydrogel spheres and on gelatin-coated 2D plates for 8 days, dissociated, and further cultured on MEF feeder layer for subsequent 4 days (average of three representative regions in the plate). p value < 1e-03.

(C) Bulk RNA-seq results of ESCs grown in 3D and 2D in absence of supporting factors for 4 days. Representative pluripotent genes are marked in red.

(D) Differentiation potential comparison of ESCs cultured for 4 days in 3D hydrogel with or without pluripotency supporting media (2i/LIF) and then moved to 2D-based culturing in the presence of RA for an additional 4 days. The heatmap highlights differentially expressed genes upregulated in each condition after differentiation.

(E) The percentage of NANOG-positive colonies compared between ESCs from 3D and 2D culture and during 10 days of reprogramming. All bar plots and box plots are represented as mean \pm SEM. See also Figure S5.

order to test this, we downregulated *Oct4* expression for 24 h, in the presence of 2i/LIF, using a TET-off *Oct4* mouse ESC line (Niwa et al., 2000). We validated that *Oct4* was no longer expressed upon addition of doxycycline (Dox) by western blot (Figure 6A). We validated that other pluripotency factors are still expressed (e.g., *Esrrb* [Festuccia et al., 2012] and *Nanog* [Mitsui et al., 2003]) and early differentiation markers were not upregulated in specific subpopulations of cells, suggesting that the cells

were not in the process of differentiation or at least only in a very early stage of differentiation decisions (Figures S6A and S6B). The cells were then encapsulated into 3D hydrogels and cultured for 4 days with Dox but without 2i/LIF, which allowed them to initiate differentiation. We then performed both CloneSeq and scRNA-seq on the dissociated clones (Figure 6A).

Overall, we sequenced 1,344 cells treated with Dox for 24 h. For differentiated cells, we collected 1,189 single cells from 3D

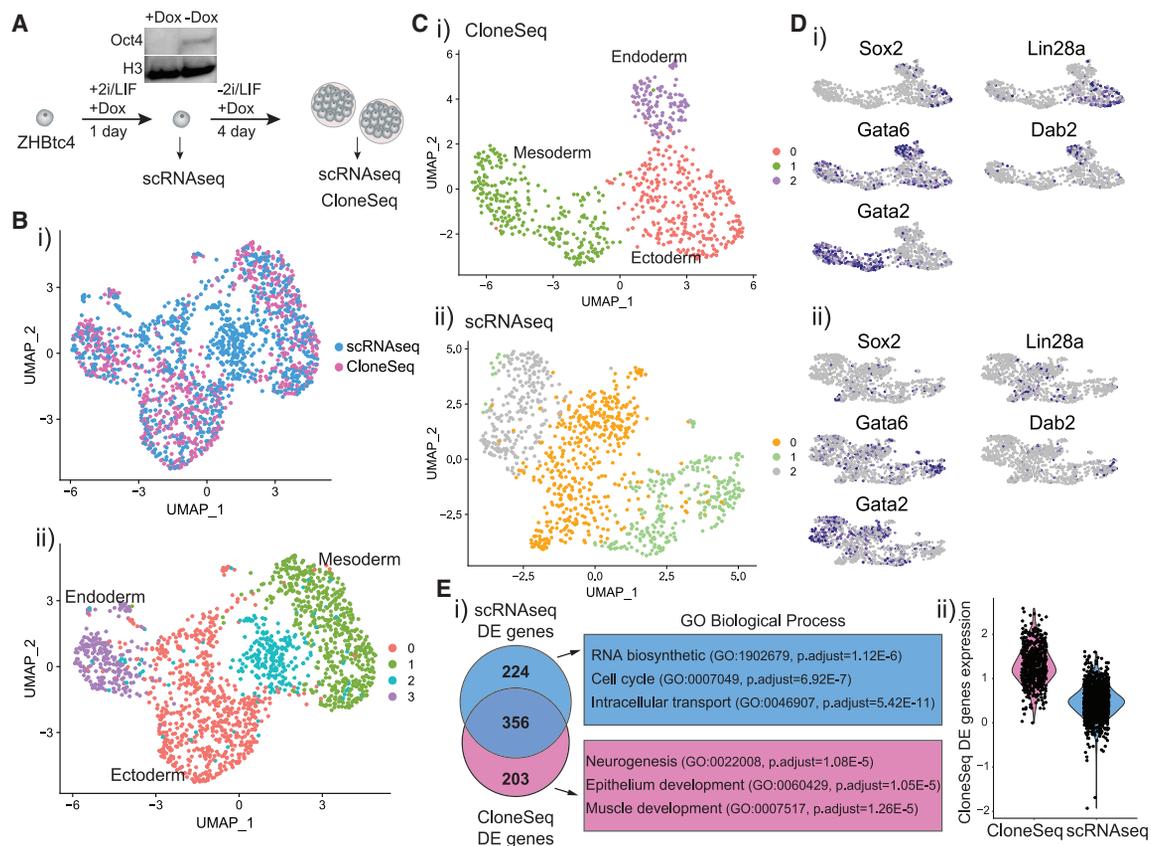


Figure 6. CloneSeq analysis of ESCs upon downregulation of Oct4 shows clonal homogeneity in differentiation potential

(A) (i) ZHBTc4 cells were first treated with Dox for 24 h in plates, the removal of Oct4 protein is shown by a western blot. The cells were then encapsulated into PEG-DEX hydrogels and cultured for 4 days in the presence of Dox but without 2i/LIF. scRNA-seq and CloneSeq were performed accordingly. (B) The transcriptional profile of clones (n = 536) and single cells dissociated from clones (n = 1,189) were obtained by CloneSeq and scRNA-seq. The combined data were mapped into UMAP space based on batch (i) and cluster associations (ii) with four distinct subpopulations in which three are associated with differentiation trajectories and the fourth associated specifically with low-complexed single cells. (C) CloneSeq (i) and scRNA-seq (ii) were analyzed separately and mapped into UMAP space. (D) Feature plots of marker genes for ectoderm (*Sox2* and *Lin28a*), endoderm (*Gata6* and *Dab2*), and mesoderm (*Gata2*) in CloneSeq (i) and scRNA-seq (ii) are marked on UMAP space. (E) (i) Number of differentially expressed genes specific for scRNA-seq and CloneSeq and their enriched functions in GO. (ii) Normalized expression levels of differentially expressed genes of CloneSeq in each dataset. See also Figure S6.

hydrogels (HydrogelSC) and 536 clones (HydrogelCl). The analysis after 4 days of differentiation showed distinct clusters. Although CloneSeq provided superior coverage compared with single-cell assays, clones and single cells clustered together (Figure 6B). This strongly supports that the differentiation decision was made during the first 24 h upon downregulation of Oct4 and that clones merely amplified the differentiation signal of each single cell. Furthermore, clones and single cells were divided into three subpopulations based on marker genes that support the formation of ectoderm, endoderm, and mesoderm cellular states. Single cells and clones were equally distributed in the different subpopulations, with marker genes of endoderm such as *Psap* (Nakazawa et al., 2011), *Crxos* (Saito et al., 2010), *Slc39a4* (*Zip4*) (Dufner-Beattie et al., 2003), *Klf5* (Moore-Scott et al., 2007), *Dab2*, and *Gata6* (Morrisey et al., 1998) and of ectoderm such as *Ssbp3* (Liu et al., 2016), *F11r* (*JAM1*) (Thomas et al., 2004), *Hmga2* (Navarra et al., 2016), *Lin28a* (Parisi et al., 2017), and *Id3* (Kowanetz et al., 2004; Kee and Bronner-Fraser, 2005). The main mesodermal marker genes

that were upregulated were *Gata2* (Orkin, 1992; Johnson et al., 2012), *Sin3b* (David et al., 2008), *Irx3* (Mahlapuu et al., 2001), *Prss8* (Sherwood et al., 2007; Popowski et al., 2017), and *Arid3a* (Tucker, 2017; Figures S6A–S6D). To further validate the specificity of our results, we performed an *in silico* test in which we summed up the expression profiles of ten randomly picked single cells without repeats and produce 118 pseudo-clones. The pseudo-clones show an averaged signal with no subpopulation structures (Figures S6E and S6F).

When single cells and clones were analyzed together, they presented similar patterns in the UMAP space. However, clones provided better sensitivity over single cells. A separate analysis of single cells and clones revealed that clones showed better separation between the clusters, compared with single cells (Figures 6C and 6D). Moreover, differentially expressed genes specifically upregulated in the CloneSeq data are significantly enriched for early differentiation of the three germ layers, while differentially expressed genes upregulated in the single-cell experiment are

associated with metabolic and cell-cycle annotations (Figure 6E). These findings prove the robustness of the CloneSeq technology in identifying cellular states with improved precision.

DISCUSSION

Single-cell technology is key in revealing the importance of heterogeneity. However, it suffers from low sensitivity and therefore it is prone to detect cell-to-cell variation based on highly expressed genes while missing differences in expression of low-abundance genes (Qiu, 2020). To overcome this obstacle, we developed an innovative technology that allows single cells to be grown into small clones in 3D hydrogel spheres and the clone to be sequenced as an entity. Our results suggest that the biochemical composition and the physical structure of the hydrogel spheres are stable, ensuring the compartmentalization of single cells during their development into clones. We have optimized 3D hydrogel sphere culture conditions for PC9 human lung cancer cells, primary glioblastoma cells, and K562 leukemia cells (Figure S1B), as well as mouse ESCs and MEFs that were reprogrammed. CloneSeq showed superior sensitivity over scRNA-seq by retrieving 10-times more transcripts. For PC9 cells, we showed that cells sharing clonal origins are more similar to each other compared with random cells, suggesting that clones are homogeneous.

The heterogeneity of the clonal expansion speed of PC9 cells results in clones of variable sizes. These clones exhibited variable expression profiles, with differentially expressed genes associated with cellular migration, growth inhibition, and metastatic state. Some of the larger clones were also associated with CSC signature. As our single-cell reference profiles were based on cells growing in 3D, we suggest that the stem-cell-like state of PC9 cells cannot be detected with a standard single-cell profiling. Together, our 3D culture system promotes stemness, a feature that we further confirmed using ESCs.

In absence of supporting factors, ESCs maintained pluripotency for 8 days in PEG-DEX hydrogel and continued as pluripotent for 4 days after being reseeded on MEFs. RNA-seq results corroborated our findings, showing significantly higher expression of pluripotency-related transcription factors. Thus, our 3D culturing system promotes stemness for ESCs as well. Furthermore, CloneSeq highlighted the dynamics of ESC differentiation by showing that the differentiation decision is taken very early upon the downregulation of *Oct4*, and better exposed the underlying differentiation signatures.

In summary, CloneSeq is an effective and general method that leverages 3D culture, drop-based microfluidics, and high-throughput sequencing to dramatically extend our ability to characterize cellular states with great statistical power. We will optimize CloneSeq to support culturing of primary tumor-derived cells to allow clonal profiling of cancer cellular states in association with different treatments. In addition, the method could be used to enable full-length RNA-seq to dissect alternative splicing events and somatic mutations in cancer samples. ChIP-seq and bisulfite-seq protocols could also be performed on clones grown in the hydrogel spheres. We expect that CloneSeq will expand our understanding of cancer biology in particular and of other biological systems involving proliferative cells.

Limitations of the study

In the current protocol, we dissolve the hydrogel and profile intact clones. Hence, the CloneSeq is limited to cells that form cell-cell adhesions and remain physically connected outside the sphere. Second, the soft 3D-hydrogel microenvironment is highly supportive of the maintenance of a pluripotent state. Hence, using CloneSeq to study the exit from the pluripotent state would require adjustment and optimization of the differentiation protocols.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Cell lines
- METHOD DETAILS
 - Confocal imaging and analysis
 - Device design and fabrication
 - Barcode and primer design
 - Production of barcoded hydrogel beads
 - Microfluidics operation
 - Clone formation within PEGDT/MALDEX hydrogels
 - Viability evaluation of 3D hydrogel clones
 - Mechanical properties of hydrogel spheres
 - Determination of the size of clones
 - Western blot
 - Clonal barcoding and sequencing
 - Single-cell/clone barcoding using inDrops
 - Single-cell/clone barcoding using Drop-seq
 - Library preparation
 - Species-mixing experiments
 - Sequencing and data filtering
 - Single clone alignment and UMI-based filtering
 - Gene set signature activation analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Differential expression analysis
 - PCA and UMAP analysis
 - Clonal barcoding analysis
 - Clone-to-Clone variation analysis for PC9 cells

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.devcel.2021.04.026>.

ACKNOWLEDGMENTS

O.R. is supported by research grants from the European Research Council (ERC, # 715260 SC-EpiCode), the Israeli Center of Research Excellence (I-CORE) program, the Israel Science Foundation (ISF, #1618/16), and Azrieli Foundation Scholar Program for Distinguished Junior Faculty. O.R. and A.A. are supported by Nofar (65883) of the Israel Innovation authority. E.M. is the Arthur Gutterman Family Chair in stem cell research and is supported by the Israel Science Foundation (ISF1140/17). This project has received funding

from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no 765966 – EpiSyStem.

AUTHOR CONTRIBUTIONS

D.B., X.S., C.K., D.E., E.M., A.B., and O.R. conceived the study, prepared the figures, and wrote the manuscript. D.B., X.S., E.M., A.B., and O.R. designed the experiments. C.K., A.A., A.M., A.B., and D.E. performed PC9, ESCs, and iPSCs tissue culture and library preparation. S.B. and D.B. performed microscopy and pipet aspiration test. D.B. and X.S. prepared the microfluidics system and performed scRNA-seq and CloneSeq experiments. X.S. and C.K. performed clonal barcoding experiments. X.S., D.E., and O.R. performed the computational analysis.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 28, 2020
Revised: March 7, 2021
Accepted: April 23, 2021
Published: May 18, 2021

REFERENCES

- Almalki, S.G., and Agrawal, D.K. (2016). Effects of matrix metalloproteinases on the fate of mesenchymal stem cells. *Stem Cell Res. Ther.* 7, 129. <https://doi.org/10.1186/s13287-016-0393-1>.
- Bainer, R., and Weaver, V. (2013). Cell biology. Strength under tension. *Science* 341, 965–966. <https://doi.org/10.1126/science.1243643>.
- Benchetrit, H., Jaber, M., Zayat, V., Sebban, S., Pushett, A., Makedonski, K., Zakheim, Z., Radwan, A., Maoz, N., Lasry, R., et al. (2019). Direct induction of the three pre-implantation blastocyst cell types from fibroblasts. *Cell Stem Cell* 24, 983–994.e7. <https://doi.org/10.1016/j.stem.2019.03.018>.
- Bhattacharyya, S., and Vratsi, S. (2015). The Malat1 long non-coding RNA is up-regulated by signalling through the PERK axis of unfolded protein response during Flavivirus infection. *Sci. Rep.* 5, 17794. <https://doi.org/10.1038/srep17794>.
- Boumahdi, S., Driessens, G., Lapouge, G., Rorive, S., Nassar, D., Le Mercier, M., Delatte, B., Caauwe, A., Lenglez, S., Nkusi, E., et al. (2014). SOX2 controls tumour initiation and cancer stem-cell functions in squamous-cell carcinoma. *Nature* 511, 246–250. <https://doi.org/10.1038/nature13305>.
- Burdick, J.A., and Murphy, W.L. (2012). Moving from static to dynamic complexity in hydrogel design. *Nat. Commun.* 3, 1269. <https://doi.org/10.1038/ncomms2271>.
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36, 411–420. <https://doi.org/10.1038/nbt.4096>.
- Carey, B.W., Markoulaki, S., Beard, C., Hanna, J., and Jaenisch, R. (2010). Single-gene transgenic mouse strains for reprogramming adult somatic cells. *Nat. Methods* 7, 56–59. <https://doi.org/10.1038/NMETH.1410>.
- Chen, C., Song, G., Xiang, J., Zhang, H., Zhao, S., and Zhan, Y. (2017). AURKA promotes cancer metastasis by regulating epithelial-mesenchymal transition and cancer stem cell properties in hepatocellular carcinoma. *Biochem. Biophys. Res. Commun.* 486, 514–520. <https://doi.org/10.1016/j.bbrc.2017.03.075>.
- Chen, S., Lake, B.B., and Zhang, K. (2019). High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat. Biotechnol.* 37, 1452–1457. <https://doi.org/10.1038/s41587-019-0290-0>.
- Chen, X., and Cubillos-Ruiz, J.R. (2021). Endoplasmic reticulum stress signals in the tumour and its microenvironment. *Nat. Rev. Cancer* 21, 71–88. <https://doi.org/10.1038/s41568-020-00312-2>.
- Cheung, H.-H., Liu, X., and Rennert, O.M. (2012). Apoptosis: reprogramming and the fate of mature cells. *ISRN Cell Biol* 2012, 1–8. <https://doi.org/10.5402/2012/685852>.
- Chung, W., Eum, H.H., Lee, H.-O., Lee, K.-M., Lee, H.-B., Kim, K.T., Ryu, H.S., Kim, S., Lee, J.E., Park, Y.H., et al. (2017). Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat. Commun.* 8, 15081. <https://doi.org/10.1038/ncomms15081>.
- Cidado, J., Wong, H.Y., Rosen, D.M., Cimino-Mathews, A., Garay, J.P., Fessler, A.G., Rasheed, Z.A., Hicks, J., Cochran, R.L., Croessmann, S., et al. (2016). Ki-67 is required for maintenance of cancer stem cells but not cell proliferation. *Oncotarget* 7, 6281–6293. <https://doi.org/10.18632/oncotarget.7057>.
- Dalerba, P., Kalisky, T., Sahoo, D., Rajendran, P.S., Rothenberg, M.E., Leyrat, A.A., Sim, S., Okamoto, J., Johnston, D.M., Qian, D., et al. (2011). Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat. Biotechnol.* 29, 1120–1127. <https://doi.org/10.1038/nbt.2038>.
- Dar, R.D., Razoogy, B.S., Singh, A., Trimeloni, T.V., McCollum, J.M., Cox, C.D., Simpson, M.L., and Weinberger, L.S. (2012). Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc. Natl. Acad. Sci. USA* 109, 17454–17459. <https://doi.org/10.1073/pnas.1213530109>.
- David, G., Grandinetti, K.B., Finnerty, P.M., Simpson, N., Chu, G.C., and Depinho, R.A. (2008). Specific requirement of the chromatin modifier mSin3B in cell cycle exit and cellular differentiation. *Proc. Natl. Acad. Sci. USA* 105, 4168–4172. <https://doi.org/10.1073/pnas.0710285105>.
- Diamandis, E.P., and Christopoulos, T.K. (1991). The biotin-(strept)avidin system: principles and applications in biotechnology. *Clin. Chem.* 37, 625–636.
- Dufner-Beattie, J., Wang, F., Kuo, Y.M., Gitschier, J., Eide, D., and Andrews, G.K. (2003). The acrodermatitis enteropathica gene ZIP4 encodes a tissue-specific, zinc-regulated zinc transporter in mice. *J. Biol. Chem.* 278, 33474–33481. <https://doi.org/10.1074/jbc.M305000200>.
- Edupuganti, R.R., Harikumar, A., Aaronson, Y., Biran, A., Sailaja, B.S., Nissim-Rafinia, M., Azad, G.K., Cohen, M.A., Park, J.E., Shivaila, C.S., et al. (2017). Alternative SET/TAFI promoters regulate embryonic stem cell differentiation. *Stem Cell Rep* 9, 1291–1303. <https://doi.org/10.1016/j.stemcr.2017.08.021>.
- Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotent cells from mouse embryos. *Nature* 292, 154–156. <https://doi.org/10.1038/292154a0>.
- Festuccia, N., Osorno, R., Halbritter, F., Karwacki-Neisius, V., Navarro, P., Colby, D., Wong, F., Yates, A., Tomlinson, S.R., and Chambers, I. (2012). Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. *Cell Stem Cell* 11, 477–490. <https://doi.org/10.1016/j.stem.2012.08.002>.
- Gutschner, T., Hämmerle, M., Eissmann, M., Hsu, J., Kim, Y., Hung, G., Revenko, A., Arun, G., Stentrup, M., Gross, M., et al. (2013). The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res* 73, 1180–1189. <https://doi.org/10.1158/0008-5472.CAN-12-2850>.
- Hochmuth, R.M. (2000). Micropipette aspiration of living cells. *J. Biomech.* 33, 15–22. [https://doi.org/10.1016/S0021-9290\(99\)00175-X](https://doi.org/10.1016/S0021-9290(99)00175-X).
- Hu, H., Song, Z., Yao, Q., Geng, X., Jiang, L., Guo, C., and Li, H. (2018). Proline-rich Protein 11 regulates self-renewal and tumorigenicity of gastric cancer stem cells. *Cell. Physiol. Biochem.* 47, 1721–1728. <https://doi.org/10.1159/000491005>.
- Jaitin, D.A., Weiner, A., Yofe, I., Lara-Astiaso, D., Keren-Shaul, H., David, E., Salame, T.M., Tanay, A., van Oudenaarden, A., and Amit, I. (2016). Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. *Cell* 167, 1883–1896.e15. <https://doi.org/10.1016/j.cell.2016.11.039>.
- Ji, Y., Xie, M., Lan, H., Zhang, Y., Long, Y., Weng, H., Li, D., Cai, W., Zhu, H., Niu, Y., et al. (2013). PRR11 is a novel gene implicated in cell cycle progression and lung cancer. *Int. J. Biochem. Cell Biol.* 45, 645–656. <https://doi.org/10.1016/j.biocel.2012.12.002>.
- Johnson, K.D., Hsu, A.P., Ryu, M.J., Wang, J., Gao, X., Boyer, M.E., Liu, Y., Lee, Y., Calvo, K.R., Keles, S., et al. (2012). Cis-element mutated in GATA2-dependent immunodeficiency governs hematopoiesis and vascular integrity. *J. Clin. Invest.* 122, 3692–3704. <https://doi.org/10.1172/JCI61623>.
- Kee, Y., and Bronner-Fraser, M. (2005). To proliferate or to die: role of Id3 in cell cycle progression and survival of neural crest progenitors. *Genes Dev* 19, 744–755. <https://doi.org/10.1101/gad.1257405>.

- Kim, K.T., Lee, H.W., Lee, H.O., Kim, S.C., Seo, Y.J., Chung, W., Eum, H.H., Nam, D.H., Kim, J., Joo, K.M., et al. (2015). Single-cell mRNA sequencing identifies subclonal heterogeneity in anti-cancer drug responses of lung adenocarcinoma cells. *Genome Biol* 16, 127. <https://doi.org/10.1186/s13059-015-0692-3>.
- Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161, 1187–1201. <https://doi.org/10.1016/j.cell.2015.04.044>.
- Kowanzet, M., Valcourt, U., Bergström, R., Heldin, C.H., and Moustakas, A. (2004). Id2 and Id3 define the potency of cell proliferation and differentiation responses to transforming growth factor β and bone morphogenetic protein. *Mol. Cell. Biol.* 24, 4241–4254. <https://doi.org/10.1128/mcb.24.10.4241-4254.2004>.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25. <https://doi.org/10.1186/gb-2009-10-3-r25>.
- Leung, E.L.H., Fiscus, R.R., Tung, J.W., Tin, V.P., Cheng, L.C., Sihoe, A.D., Fink, L.M., Ma, Y., and Wong, M.P. (2010). Non-small cell lung cancer cells expressing CD44 are enriched for stem cell-like properties. *PLoS One* 5, e14062. <https://doi.org/10.1371/journal.pone.0014062>.
- Li, C.Y., Stevens, K.R., Schwartz, R.E., Alejandro, B.S., Huang, J.H., and Bhatia, S.N. (2014). Micropatterned cell-cell interactions enable functional encapsulation of primary hepatocytes in hydrogel microtissues. *Tissue Eng. Part A* 20, 2200–2212. <https://doi.org/10.1089/ten.tea.2013.0667>.
- Li, H., Courtois, E.T., Sengupta, D., Tan, Y., Chen, K.H., Goh, J.J.L., Kong, S.L., Chua, C., Hon, L.K., Tan, W.S., et al. (2017). Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat. Genet.* 49, 708–718. <https://doi.org/10.1038/ng.3818>.
- Li, H., Yin, C., Zhang, B., Sun, Y., Shi, L., Liu, N., Liang, S., Lu, S., Liu, Y., Zhang, J., et al. (2013). PTTG1 promotes migration and invasion of human non-small cell lung cancer cells and is modulated by miR-186. *Carcinogenesis* 34, 2145–2155. <https://doi.org/10.1093/carcin/bgt158>.
- Litzenburger, U.M., Buenrostro, J.D., Wu, B., Shen, Y., Sheffield, N.C., Kathiria, A., Greenleaf, W.J., and Chang, H.Y. (2017). Single-cell epigenomic variability reveals functional cancer heterogeneity. *Genome Biol* 18, 15. <https://doi.org/10.1186/s13059-016-1133-7>.
- Liu, J., Luo, X., Xu, Y., Gu, J., Tang, F., Jin, Y., and Li, H. (2016). Single-stranded DNA binding protein Ssbp3 induces differentiation of mouse embryonic stem cells into trophoblast-like cells. *Stem Cell Res. Ther.* 7, 79. <https://doi.org/10.1186/s13287-016-0340-1>.
- Liu, X., and Fan, D. (2015). The epithelial-mesenchymal transition and cancer stem cells: functional and mechanistic links. *Curr. Pharm. Des.* 21, 1279–1291. <https://doi.org/10.2174/138161282166614121115611>.
- Losick, R., and Desplan, C. (2008). Stochasticity and cell fate. *Science* 320, 65–68. <https://doi.org/10.1126/science.1147888>.
- Ma, I., and Allan, A.L. (2011). The role of human aldehyde dehydrogenase in normal and cancer stem cells. *Stem Cell Rev. Rep.* 7, 292–306. <https://doi.org/10.1007/s12015-010-9208-4>.
- Ma, K.-Y.Y., Schonnesen, A.A., Brock, A., Van Den Berg, C., Eckhardt, S.G., Liu, Z., and Jiang, N. (2019). Single-cell RNA sequencing of lung adenocarcinoma reveals heterogeneity of immune response-related genes. *JCI Insight* 4. <https://doi.org/10.1172/jci.insight.121387>.
- Macosko, E.Z., Basu, A., Satija, R., Nemes, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A., Kamitaki, N., Martersteck, E., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161, 1202–1214. <https://doi.org/10.1016/j.cell.2015.05.002>.
- Mahlapuu, M., Ormestad, M., Enerbäck, S., and Carlsson, P. (2001). The forkhead transcription factor Foxf1 is required for differentiation of extra-embryonic and lateral plate mesoderm. *Development* 128, 155–166.
- Martello, G., and Smith, A. (2014). The nature of embryonic stem cells. *Annu. Rev. Cell Dev. Biol.* 30, 647–675. <https://doi.org/10.1146/annurev-cellbio-100913-013116>.
- Martin, B., Paesmans, M., Mascaux, C., Berghmans, T., Lothaire, P., Meert, A.P., Lafitte, J.J., and Sculier, J.P. (2004). Ki-67 expression and patients survival in lung cancer: systematic review of the literature with meta-analysis. *Br. J. Cancer* 91, 2018–2025. <https://doi.org/10.1038/sj.bjc.6602233>.
- Menon, D.R., and Fujita, M. (2019). A state of stochastic cancer stemness through the CDK1-SOX2 axis. *Oncotarget* 10, 2583–2585. <https://doi.org/10.18632/oncotarget.26819>.
- Migita, T., Narita, T., Nomura, K., Miyagi, E., Inazuka, F., Matsuura, M., Ushijima, M., Mashima, T., Seimiya, H., Satoh, Y., et al. (2008). ATP citrate lyase: activation and therapeutic implications in non-small cell lung cancer. *Cancer Res* 68, 8547–8554. <https://doi.org/10.1158/0008-5472.CAN-08-1235>.
- Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* 113, 631–642. [https://doi.org/10.1016/S0092-8674\(03\)00393-3](https://doi.org/10.1016/S0092-8674(03)00393-3).
- Moore-Scott, B.A., Opoka, R., Lin, S.C., Kordich, J.J., and Wells, J.M. (2007). Identification of molecular markers that are expressed in discrete anterior-posterior domains of the endoderm from the gastrula stage to mid-gestation. *Dev. Dyn.* 236, 1997–2003. <https://doi.org/10.1002/dvdy.21204>.
- Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., et al. (2003). PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* 34, 267–273. <https://doi.org/10.1038/ng1180>.
- Morrisey, E.E., Tang, Z., Sigrist, K., Lu, M.M., Jiang, F., Ip, H.S., and Parmacek, M.S. (1998). GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes Dev* 12, 3579–3590. <https://doi.org/10.1101/gad.12.22.3579>.
- Nakazawa, F., Alev, C., Jakt, L.M., and Sheng, G. (2011). Yolk sac endoderm is the major source of serum proteins and lipids and is involved in the regulation of vascular integrity in early chick development. *Dev. Dyn.* 240, 2002–2010. <https://doi.org/10.1002/dvdy.22690>.
- Navarra, A., Musto, A., Gargiulo, A., Petrosino, G., Pierantoni, G.M., Fusco, A., Russo, T., and Parisi, S. (2016). Hmga2 is necessary for Otx2-dependent exit of embryonic stem cells from the pluripotent ground state. *BMC Biol* 14, 24. <https://doi.org/10.1186/s12915-016-0246-5>.
- Nie, W., Xu, M.D., Gan, L., Huang, H., Xiu, Q., and Li, B. (2015). Overexpression of stathmin 1 is a poor prognostic biomarker in non-small cell lung cancer. *Lab. Invest.* 95, 56–64. <https://doi.org/10.1038/labinvest.2014.124>.
- Nitzan, M., Karaiskos, N., Friedman, N., and Rajewsky, N. (2019). Gene expression cartography. *Nature* 576, 132–137. <https://doi.org/10.1038/s41586-019-1773-3>.
- Niwa, H., Miyazaki, J.I., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat. Genet.* 24, 372–376. <https://doi.org/10.1038/74199>.
- Noto, A., De Vitis, C., Pisanu, M.E., Roscilli, G., Ricci, G., Catizone, A., Sorrentino, G., Chianese, G., Tagliatalata-Scafati, O., Trisciuglio, D., et al. (2017). Stearoyl-CoA-desaturase 1 regulates lung cancer stemness via stabilization and nuclear localization of YAP/TAZ. *Oncogene* 36, 4573–4584. <https://doi.org/10.1038/ncr.2017.75>.
- Nuttelman, C.R., Moltisen, D.J., Henry, S.M., and Anseth, K.S. (2001). Attachment of fibronectin to poly(vinyl alcohol) hydrogels promotes NIH3T3 cell adhesion, proliferation, and migration. *J. Biomed. Mater. Res.* 57, 217–223. [https://doi.org/10.1002/1097-4636\(200111\)57:2<217::aid-jbm1161>3.0.co;2-i](https://doi.org/10.1002/1097-4636(200111)57:2<217::aid-jbm1161>3.0.co;2-i).
- Obayashi, S., Horiguchi, J., Higuchi, T., Katayama, A., Handa, T., Altan, B., Bai, T., Bao, P., Bao, H., Yokobori, T., et al. (2017). Stathmin1 expression is associated with aggressive phenotypes and cancer stem cell marker expression in breast cancer patients. *Int. J. Oncol.* 51, 781–790. <https://doi.org/10.3892/ijo.2017.4085>.
- Orkin, S.H. (1992). GATA-binding transcription factors in hematopoietic cells. *Blood* 80, 575–581.
- Pan, X.N., Chen, J.J., Wang, L.X., Xiao, R.Z., Liu, L.L., Fang, Z.G., Liu, Q., Long, Z.J., and Lin, D.J. (2014). Inhibition of c-Myc overcomes cytotoxic

- drug resistance in acute myeloid leukemia cells by promoting differentiation. *PLoS One* 9, e105381. <https://doi.org/10.1371/journal.pone.0105381>.
- Parisi, S., Passaro, F., Russo, L., Musto, A., Navarra, A., Romano, S., Petrosino, G., and Russo, T. (2017). Lin28 is induced in primed embryonic stem cells and regulates let-7-independent events. *FASEB J* 31, 1046–1058. <https://doi.org/10.1096/fj.201600848R>.
- Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344, 1396–1401. <https://doi.org/10.1126/science.1254257>.
- Perumal, D., Singh, S., Yoder, S.J., Bloom, G.C., and Chellappan, S.P. (2012). A novel five gene signature derived from stem-like side population cells predicts overall and recurrence-free survival in NSCLC. *PLoS One* 7, e43589. <https://doi.org/10.1371/journal.pone.0043589>.
- Qiu, P. (2020). Embracing the dropouts in single-cell RNA-seq analysis. *Nat. Commun.* 11, 1169. <https://doi.org/10.1038/s41467-020-14976-9>.
- Raj, A., and van Oudenaarden, A. (2009). Single-molecule approaches to stochastic gene expression. *Annu. Rev. Biophys.* 38, 255–270. <https://doi.org/10.1146/annurev.biophys.37.032807.125928>.
- Rotem, A., Ram, O., Shores, N., Sperling, R.A., Schnall-Levin, M., Zhang, H., Basu, A., Bernstein, B.E., and Weitz, D.A. (2015). High-throughput single-cell labeling (Hi-SCL) for RNA-seq using drop-based microfluidics. *PLoS One* 10, e0116328. <https://doi.org/10.1371/journal.pone.0116328>.
- Saito, K., Abe, H., Nakazawa, M., Irokawa, E., Watanabe, M., Hosoi, Y., Soma, M., Kasuga, K., Kojima, I., and Kobayashi, M. (2010). Cloning of complementary DNAs encoding structurally related homeoproteins from preimplantation mouse embryos: their involvement in the differentiation of embryonic stem cells. *Biol. Reprod.* 82, 687–697. <https://doi.org/10.1095/biolreprod.108.075697>.
- Schlesinger, S., and Meshorer, E. (2019). Open chromatin, epigenetic plasticity, and nuclear organization in pluripotency. *Dev. Cell* 48, 135–150. <https://doi.org/10.1016/j.devcel.2019.01.003>.
- Sherwood, R.I., Jitianu, C., Cleaver, O., Shaywitz, D.A., Lamenza, J.O., Chen, A.E., Golub, T.R., and Melton, D.A. (2007). Prospective isolation and global gene expression analysis of definitive and visceral endoderm. *Dev. Biol.* 304, 541–555. <https://doi.org/10.1016/j.ydbio.2007.01.011>.
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. *Cell* 177, 1888–1902.e21. <https://doi.org/10.1016/j.cell.2019.05.031>.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* 102, 15545–15550. <https://doi.org/10.1073/pnas.0506580102>.
- Theret, D.P., Levesque, M.J., Sato, M., Nerem, R.M., and Wheeler, L.T. (1988). The application of a homogeneous half-space model in the analysis of endothelial cell micropipette measurements. *J. Biomech. Eng.* 110, 190–199. <https://doi.org/10.1115/1.3108430>.
- Thomas, F.C., Sheth, B., Eckert, J.J., Bazzoni, G., Dejana, E., and Fleming, T.P. (2004). Contribution of JAM-1 to epithelial differentiation and tight-junction biogenesis in the mouse preimplantation embryo. *J. Cell Sci.* 117, 5599–5608. <https://doi.org/10.1242/jcs.01424>.
- Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 352, 189–196. <https://doi.org/10.1126/science.aad0501>.
- Popowski, M., Lee, B.K., Rhee, C., Iyer, V.R., and Tucker, H.O. (2017). Arid3a regulates mesoderm differentiation in mouse embryonic stem cells. *J. Stem Cell Ther. Transplant.* 1, 52–62. <https://doi.org/10.29328/journal.jsctt.1001005>.
- Tung, P.Y., Blischak, J.D., Hsiao, C.J., Knowles, D.A., Burnett, J.E., Pritchard, J.K., and Gilad, Y. (2017). Batch effects and the effective design of single-cell gene expression studies. *Sci. Rep.* 7, 39921. <https://doi.org/10.1038/srep39921>.
- Wagner, D.E., Weinreb, C., Collins, Z.M., Briggs, J.A., Megason, S.G., and Klein, A.M. (2018). Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* 360, 981–987. <https://doi.org/10.1126/science.aar4362>.
- Wang, H., and Unternaehrer, J.J. (2019). Epithelial-mesenchymal transition and cancer stem cells: at the crossroads of differentiation and dedifferentiation. *Dev. Dyn.* 248, 10–20. <https://doi.org/10.1002/dvdy.24678>.
- Wu, J., Yang, L., Shan, Y., Cai, C., Wang, S., and Zhang, H. (2016). AURKA promotes cell migration and invasion of head and neck squamous cell carcinoma through regulation of the AURKA/Akt/FAK signaling pathway. *Oncol. Lett.* 11, 1889–1894. <https://doi.org/10.3892/ol.2016.4110>.
- Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J., et al. (2013). Single-cell RNA-seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* 20, 1131–1139. <https://doi.org/10.1038/nsmb.2660>.
- Yang, L., Zhou, Q., Chen, X., Su, L., Liu, B., and Zhang, H. (2016). Activation of the FAK/PI3K pathway is crucial for AURKA-induced epithelial-mesenchymal transition in laryngeal cancer. *Oncol. Rep.* 36, 819–826. <https://doi.org/10.3892/or.2016.4872>.
- Yang, Y.F., Chang, Y.C., Jan, Y.H., Yang, C.J., Huang, M.S., and Hsiao, M. (2020). Squalene synthase promotes the invasion of lung cancer cells via the osteopontin/ERK pathway. *Oncogenesis* 9, 78. <https://doi.org/10.1038/s41389-020-00262-2>.
- Yasumoto, Y., Miyazaki, H., Vaidyan, L.K., Kagawa, Y., Ebrahimi, M., Yamamoto, Y., Ogata, M., Katsuyama, Y., Sadahiro, H., Suzuki, M., and Owada, Y. (2016). Inhibition of fatty acid synthase decreases expression of stemness markers in glioma stem cells. *PLoS ONE* 11, e0147717. <https://doi.org/10.1371/journal.pone.0147717>.
- Yi, M., Li, J., Chen, S., Cai, J., Ban, Y., Peng, Q., Zhou, Y., Zeng, Z., Peng, S., Li, X., et al. (2018). Emerging role of lipid metabolism alterations in cancer stem cells. *J. Exp. Clin. Cancer Res.* 37, 118. <https://doi.org/10.1186/s13046-018-0784-5>.
- Ying, Q.L., and Smith, A. (2017). The art of capturing pluripotency: creating the right culture. *Stem Cell Rep* 8, 1457–1464. <https://doi.org/10.1016/j.stemcr.2017.05.020>.
- Ying, Q.L., Wray, J., Nichols, J., Battle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature* 453, 519–523. <https://doi.org/10.1038/nature06968>.
- Yu, Y.H., Chiou, G.-Y., Huang, P.-I., Lo, W.-L., Wang, C.-Y., Lu, K.-H., Yu, C.-C., Alterovitz, G., Huang, W.-C., Lo, J.-F., et al. (2012). Network biology of tumor stem-like cells identified a regulatory role of cbx5 in lung cancer. *Sci. Rep.* 2, 584. <https://doi.org/10.1038/srep00584>.
- Zhang, R., Gao, X., Zuo, J., Hu, B., Yang, J., Zhao, J., and Chen, J. (2020). STMN1 upregulation mediates hepatocellular carcinoma and hepatic stellate cell crosstalk to aggravate cancer by triggering the MET pathway. *Cancer Sci* 111, 406–417. <https://doi.org/10.1111/cas.14262>.
- Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., and Enard, W. (2017). Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* 65, 631–643.e4. <https://doi.org/10.1016/j.molcel.2017.01.023>.
- Zilionis, R., Nainys, J., Veres, A., Savova, V., Zemmour, D., Klein, A.M., and Mazutis, L. (2017). Single-cell barcoding and sequencing using droplet microfluidics. *Nat. Protoc.* 12, 44–73. <https://doi.org/10.1038/nprot.2016.154>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	Identifier
Antibodies		
Goat polyclonal to Oct4	Abcam	Cat# ab27985; RRID: AB_776898
Rabbit polyclonal to Histone H3	Abcam	Cat# ab1791; RRID: AB_302613
Chemicals, peptides, and recombinant proteins		
LIF	Merck	Cat# ESG1107
PD0325901	PeproTech	Cat #3911091
CHIR99021	PeproTech	Cat# 2520691
Doxycycline hydrochloride	Sigma-Aldrich	Cat #D3447
all-trans-retinoic acid	Sigma-Aldrich	Cat #R2625
Critical commercial assays		
The 3-D Life Product	Cellendes GmbH	N/A
Experimental models: cell lines		
Human: PC-9 cells	Prof. Ravid Straussman (Weizmann Institute, Israel)	RCB Cat# RCB4455, RRID: CVCL_B260
Mouse: ES-R1 cells	Prof. Eran Meshorer (The Hebrew University of Jerusalem, Israel)	ATCC Cat# SCRC-1011, RRID: CVCL_2167
Mouse: ZHBTc4 cells	Prof. Eran Meshorer (The Hebrew University of Jerusalem, Israel)	RCB Cat# AES0136, RRID: CVCL_C715
Mouse: BYKE ESCs	Dr. Yosef Buganim (The Hebrew University of Jerusalem, Israel) Benchetrit et al., 2019	N/A
Recombinant DNA		
Plasmid: CRISPseq BFP w gRNA UGI	Jaitin et al., 2016	N/A
Software and algorithms		
Bowtie2	Langmead et al., 2009	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
MSigDB GSEA	Mootha et al., 2003 ; Subramanian et al., 2005	http://www.gsea-msigdb.org/gsea/msigdb/index.jsp
Seurat v2.4	Stuart et al., 2019	https://satijalab.org/seurat/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Oren Ram (oren.ram@mail.huji.ac.il).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The accession number for the data reported in this paper is GEO: GSE155888.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines

Human PC-9 lung adenocarcinoma cells that express GFP were kindly provided by Prof. Ravid Straussman (Weizmann Institute, Israel). ES-R1 and ZHBTc4 cells were kindly provided by Prof. Eran Meshorer, BYKE ESCs were kindly provided by Dr. Yosef

Buganin (The Hebrew University of Jerusalem, Israel). PC9 cells were grown in DMEM (Sigma-Aldrich, cat #D5671) supplemented with 10% fetal bovine serum (FBS; Biological Industries Israel, cat #04-007-1A), 50 $\mu\text{g/ml}$ penicillin-streptomycin (Biological Industries Israel, cat #03-031-1B), 2 mM L-glutamine (Biological Industries Israel, cat #03-020-1B), and 1 mM sodium pyruvate (Biological Industries Israel, cat #03-042-1B). R1, BYKE and ZHBTc4 ESCs cells were grown on 0.1% gelatin-coated standard tissue culture dishes and maintained in ESC medium (DMEM, 15% ESC-grade FBS, 50 $\mu\text{g/ml}$ penicillin-streptomycin, 2 mM L-glutamine, 1 mM sodium pyruvate, 0.1 mM non-essential amino acids (Biological Industries Israel, cat #01-340-1B), 0.1 mM β -mercaptoethanol (Sigma-Aldrich, cat #M3148)). To maintain pluripotency, 1000 U/ml Leukemia Inhibitory Factor (LIF; Merck, cat# ESG1107) and 2i (1 μM PD0325901, PeproTech cat #3911091; 3 μM CHIR99021, PeproTech cat# 2520691) were added to the culture medium. For differentiation experiments, cells or hydrogel spheres were washed with PBS, and resuspended in basal ESC medium 1) without 2i, 2) without 2i and LIF, or and 3) without 2i and LIF supplemented with 0.25 M all-trans-retinoic acid (RA, Sigma-Aldrich, Cat #R2625), and cultured for 4 days. Specifically, in the differentiation experiment of ZHBTc4 cells, cells were first cultured with 2i and LIF and with 1 $\mu\text{g/ml}$ Doxycycline hydrochloride (Dox, Sigma-Aldrich, cat #D3447) for 24 h, then without 2i and LIF but still with doxycycline for 4 days, either in PEG-DEX hydrogel, or as spheroid in suspension on Petri Dish.

METHOD DETAILS

Confocal imaging and analysis

The spheres were imaged using the revolution spinning disk (CSUX; Yokogawa) confocal microscope, equipped with the iXon DU-897-BV monochrome EMCCD camera (Andor, UK). The images were acquired in z-stack and the same parameters were kept along the different experiments for each marker. Image analysis was performed using a custom MATLAB script (MathWorks). Images were processed by manual segmentation of cells outline using phase-contrast images. Fluorescence intensities were quantified from a single focal plane by subtracting each fluorescent channel's background levels and calculating the average intensity in a specified area.

Device design and fabrication

Four microfluidics devices were used: (I) to produce acrylamide hydrogel microparticles with acrydite-modified DNA primers for barcoding; (II) to encapsulate single cells within PEGDT/MALDEX hydrogels; (III) to encapsulate single cells with barcodes, lysis buffer, and reverse transcriptase (RT) enzyme within droplets; and (IV) to encapsulate clones with barcodes, lysis buffer, and RT enzyme within droplets (Figures S3A–S3D). Devices were designed using the AutoCAD software (Autodesk). All chips were fabricated by photolithographically defining SU8 (SU-8 2050, MicroChem) on silicon wafers at the Harvey Krueger Center of Nanotechnology at the Hebrew University of Jerusalem. The depths of the photoresist layers were $50.4 \pm 1 \mu\text{m}$ for device I, $81.2 \pm 1 \mu\text{m}$ for device II, $79.1 \pm 1 \mu\text{m}$ for device III, and $123.6 \mu\text{m} \pm 1 \mu\text{m}$ for device IV. Designs for devices I and III were adapted from inDrops (Zilionis et al., 2017). Device II was modified from the co-flow drop maker of Hi-SCL (Rotem et al., 2015), whereas device IV is a modification of the standard inDrops chip. The designs used to fabricate the devices are available in CAD format (Data S1). Polydimethylsiloxane (PDMS) at a ratio of 10:1 of base and crosslinker, respectively, was formed by curing the prepolymer (Sylgard 184, Dow-Corning) on the silicon templet at 65°C for 2 h. PDMS devices were covalently bound to $\text{N}^\circ 1$ glass coverslips using Femto oxygen plasma activation (Diener) for 15 s at 50 W. The PDMS devices were treated with Aquapel (Rider) water repellent and dried under air in order to make the devices more hydrophobic and prevent wetting of drops on the channel walls. Drop volume (v) calculation for CloneSeq devices was based on still images of droplets at the outlet of the microfluidic device using the equation: $v = \frac{\pi D^3 h}{6}$, where h is the height of the channel and D is the droplet diameter in μm (Zilionis et al., 2017).

Barcode and primer design

The acrydite-modified DNA primers used for the hydrogel barcode beads are based on a previously published protocol (Zilionis et al., 2017); primers were supplied by IDT. We modified the barcode plates (eight 96-well plates) to expand barcode complexity by changing the length of barcode 1 to a variable length of 7 to 10 bases. DNA oligonucleotide sequences are listed in Table S5. We typically used a 10-nmol scale normalization and standard desalting, and ordered oligonucleotides dissolved to a final concentration of 50 μM in 10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA (TE buffer).

Production of barcoded hydrogel beads

Hydrogel beads carrying barcoded DNA primers were produced using a method described previously (Zilionis et al., 2017). The hydrogel beads were composed of a 4xAB solution (2.6 ml 40% acrylamide in water (Sigma-Aldrich, cat #01697), 3.6 ml 40% total 19:1 acrylamide:bis-acrylamide aqueous solution (Bio-lab, cat #1352335), 3.8 ml water) supplemented with an acrydite-modified DNA primer (5'-ACryd/iSpPC/CGATGACGTAATACGACTCACTATAGGGATACCATGGCTCTTCCCTACACGACGCTCTTC-3', where ACryd represents acrydite and iSpPC represents the photo-cleavable Int PC). DNA primers on polymerized hydrogel beads were barcoded using a combination of the split-and-pool method and a primer extension reaction using an automated liquid handling system (Biomek 4000, Beckman Coulter Life Science). The final barcode library complexity was around 147,456 unique barcodes repeated across around 40 million hydrogel beads per synthesis batch with an average of 109 copies of fully extended DNA primers per single bead (Zilionis et al., 2017). Hydrogel beads were produced using the flow-focusing microfluidic device I (Figure S3A) as previously described (Wagner et al., 2018). Flow rates used for the hydrogel bead synthesis were 1000 $\mu\text{l/h}$ for a 4xAB solution

supplemented with an acrydite-modified DNA primer and 1600 $\mu\text{l/h}$ for the oil phase (Table S1). After barcode synthesis, the bar-coded beads were filtered twice with a cell strainer of 70 μm (pluriSelect, Cat #43-10070-50) to obtain homogeneously sized beads with a diameter of around 60 μm , as shown in Figure S3E.

Microfluidics operation

Droplet formation and cell encapsulation were performed using EZ pressure-derived pumps (Fluigent), controlled by the A-I-O software at pressures ranged from 69 mbar to 2 bar. The continuous oil phase for all droplet microfluidics experiments was Novec HFE-7500 fluorinated oil (3M) containing 2% w/w 008-FluoroSurfactant (RAN Biotechnologies). For all experiments, cells were kept in a tube surrounded by ice and were gently agitated with a micro-stir bar placed inside the tube and rotated using a magnet attached to a rotating motor to prevent sedimentation and clumping. The flow was visualized under an optical microscope (NIKON Ti-U) at 10x magnification and imaged at 1000-2000 frames per second using a Hispec1 camera (FASTEC Imaging). Table S1 summarizes the flow rates and the pressure pumps used for operating the different devices used.

Clone formation within PEGDT/MALDEX hydrogels

All materials used to synthesize and dissolve different hydrogel spheres used to grow PC9 and ESCs cells were from Cellendes GmbH, Germany. We used PEGDT (Mn \sim 10,000; cat #L50-1) and MALDEX (cat #M92-3) hydrogel chemistry to encapsulate cells within spheres and grew them into clones. PC9 cells were encapsulated in PEGDT/MALDEX hydrogel with no additional cell adhesion peptides or remodeling supplements. ESCs cells were encapsulated in MALDEX and cell adhesion peptides containing cell recognition motifs of the extracellular matrix (RGD (cat #P10-3); peptide sequence: Acetyl-Cys-Doa*-Doa-Gly-Arg-Gly-Asp-Ser-Pro-NH₂ [*Doa:8-amino-3,6-dioxaoctanoic acid]) and MMP-cleavable peptide modified PEGDT (cat #L60-1; MMP sequence: Pro-Leu-Gly-Leu-Trp-Ala). Hydrogel spheres were dissolved using a 1:20 dilution of dextranase from *Chaetomium gracile* (cat #D10-1) in PBS incubated for 30 min at 37 °C. The gelation buffer (GB; cat #B20-3) used in all cell encapsulations contained 10 g/l glucose, 0.5 M HEPES (pH 7.2), 0.05 M KCl, 1.1 M NaCl, 0.2 M NaH₂PO₄, and 0.2 g/l phenol red. Before use, MALDEX, PEGDT, MPEGDT, and RGD peptide were briefly spun down to make sure that the lipolysis material was at the bottom of the reaction tube. MALDEX was resuspended in 170 μl of double distilled water to a concentration of 30 mM maleimide groups. PEGDT and MPEGDT were resuspended in 188 μl of double distilled water to a concentration of 20 mM thiol groups. RGD peptide was resuspended in 48 μl of double distilled water to a concentration of 20 mM of peptide and thiol groups.

Device II (Figure S3B) was used to encapsulate the single cells within the hydrogel spheres. For PC9 cells, inlet 1 consisted of 310 μl 0.1% w/v gelatin in double distilled water, 50 μl of GB, 67.5 μl of 20 mM PEGDT, and 120 μl PBS. For ES cells, inlet 1 consisted of 300 μl 0.1% w/v gelatin in water, 10 μl of 20 mM RGD peptide, 50 μl of GB, 67.5 μl of 20 mM MPEGDT, and 120 μl PBS. For both cell types, inlet 2 consisted of 335 μl 0.1% w/v gelatin in water, 50 μl of GB, 45 μl of 30 mM MALDEX, and 120 μl of cell suspension containing around 1 million cells in 99 μl PBS, 10 μl Extracellular matrix (Sigma-Aldrich, cat #E1270), and 11 μl OptiPrep Density Gradient Medium (Sigma-Aldrich, cat #D1556) to minimize cell clumping. The resulting hydrogel-cell mix was subsequently enveloped in the device in HFE 7500 oil with 2% surfactant (inlet 3) to produce single-cell hydrogel spheres of 50-55 μm in diameter. Flow rates were 500 $\mu\text{l/h}$ for inlets 1 and 2 and 2200 $\mu\text{l/h}$ for inlet 3 (Table S1). The encapsulation efficiency is about 1 cell per 5 droplets. The device is designed to generate \sim 700 hydrogel spheres per second with a diameter of 60 ± 3 μm . (Video S1) The resulting single cell-containing hydrogel spheres were allowed to cure for 5 min at 37 °C, and the upper hydrogel fraction (\sim 500 μl) was demulsified by incubating 1 min at 37 °C in the demulsifying solution containing 400 μl of neat HFE 7500 oil, 100 μl perfluoro-1-octanol (PFO, Sigma-Aldrich, cat #370533), 280 μl PBS, and 20 μl of 1 g/ml methoxy PEG thiol (average Mn \sim 800; Sigma-Aldrich, cat #729108). Methoxy PEG thiol was used to mask unbound maleimide groups and to prevent aggregation of spheres while demulsifying. The hydrogel spheres fraction was washed three times with 1 ml PBS with centrifugation at 250 rcf for 2 min. For culturing encapsulated cells into clones, 500- μl aliquots of hydrogel spheres were grown in a standard 48-well cell culture dish well with 500 μl standard cell growth medium suitable for each cell type and growth condition. A confocal microscope video showing the clonal expansion of ESCs for 3 days can be found in Video S2.

Viability evaluation of 3D hydrogel clones

Cell viability within PEGDT/MALDEX hydrogels was evaluated by counting cells stained with trypan blue (Biological Industries, Israel, cat #03-102-1B). Cells were encapsulated in PEGDT/MALDEX based hydrogels as single cells and then left to grow into clones for 4 days. The media was replaced daily for ESCs and every 3 days for PC9 cells. After 4 days, hydrogel spheres were washed three times with 1 ml PBS with centrifugation at 250 rcf for 2 min. Beads were then suspended in 300 μl of 1:20 dilution of dextranase from *Chaetomium gracile* in PBS and incubated for 30 min at 37 °C. Following hydrogel spheres degradation, cells were centrifuged at 500 rcf and treated with 100 μl 1x trypsin/EDTA solution for 5 min at 37 °C, to break aggregates. The trypsin was then quenched by adding an equal volume of medium. Trypan blue was added at a 1:1 ratio with the cell medium. Live/dead cell viability was assessed using Countess II FL Automated Cell Counters (Thermo Fisher Scientific). For each condition, cell counts were obtained for three different measurements averaged over two replicates.

Mechanical properties of hydrogel spheres

To evaluate the microenvironment in hydrogel spheres, we used GFP-based R1 ESCs and PC9 cells encapsulated in rhodamine-modified PEGDT/MALDEX hydrogel spheres. To modify the spheres, we added 12.5 μl of 1 mg/ml of Biotin PEG thiol, MW 400

(NANOCS, cat #PG2-BNTH-400) to inlet 1, which contained the PEGDT when encapsulating single cells within the hydrogel spheres. Following sphere formation, the biotin-modified hydrogel spheres were washed three times with PBS, modified with a 1:1 volume ratio of packed hydrogel spheres and 1 mg/ml streptavidin-rhodamine (Jackson ImmunoResearch, Cat # 016-290-084) for 5 min, and washed three times with PBS.

To analyze sphere homogeneity, empty biotin modified hydrogel spheres were modified with 1:1 volume ratio of 1 mg/ml streptavidin (Sigma-Aldrich, cat #85878) for 5 min, washed three times with PBS, and then modified with 1:1 volume ratio with 1 mg/ml biotin-5-fluorescein conjugate (Sigma-Aldrich, cat #53608) followed by three washes with PBS. Cross-sectional images were taken using NIKON A1 confocal microscope to evaluate the distribution of molecules within the hydrogel spheres.

Hydrogel spheres mechanics were measured using a micropipette aspiration system made in-house using a micromanipulator holder (Narishige, NT-88-V3) connected to a manual hydraulic, oil-filled microinjector (Eppendorf, CellTram) and a pressure-sensing diaphragm (Validyne, DP15, CD379). Micropipettes were fabricated by pulling borosilicate capillaries (Sutter Instruments, MicroPulser P-1000) and then forging them to 3- μ m inner diameter tips (Narishige, MF-830 Micro-Forging). Suspended beads were then placed on a glass microscope slide and mounted onto an inverted fluorescent microscope (Nikon Eclipse, Ti-E). The pipette tip was aligned with the hydrogel spheres, and basal negative pressure was applied to capture it stably. Aspiration dynamics in response to applied pressure (relative to the basal levels) inside the pipette were recorded (Andor Zyla 4.2 sCMOS) by imaging the Cy3 fluorescent channel using a CFI Super Plan Fluor ELWD 40XC (Nikon). The mechanical properties of the hydrogel spheres were evaluated based on the relationship between the aspirated based length $L(t)$ and the applied pressure ΔP using the half-space model (Theret et al., 1988):

$$E = \varphi \frac{3R}{2\pi} \left(\frac{\Delta P}{L} \right)$$

where R is the inner pipette radius and WD 40XC (Nikon). The mechanic

Determination of the size of clones

In order to measure the size of clones developed in PEGDT/MALDEX hydrogel spheres, we took confocal fluorescence microscopy images of clones in hydrogel spheres and counted the number of cells per clone. ESCs and PC9 cells were encapsulated in PEG-DEX hydrogel and grown for 3 to 4 days for ESCs with 2i+LIF and 6 to 7 days for PC9 cells. Spheres containing clones were selected randomly from the tissue culture plate then fixed with 10% v/v formaldehyde (Sigma-Aldrich, cat #F8775). The cells were stained with DAPI Fluoromount-G® (ENCO, cat #0100-20). Cross-sectional images were taken using a NIKON A1 confocal microscope. The numbers of cells per clone were determined by counting the number of nuclei. The surface areas of clones were determined using ImageJ image processing software (NIH, <https://imagej.nih.gov/ij/>).

Western blot

Proteins were separated by SDS-PAGE on 4%–20% polyacrylamide gradient gels and transferred to 0.45- μ m membranes (iBlot2, PVDF, mini Transfer Stacks, Thermo Scientific, IB24002v). The membranes were incubated with the appropriate primary and secondary antibodies and washed with PBS-Tween 20. Horseradish-peroxidase-conjugated secondary antibodies were detected by SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific, PI-34080). Antibodies used were goat polyclonal anti-Oct4 (Abcam, ab27985) and rabbit polyclonal to Histone H3 (Abcam, ab1791).

Clonal barcoding and sequencing

To assess the impact of clonal cell origin on cellular states and variation, we produced PC9 cell lines that carry genomic barcodes. The plasmid CRISPseq BFP w gRNA UGI was kindly given by Prof. Ido Amit (Jaitin et al., 2016). The plasmid contains a UGI region of 8 bp as a genomic barcode about 600 bp upstream of the BFP polyA region; BFP expression allowed us to detect plasmid integration. The design of plasmid can be found in Data S2. To validate the number of unique barcodes and their even distribution, we produced a sequencing library from about 50,000 cells by amplifying the plasmid by 12 PCR cycles (2 min at 98 °C, 2 x (98 °C 20 s, 55 °C 30 s, 72 °C 40 s), 10 x (98 °C 20 s, 65 °C 30 s, 72 °C 40 s)). For transfection, we grew 293T cells to 80% confluency and then incubated the cells for 30 min in conditioning medium (50 ml 293T medium supplemented with 500 μ l L-glutamine and 500 μ l Sodium-Pyruvate). The transfection solution contained 34.5 μ l TransIT®-LT1 Transfection Reagent (Mirus, cat #MIR-2300), 1 μ g transfer plasmid (psPAX.2), 7 μ g VSV-G (PMD2.G), and 3.5 μ g of our BFP-UGI plasmid diluted to 1.5 ml with Opti-MEM I Reduced Serum Medium (Gibco, cat #31985088). The transfection solution was incubated for 30 min at room temperature and then added dropwise onto the cells. Cells were incubated with gentle shaking, and media containing viruses were collected after 48 h and 72 h. For virus concentration, the PEG Virus Precipitation Kit (BioVision, cat #K904) was used. The collected media was PEG precipitated overnight at 4 °C, filtered through a 0.45- μ m pore size filter (MF-Millipore), and centrifuged at 2500 rcf for 30 min. The resulting virus pellet was resuspended in 20 μ l Virus Resuspension Solution supplied with the kit.

For infection, PC9 cells at 50% confluency were incubated for 30 min in 8 μ g/ml polybrene (Sigma-Aldrich, cat #107689) solution in culture medium. The cells were then loaded dropwise with the virus solution and incubated for 2.5 h in a humidified incubator at 37 °C, 5% CO₂. PC9 cells were infected with the virus at an MOI of 0.05. Finally, BFP-positive cells were sorted by FACS to obtain the BFP PC9 cell line containing genetic barcodes. We encapsulated BFP PC9 cells in PEGDT/MALDEX hydrogel and grew them for 7 days. Hydrogel spheres were washed and degraded as described above, and the released cells were trypsinized and resuspended to obtain a single-cell solution, which was subjected to scRNA-seq. Three biological replicates were performed for this experiment.

Single-cell/clone barcoding using inDrops

Single cells and single clone transcriptomes were barcoded using inDrops as previously reported (Klein et al., 2015). **Device III** was used for scRNA-seq, and **device IV** was used for CloneSeq (Figures S3C and S3D). The devices have four inlets: 1) **Cell/encapsulated clone inlet**: For *scRNA-seq experiments*, cells were loaded at 200,000 cells/ml in PBS containing 10% v/v OptiPrep and maintained in suspension using a magnetic micro-stirrer bar placed within the tube. For *CloneSeq experiments*, the encapsulated clones were first released from the PEG-DEX hydrogel using Dextranase. Dextranase was added to hydrogels in 1:20 (v/v) and incubated for 30 min at 37 °C. Then the clone suspension was passed through a 70 μm strainer and washed twice with PBS. Finally, around 20,000 clones were resuspended in 900 μl PBS and 100 μl OptiPrep and maintained in suspension using a magnetic micro-stirrer bar placed within the tube. 2) **Barcoding acrylamide beads inlet**: Barcoding beads were prepared as previously described (Zilionis et al., 2017) and kept in dark at 4 °C in 50% (v/v) 10 mM Tris-HCl (pH 8), 0.1 M EDTA, 0.1% (v/v) Tween-20. Around 100-150 μl barcoded acrylamide beads were centrifuged in a 1.5-ml Eppendorf tube at 1500 rcf for 2 min to obtain packed beads. After aspirating residual buffer from the pelleted beads, the tube was loaded onto the corresponding inlet in the microfluidics setup. 3) **Reverse transcription/lysis mix inlet**: For *both single-cell and CloneSeq experiments*, the RT/lysis mix consisted of 180 μL 5X First-Strand buffer (SuperScript™ III Reverse Transcriptase Kit, Invitrogen Cat #18080044), 27 μL 10% (v/v) IGEPAL CA-630 (Sigma-Aldrich, Cat #18896), 20 μL 25 mM dNTPs (NEB, Cat #N0446S), 30 μL 0.1 M DTT (SuperScript™ III Reverse Transcriptase Kit, Invitrogen Cat #18080044), 45 μL 1 M Tris-HCl (pH 8.0) (Sigma-Aldrich, Cat #T2319), 30 μL murine RNase inhibitor (NEB, Cat #M0314), 45 μL SuperScript™ III RT enzyme (200 U/μL, Invitrogen Cat #18080044), and 73 μL nuclease-free water (Sigma-Aldrich, Cat #W4502). 4) **Carrier oil inlet**: The carrier oil was 3 ml of HFE-7500 with 2% (w/w) fluorosurfactant. During microfluidics runs, cell suspension/encapsulated cells and collection tubes were kept on ice. The device generates monodispersed droplets with volumes in the range of 2 nl for scRNA-seq and around 4-5 nl for CloneSeq (Video S3). The flow rates used for sequencing are shown in Table S1.

Single-cell/clone barcoding using Drop-seq

Single cells and single clone transcriptomes were barcoded using Drop-seq as previously reported (Macosko et al., 2015). We used the same devices used for inDrops (**Device III** for scRNA-seq; **device IV** for CloneSeq) while plugging port No. 7 (Figures S3C and S3D). The devices have three inlets: 1) **Cell/clone inlet**: Cells were loaded at 12,000 cells/ml, and clones were loaded at 5,000 clones/ml in PBS containing 10% v/v OptiPrep and were maintained in suspension using a magnetic micro-stirrer bar placed within the tube. 2) **Barcoding/lysis mix inlet**: An aliquot of barcode beads (Chemgenes Corp.) containing 300,000 barcodes at a concentration of ~400 beads/μL was removed from the stock tube and washed twice with 1 ml lysis solution made of 67.5 μL 10% (v/v) IGEPAL CA-630, 112.5 μL 1 M Tris-HCl (pH 8.0), and 820 μL nuclease-free water. The beads were then resuspended in 1.5 ml 100 μL 10% (v/v) IGEPAL CA-630, 112 μL 0.1 M DTT, 170 μL 1 M Tris-HCl (pH 8.0), 118 μL murine RNase inhibitor, and 1 ml nuclease-free water leading to a concentration of 200,000 barcode beads/ml. Following resuspension, the sample was loaded onto the corresponding inlet in the microfluidics setup. 3) **Carrier oil inlet**: The carrier oil was 10 ml of HFE-7500 with 2% (w/w) fluorosurfactant. During microfluidics runs, cell suspension/encapsulated cells, lysis mix, and collection tubes were kept on ice. The device generates monodispersed droplets with volumes in the range of 0.5 nl for scRNA-seq and 1 nl for CloneSeq. Flow rates used for Drop-seq are shown in Table S1.

Library preparation

The Drop-seq libraries were prepared following a previously published protocol (Macosko et al., 2015). All the primers used in the library preparation are listed in Table S2. The inDrops libraries were prepared using the following procedure: After completion of the microfluidics stage, the collection tubes were exposed to 6.5 J/cm² of a 365-nm UV lamp for 10 min to release photocleavable barcoding primers from the barcoding beads. Next, the collection tubes containing the UV-exposed emulsion were transferred to a reverse transcription reaction at 50 °C for 2 h followed by 15 min at 70 °C to stop the reaction. Each sample was then demulsified by adding 50 μl PFO to release the barcoded cDNA from the droplets. After clear separation of the two phases was observed, the upper aqueous phase containing the barcoded cDNA was transferred to a new well. To remove unused primers and primer dimers, a 1:1 digestion mix was added, containing 20 U/μL Exonuclease I (NEB, Cat #M0293), 20 U/μL HinfI enzyme (NEB, Cat #R0155), x1 Exonuclease I Reaction Buffer (NEB, Cat #B0293), x1 CutSmart buffer (NEB, Cat #B7204), and 30 μl of nuclease-free water. Samples were incubated for 1 h at 37 °C and 10 min at 80 °C. The reaction product (in the form of a cDNA:RNA hybrid) was purified with a 1.5X reaction volume of AMPure XP beads (Beckman Coulter, Cat #A63882) and eluted in 13.5 μl TE buffer. For second strand synthesis, 13.5 μl digestion reaction product was combined with 1.5 μl second-strand synthesis (SSS) buffer and 1 μl of SSS enzyme mix from the NEBNext mRNA Second Strand Synthesis Module (NEB, Cat #E6111) and incubated at 16 °C for 2.5 h, followed by 20 min at 65 °C. For linear amplification by in vitro transcription, SSS reaction products (16 μl) were combined with 24 μl T7 High Yield RNA Synthesis Kit (NEB, Cat #E2040) reagent mix containing 4 μl T7 Buffer, 4 μl ATP, 4 μl CTP, 4 μl GTP, 4 μl UTP, and 4 μl T7 enzyme mix. The reaction was incubated at 37 °C for 13 h, and the resulting RNA was purified with 1.3X reaction volume of AMPure XP beads and eluted with 20 μl TE buffer. An aliquot of 9 μl was frozen for backup at -80 °C, a 2-μl sample was taken for direct analysis, and the remaining 9 μl was used in subsequent library preparation steps. Next, RNA was fragmented using an RNA fragmentation kit (Invitrogen, Cat #AM8740). The 9-μl aliquot of RNA were combined with 1 μl of RNA fragmentation reagent and incubated at 70 °C for 2 min, transferred to ice, and 40 μl fragmentation stop mix containing 5 μl fragmentation stop solution and 35 μl TE buffer was added. Fragmented RNA was purified with a 1.3X reaction volume of AMPure XP beads and eluted in 10 μl TE buffer. The resulting amplified and fragmented RNA was reverse transcribed using a random hexamer primer as follows: first, 10 μl RNA was mixed with 2 μl of

100 μ M PE2-N6-v2 random hexamer primer (PvG748-SBS12-RT) and 1 μ l of 10 mM dNTPs, incubated for 3 min at 65 °C and transferred to ice. Then the following components were added to the reaction: 4 μ l of 5X First-Strand buffer, 1 μ l of 0.1 M DTT, 1 μ l murine RNase inhibitor, and 1 μ l of SuperScript™ III RT enzyme (200 U/ μ L). Samples were incubated at 25 °C for 5 min, 50 °C for 60 min, and 70 °C for 15 min. For the clonal barcoding library, UGI-shifted primers (5'-AGACGTGTGCTCTTCCGATCTGTGACGGATCC-3') were used instead of random primers for half of the sample to amplify the genetic barcode region with incubation at 48 °C for 5 min, 55 °C for 60 min, and 70 °C for 15 min. Following reverse transcription, the reaction volume was raised to 50 μ l by adding 30 μ l nuclease-free water, and the resulting cDNA was purified with 1.2X reaction volume of AMPure XP beads and eluted in 11.5 μ l TE buffer. The resulting libraries were PCR amplified using standard PE1/PE2 full-length primer mix (2p fixed; 2p fixed (+barcode)). The primers contain Illumina library indices for multiplexing. Each PCR reaction consisted of 14 amplification cycles and contained 11.5 μ l post-reverse transcription cDNA library, 12.5 μ l 2x KAPA HiFi HotStart ReadyMix (Roche), and 1 μ l of 25 μ M PE1/PE2 index primer mix. Amplified libraries were purified using a 0.7X reaction volume of AMPure XP beads and eluted in 30 μ l nuclease-free water. Aliquots of 15 μ l of each resulting library were run in 2% agarose gels, and the desired 200–800 bp DNA library fragments were isolated using PureLink™ Quick Gel Extraction Kit (Invitrogen, cat #K210012). For the clonal barcoding library, we aimed for amplification of a band around 600 bp. Library quality was confirmed by Agilent 2200 TapeStation nucleic acid system (Agilent) using the Agilent High Sensitivity D1000 DS DNA kit. The resulting libraries had an average size of 350–550 bp. Size-selected libraries were diluted to 4 nM and combined into a pool for paired-end, single index sequencing on the Illumina NextSeq 550 instrument, using an Illumina 550 High Output v2 (75 cycles) kit. Cycle distribution was 45 cycles for Read 1, 35 cycles for Read 2, and 8 cycles for library index read.

Species-mixing experiments

To determine off-species contamination in our single-cell and clonal preparations, we performed inDrops as described above with a PC9/R1 ESCs cell suspension mixture. The suspension mixtures were 100,000 cells/ml in total (1:1 human:mouse ratio) for the single-cell experiment and 20,000 clones/ml in total (1:1 ratio) for single clone experiment. PC9 cells were identified as those barcodes with greater than 15,000 human transcripts, and R1 mESCs were identified as those with greater than 15,000 mouse transcripts.

Sequencing and data filtering

Paired-end sequencing was performed on Illumina NextSeq 500. Read 1 was used to obtain the sample barcode and UMI sequences and read 2 was mapped to a reference transcriptome. The reads were first filtered based on the presence of two sample barcode components separated by the W1 adaptor sequence in Read 1. Barcodes for each read were matched against a list of the 3842 pre-determined barcodes, and errors of up to two nucleotides mismatch were corrected. Reads with a barcode separated by more than two nucleotides from the reference list were discarded. The reads were then split into barcode-specific files for mapping and UMI filtering.

Single clone alignment and UMI-based filtering

Reads split into clone barcode-specific files were aligned using Bowtie2 (Langmead et al., 2009) to the mouse or human reference transcriptome. Alignments from Bowtie were filtered as follows: (1) For each read, we retained at most one alignment per gene, across all isoforms, by choosing the alignment closest to the end of the transcript. (2) If a read aligned to multiple genes, we excluded any alignments more than 400 bp away from the end of the transcript. This step results in an approximately 5% increase in the number of final UMI reads obtained, as compared to simply discarding any ambiguous read. (3) If a read still aligned to more than two genes after UMI filtering, we excluded the read altogether.

Gene set signature activation analysis

We used MSigDB GSEA mapped to NCI-60 cell lines using GO biological processes and REACTOME gene sets with FDR q-value less than 0.01 as described previously (Mootha et al., 2003; Subramanian et al., 2005).

QUANTIFICATION AND STATISTICAL ANALYSIS

Differential expression analysis

For bulk data analysis the transcript per million (TPM) values were used to compare libraries. Differential gene expression was visualized using xy plots. Statistical analysis was performed for replicates using a two-sided t test, and p values of <0.05 were deemed significant.

PCA and UMAP analysis

scRNA-seq data were analyzed using the Seurat v2.4 pipeline (Stuart et al., 2019). For single cells, cells with more than 5,000 unique molecular identifiers were retained for further analysis. Clones with more than 15,000 unique molecular identifiers were retained for further analysis. A global-scaling normalization was performed on the filtered dataset using “LogNormalize” with a scale factor of 10,000. Identification of highly variable genes was performed with the following parameters: x.low.cutoff = 0.2, x.high.cutoff = 5, y.cutoff = 0.5, and y.high.cutoff = 10. Cell-to-cell variation in gene expression driven by batch, cell alignment rate, and the number of detected molecules were regressed out and a linear transformation was applied. A principal component analysis was performed on the scaled data with 15 principal components. Identification of clusters of cells was done by a SNN modularity optimization-based

clustering algorithm. We first calculated k-nearest neighbors and then constructed the SNN graph. The modularity function was optimized to identify clusters. Clustering was done with resolution of 0.6, and UMAP was used for visualization.

Clonal barcoding analysis

For clonal barcoding experiments, two libraries were generated from the same sample: One was a general library made using a random primer that showed the transcription profile background, whereas the second was made using the UGI-shifted primer that only presented a narrow region around the clone barcodes. The list was matched against the pre-determined barcode reference list, and errors of up to two nucleotides were corrected. Reads with a barcode separated by more than two nucleotides from the reference list were discarded. Some cells carried more than one clone barcode as a result of an MOI greater than 1 at the infection step. Clonal cell origin was determined by matching cell UMIs from the random library and the UGI library. To compare similarities between clonal cells to cells picked at random from the whole population, we first created a UMAP of the random primer library as background, then marked the clonal origins of cells on the plot. The Euclidian distance between cells was calculated and compared between clonal cells and random cells first by coordinates on UMAP then by the entire gene expression matrix. The significance of the difference between the distance within clones and random cells was tested by a Wilcoxon rank sum test.

Clone-to-Clone variation analysis for PC9 cells

Two special analysis were done on the raw RNA expression matrix of PC9 scRNA-seq and CloneSeq data. The first analysis tested the coverage of the two method. We first converted the expression matrix into binary format, in which 0 was kept as 0, and any expression more than 0 was converted to 1. Then we randomly sampled the same number of columns from the clone matrix as the single cell matrix. Next, we calculated the sum of row, which represents the number of cells that the gene is detected to be expressed in and recorded the percentage of the genes passed the defined threshold. Finally, we made line plot of the percentage of genes passing threshold according to the number of cells that the gene has expression in.

The second analysis investigated the expression level of the CSC-related genes in CloneSeq and scRNA-seq data. We first withdraw from the raw matrix the submatrix of CSC-related genes. Then we calculated the sum of columns in each submatrix, which represents the overall expression level of those CSC-related genes in a certain cell/clone. Finally, we made a histogram showing the frequency of the overall CSC-related genes expression in scRNA-seq and CloneSeq, respectively.