

POLYseq: A poly(β -amino ester)-based vector for multifunctional cellular barcoding

Andrew Dunn,¹ Yuqi Cai,¹ Kentaro Iwasawa,¹ Masaki Kimura,¹ and Takanori Takebe^{1,2,3,4,*}

¹Division of Gastroenterology, Hepatology & Nutrition, Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

²Center for Stem Cell and Organoid Medicine (CuSTOM), Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

³Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

⁴Institute of Research, Tokyo Medical and Dental University (TMDU), 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, Japan

*Correspondence: takanori.takebe@cchmc.org

<https://doi.org/10.1016/j.stemcr.2021.07.020>

SUMMARY

Despite evolving biological application of next-generation sequencing (NGS) at single-cell level, current techniques in NGS library preparation restrict multiplexing, necessitating the costly preparation of distinct libraries for each sample. Here, we report the development of a novel poly(β -amino) ester labeling system synthesized with inexpensive, common reagents, termed POLYseq, capable of efficiently delivering fluorescent molecules or sample-distinguishing DNA barcodes through non-covalent binding enabling rapid creation of custom sample pools. Chemical formulation was found to determine cellular labeling propensity. Live image-based tracking of fluorescent conjugated POLYseq vectors demonstrated lysosomal compartmentalization. Barcode labeling was uniformly detected across 90% of cells by single-cell RNA sequencing, allowing for the successful identification of human and mouse cultured cell lines from a single pool. These findings highlight the multifunctional applications of POLYseq in live cell imaging and NGS in a scalable and cost-effective manner.

INTRODUCTION

Next-generation sequencing (NGS) has allowed for unparalleled investigation of genetic characteristics through genomic and transcriptomic landscapes. Single-cell sequencing, advanced by NGS, has revolutionized transcriptomic analysis (Navin et al., 2011; Picelli et al., 2013; Shapiro et al., 2013), giving insight into rare and previously uncharacterized populations in multicellular systems (Buettner et al., 2015). Single-cell RNA sequencing (scRNA-seq) uses a dual barcoding scheme such that every RNA strand captured for sequencing receives its own strand-specific barcode while all RNA strands captured for a single cell receive their own cell-specific barcode (Zheng et al., 2017). As larger sequencers possess the capacity to run multiple single-cell experiments in parallel with adequate sequencing depth, scRNA-seq preparation benefits from the inclusion of sample-specific barcodes affixed prior to library generation as this allows for separate, distinct samples to be pooled together (multiplexed) and prepared in parallel. Multiplexing prior to single-cell processing necessitates a methodology capable of heterogeneously tagging samples with barcodes readable by NGS platforms. Previously investigated techniques relied upon genetic diversity to drive demultiplexing through bioinformatic processing or the expression of barcoding sequences from the creation and generation of viral libraries (Guo et al., 2019; Kang et al., 2018; Kobschull et al., 2016; Kester and van Oudenaarden, 2018). While viral methods are convenient for long-term lineage tracing, the generation and application of viral libraries with high transduc-

tion efficiency for sufficient barcode representation in multiplex applications are time consuming and restrictive for short-term labeling.

Alternative techniques label cells with single-stranded DNA (ssDNA) barcodes for direct capture during single-cell preparation, circumventing the dependence on vector expression at the cost of reduced barcoding longevity. The barcode-conjugated antibody is one common strategy employing ssDNA labeling (Stoeckius et al., 2017, 2018). Antibody methods take advantage of specificity for target differentiation and expression quantification. Introduction of innate barcode heterogeneity through the application of multiple distinct ssDNA sequences (each requiring individualized conjugation), super-loading, and sample multiplexing becomes possible. However, the main detractor of antibody-based methods for ubiquitous labeling is the potential lack of robust universally expressed surface antigens. To overcome this, a complementary technology employs a modification of fatty acids for non-selective integration into cell membranes (Weber et al., 2014). Juxtaposed with antibody methods, lipid labeling seeks to enhance targeting ubiquity at the expense of specificity. Regardless of the chosen method, prior ssDNA labeling techniques have relied upon covalent conjugation of either the barcode directly or a universal annealing oligo labeling mediators through the utilization of click chemistry and solid-phase approaches. There exists an opportunity for the development of a fast, efficient, sample-specific barcoding tool allowing for the creation of custom multiplexed barcoding pools without covalent conjugation restrictions to significantly enhance sequencing throughput and





reduce cost. Therefore, the POLYseq system seeks to provide a robust, ubiquitous labeling system, relying upon charged-based interaction with cells (Dunn et al., 2018), without the requirement of covalent conjugation for enhancing custom multiplex applications.

RESULTS

Synthesis and characterization

Synthesis and application scheme for POLYseq vectors is detailed in Figure 1A using commercially available monomers (Table S1) mixed in specific ratios detailed in Table S1. Acrylate monomers (D or V) mixed with an amino alcohol (S) are heated to form the uncapped acrylate-terminated vector. Vectors are then capped through the addition of a primary or secondary amine-containing small molecule (C). POLY1–4 and POLY5–8 were each created with the same acrylate backbone, only differing by the capping molecule. This capping step imparts the ability for POLYseq vectors to adhere to cells (labeled cells). Labeled cells may then be processed using standard single-cell techniques. All respective reagents are commercially available (Figure 1B). ^1H NMR confirmed the presence of terminal acrylate groups following the production of the acrylate terminated product; resonant peaks for these groups were observed at δ 6.2–5.6 and disappeared upon successful conjugation with capping reagents (Figure 1C). The ability to bind single-strand DNA barcodes used in cell hashing experiments was found to be dependent upon capping reagent and backbone structure (Figure 1D). Complete binding of ssDNA is directly observed by the absence of DNA migration with weak binding observed from band smearing. Vectors capped with molecules **C2** and **C3** more readily retain ssDNA barcodes during electrophoresis than those capped with **C1** or **C4**. Moreover, the inclusion of branching acrylate **V5** reduced the mass ratio (w/w) at which complete barcode retention was observed (POLY2 versus POLY6, POLY3 versus POLY7).

Cell targeting

Targeting propensity of POLYseq vectors 1–4 was initially tested using fluorescence-activated cell sorting (FACS) analysis of labeled anterior and posterior gut spheroids in a fusion model (Koike et al., 2019, 2021). Instead of monolayer and single lineage culture, this 3D fusion model was used to rigorously assess labeling efficiency and fidelity of POLYseq candidates as individual spheroids were labeled with POLYseq vectors conjugated with distinct fluorescent molecules allowing for cross-labeling identification following fusion. Dye conjugation efficiency was quantified through electrophoresis. Fluorescent intensity of free and bound dye was used to calculate percentage

conjugation and was found to be $73.4\% \pm 5.6\%$, $90.1\% \pm 0.6\%$, $89.1\% \pm 1.0\%$, and $74.7\% \pm 1.7\%$ for POLY1, POLY2, POLY3, and POLY4 respectively (Figure S1A). Gating analysis for isolated single cells is shown in Figure 2A. Variance in the extent of total labeling as well as double labeling was observed to be dependent on vector formulation (Figure 2B). No significant differences in targeting propensity were observed between POLY1 and POLY3 24 h post spheroid fusion. POLY4 provided a significantly lower percentage of total targeted cells juxtaposed with either POLY1, POLY2, or POLY3 ($p < 0.01$, $n = 3$). Vector POLY3 provided the greatest extent of double labeling and was significantly higher than POLY1, POLY2, and POLY4 ($p < 0.01$, $n = 3$) (Figure 2B). Spheroids fused following labeling with POLY2 showed distinct labeling with a visible boundary (Figure S1B).

Human liver organoids (HLOs) were generated from iPSCs (Ouchi et al., 2019; Shinozawa et al., 2021). Utility of POLY2 complexed with 1 μg of barcode in binding HLOs was further examined using FACS analysis of isolated single cells from mixed cultures tagged with three separate colors: DyLight 488, 550, and 650 (Figure 2C); a total labeling percentage of $94.3\% \pm 1.3\%$ was calculated from FACS of single cells isolated from HLO cultures (Figure 2D). Double-labeled cells within the mixed culture by FACS analysis was negligible ($<1\%$). It was next determined if HLOs may be rapidly stained in basal hepatocyte culture medium (HCM). HLOs were incubated with POLY2 complexed with 1 μg of barcode oligo at a concentration of 10 or 20 $\mu\text{g}/\mu\text{L}$ for 1 h (10 \times or 20 \times w/w loading respectively); 10 \times w/w loading achieved $84.7\% \pm 3.0\%$ while 20 \times w/w provided $94.1\% \pm 0.6\%$ within this time window (Figure 2E). Impact on cellular viability of POLY2 was assessed on ESH9 cells. At 90% confluency, cultures were tagged with POLY2 complexed with barcode oligo at a w/w ratio of 10 over a POLY2 concentration range of 0–50 $\mu\text{g}/\mu\text{L}$ for 1 h. Viability was measured 24 h later. Significant impact on cell viability was found only at 50 $\mu\text{g}/\mu\text{L}$ juxtaposed with controls (one-way ANOVA $p < 0.001$, t test $p < 0.05$) (Figure 2F). Confocal analysis revealed strong colocalization with lysosomes (Figure 2G) for POLY2 and POLY3, while POLY4 had lower internalization at 3 h, mirroring weaker labeling found by flow cytometry of fused spheroids. These results suggest a correlation between each vector's ability to bind barcodes and interact with cells.

Barcoding in 10 \times single-cell RNA sequencing

To test the ability for POLYseq vectors to deliver barcodes that may be amplified by the standard 10 \times Chromium workflow and read by common next-generation sequencers, and three HLO samples were individually tagged with three distinct barcodes using vector POLY2 for 1 h prior to being run individually on the 10 \times Chromium

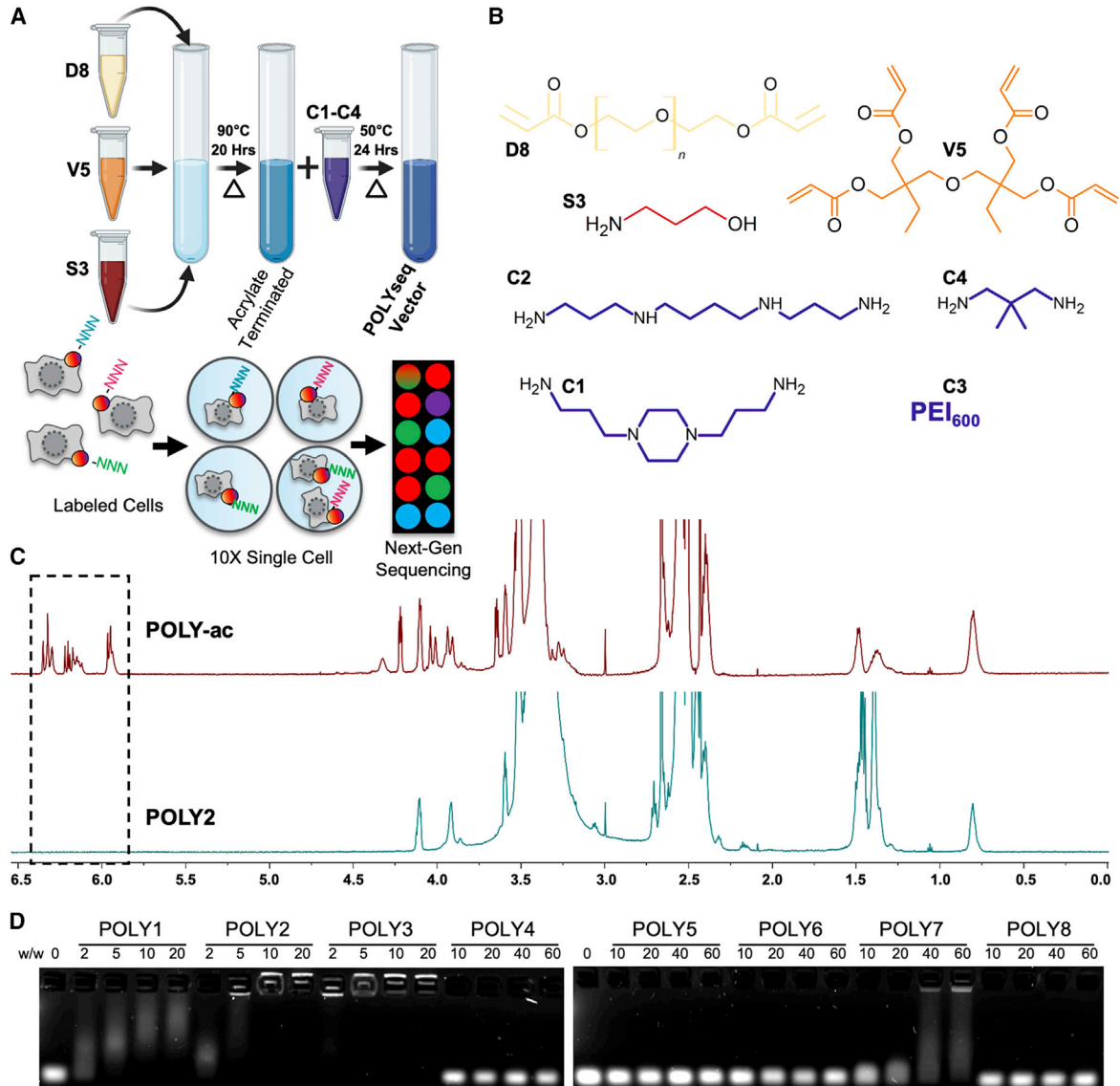


Figure 1. POLYseq characterization

(A) Synthesis and barcoding schematic. Three reagents are used to generate the acrylate-terminated polymer: the poly(ethylene glycol) diacrylate $M_n = 250$ (D8), di(trimethylolpropane) tetraacrylate (V5), and 3-amino-1-propanol (S3). Polymers are then capped with one of four reagents (C1–C4). The final POLYseq system is then used to bind ssDNA barcodes to cells for NGS applications by 10× single-cell processing.

(B) Reagents used in the creation of the POLYseq system.

(C) ^1H NMR spectrum of acrylate-terminated (POLY-ac) and spermine capped POLY2 vectors with resonance from terminal alkenes highlighted by the dashed box.

(D) Gel electrophoresis of ssDNA barcodes bound by POLYseq at indicated mass ratios.

platform. Single-cell analysis of barcoded HLOs containing all sequenced barcodes revealed a high extent of labeling across the three populations with a total extent of labeling near 90% (Figures 3A and 3B). Barcoding accuracy for all three barcodes was 94%. Barcoding uniformity across clusters was confirmed for all three samples with an average la-

beling per cluster of $85\% \pm 4.7\%$ (Figures S2 and S3), with the majority of reads arising from singlets (Figure S2D). Multicellularity has been demonstrated in the HLO culture system (Koike et al., 2021). Heterogeneous barcoding potential was further demonstrated through HLO lineage identification (MacParland et al., 2018). Hepatocytes,

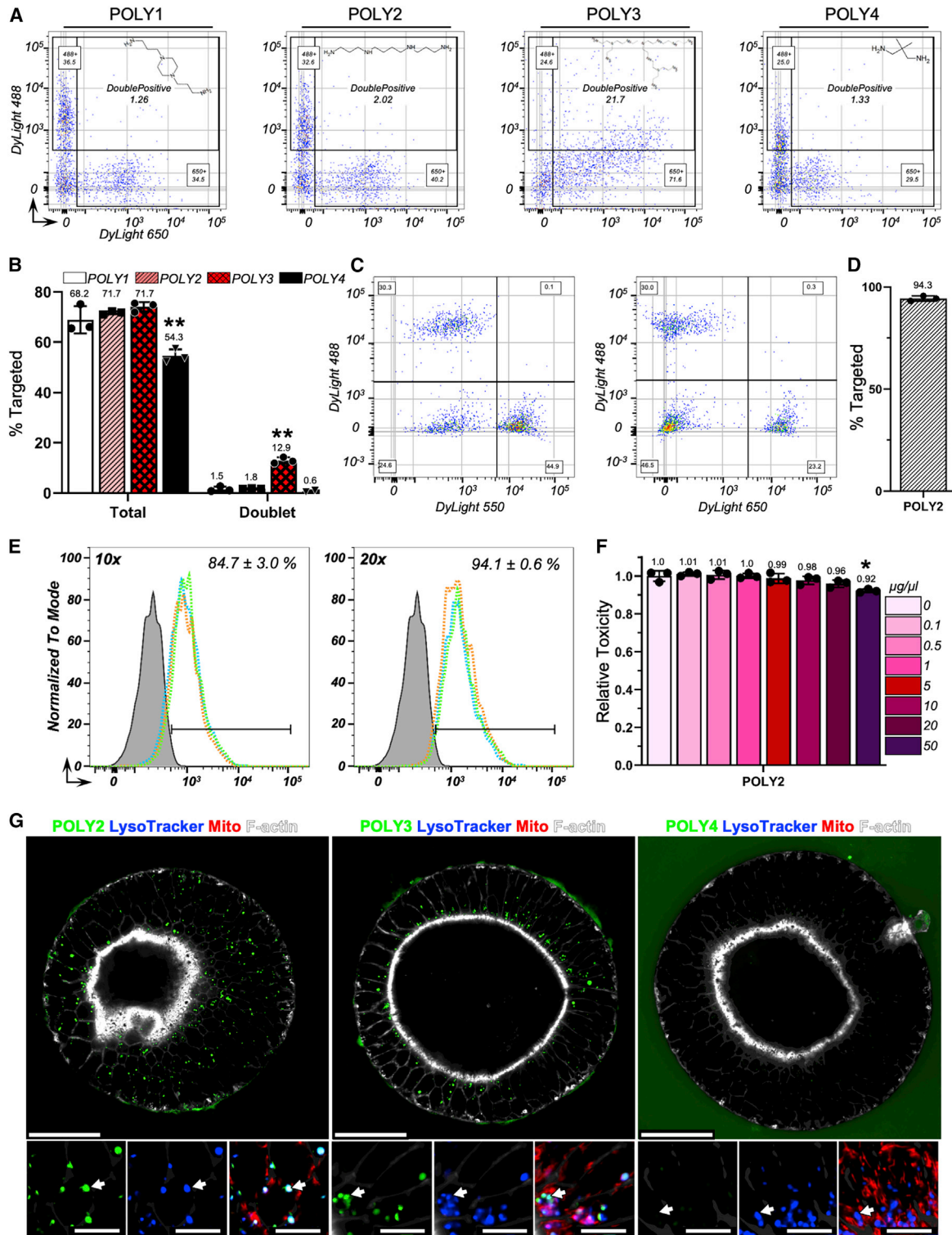


Figure 2. POLYseq cellular targeting and internalization

(A) FACS analysis gating of fused spheroids pre-tagged with DyLight 488 or DyLight 650 conjugated POLYseq vectors demonstrating singlet and double labeling.

(B) Total labeled and double-labeled cells by FACS analysis for 24 h post spheroid fusion (mean \pm SD, n = 3 independent replicates).

(legend continued on next page)



stellate cells, and biliary cells possessed a significant degree of representation among the barcoded population (Figure 3C). Barcode reads were found to be uniform across these populations (Figure 3D).

To determine the ability of POLYseq to multiplex samples in a single run, human and mouse cultured cell lines with distinct transcriptomes were utilized to examine the ability of POLYseq to correctly barcode heterogeneous pools and juxtaposed with the antibody-based method TotalSeq (Figure 4A). Flow cytometry analysis identified a labeling time of 5 min at 37°C utilizing either 0.2 µg or 1 µg of barcode complexed with 2 µg or 10 µg of POLY2 respectively and achieved 99% labeling (Figure 4B); this labeling time with 0.2 µg of barcode was therefore chosen for cell hashing. UMAP (uniform manifold approximation and projection) clustering, automatic cell type identification based on barcode reads (Figure S3A), singlet/doublet identification, and subsequent doublet exclusion, a process specific to hashing (Figures S3B and S3C), were performed in Seurat for both the POLYseq (Figure 4C) and TotalSeq (Figure 4E) libraries constructed from seven cell lines (3T3, B16-F10, embryonic stem [ES], ESH9, HepG2, HUVEC, and MEG-01). The POLYseq library consisted of two conditions for HepG2 cultures, low and high glucose, for a total of eight pooled samples. Distinct transcriptomes of cultured lines allowed for automatic clustering when performing UMAP analysis of all cells in Seurat. Verification of cell populations identified by barcoding through quantification of gene expression highly correlated with each cell line was achieved for pools hashed by POLYseq and TotalSeq staining strategies. Chosen genes for 3T3, B16-F10, ES, ESH9, HepG2, HUVEC, and MEG-01 were COL1A2, PAX3, SOX2, POU5F1, ALB, CDH5, and HBG2 respectively. Heatmap gene expression of identified cell lines from both POLYseq and TotalSeq libraries showed distinct upregulation of gene clusters (Figure 4G). Average labeling accuracy from all identified, barcoded singlets was calculated to be 89.5% ± 6.7% and 87.0% ± 9.3% for POLYseq and TotalSeq libraries respectively and was found not to be statistically significant (Figure 4H, $p = 0.56$).

To demonstrate the applicability of POLYseq to differentiate cells with similar transcriptomes for cluster analysis, HepG2 cultured in minimal essential medium (MEM) was compared with HepG2 cultured under high glucose. HepG2 cultured in high glucose (25 mM) was identified from hashing (HepG2 + glucose) and juxtaposed with HepG2 cultured in low glucose (5.6 mM) MEM (Figure 4I). Identification of significantly upregulated pathways under high glucose were identified using generally applicable gene set enrichment analysis (GAGE) comparing both conditions following normalization over library size, sequencing depth, and log₂ transformation. Glycolysis was found to be significantly upregulated (Figure S4A, Table S3, hsa00010, $q = 4.7 \times 10^{-2}$). Differential expression analysis of upregulated metabolic markers (Figure 4J) revealed G6PC (logFC = 0.43, $p = 1.2 \times 10^{-17}$), ALDOA (logFC = 0.55, $p = 1.4 \times 10^{-206}$), ENO1 (logFC = 0.64, $p = 1.7 \times 10^{-197}$), LDHA (logFC = 0.9, $p = 2.1 \times 10^{-210}$). Clustering within the control group revealed differential upregulation of SOX4 ($p = 3.5 \times 10^{-24}$) and S100A6 ($p = 5 \times 10^{-28}$) (Figure 4K). Interestingly, upregulation of proliferation-associated genes within cluster 0 was found to be differentially expressed within the high-glucose condition (Figure 4L); CDK1 (logFC = 0.89, $p = 1.4 \times 10^{-36}$), STMN1 (logFC = 1.1, $p = 1.5 \times 10^{-55}$), TOP2A (logFC = 2.1, $p = 9.6 \times 10^{-152}$), and TPX2 (logFC = 1.5, $p = 6.1 \times 10^{-107}$). Top up-regulated pathways from GAGE were found to be DNA replication (Figure S4B, Table S3, hsa03030, $q = 2.9 \times 10^{-12}$) and cell cycle (Figure S4C, hsa04110, $q = 4.96 \times 10^{-10}$).

DISCUSSION

The synthesis of a cationic polymer is a necessary step in the creation of a vector capable of binding nucleic acids (Dunn et al., 2018). The ability of POLYseq vectors to rapidly bind and retain hashing ssDNA barcodes for single-cell applications was examined using gel electrophoresis. Vectors with branching acrylate monomers (V5) and capped with monomers containing

(C) FACS dot plots from a single mixed sample of HLOs individually tagged with DyLight conjugated POLYseq vectors over 24 h. (Left) Dot plot of DyLight 488 versus DyLight 550 fluorescence. (Right) Dot plot of DyLight 488 versus DyLight 650 fluorescence.

(D) Quantified total percentage of targeted cells by FACS isolated from HLO cultures (mean ± SD, $n = 3$ independent replicates).

(E) Fluorescence of isolated single cells from HLOs tagged with POLYseq vectors in 1 h using either 10× or 20× w/w loading (1 µg barcode, 10 µg POLYseq or 1 µg barcode, 20 µg POLYseq).

(F) Toxicity imparted by POLY2 on ESH9 cultures following 1 h of tagging and 24 h of culture. Significant difference identified by one-way ANOVA (mean ± SD, $n = 3$ independent replicates).

(G) Lysosomes (blue), POLYseq vectors (green), mitochondria (red), and F-actin (white) are used to track the localization of vectors within HLOs by confocal microscopy 3 h post tagging. Whole HLOs are shown with POLYseq fluorescence and F-actin staining. Scale bar, 50 µm. Inset images show lysosomal colocalization. Scale bar, 10 µm. * $p < 0.05$, ** $p < 0.01$ calculated by an unpaired t test assuming unequal variance.

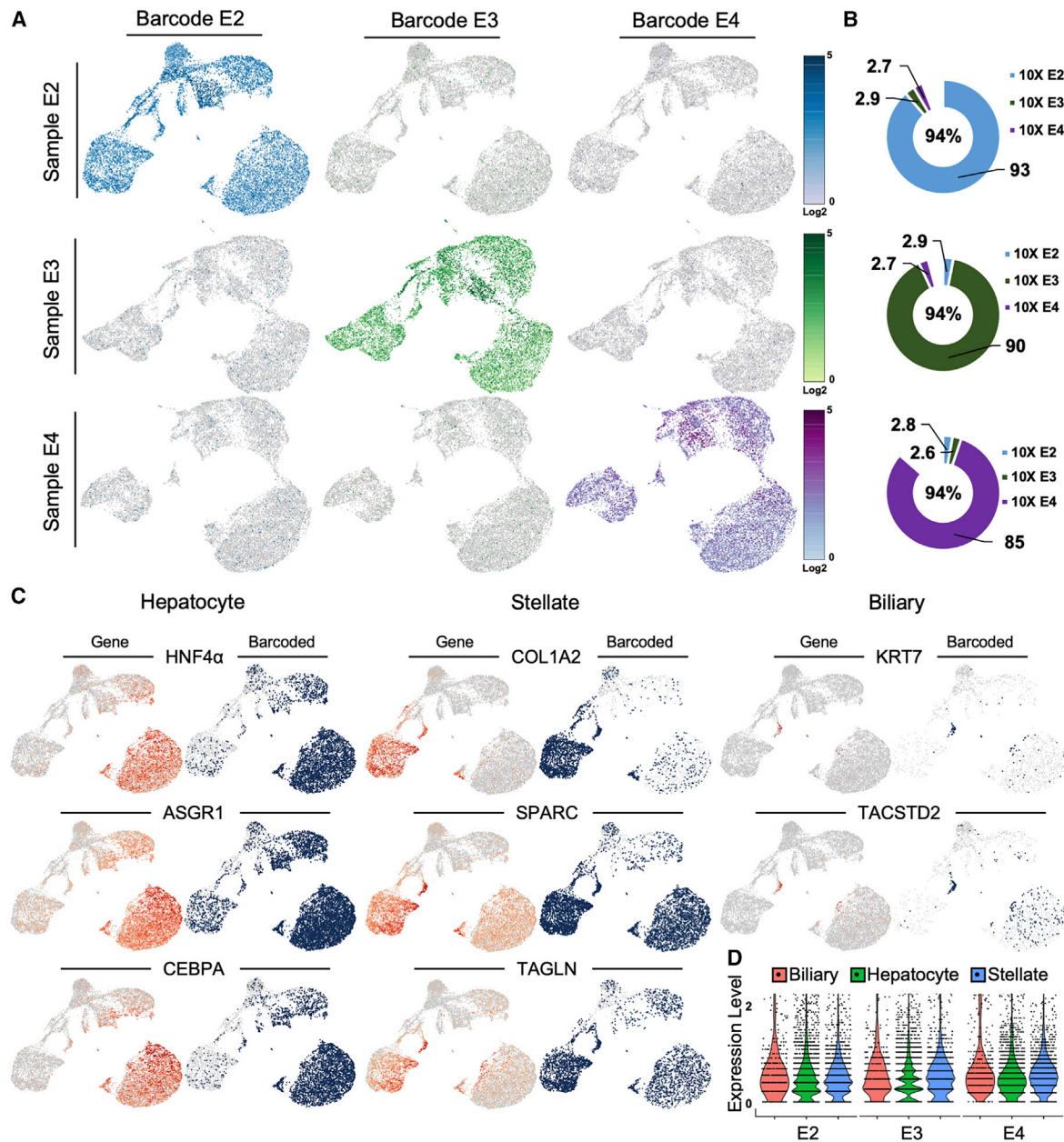


Figure 3. Cellular coverage of POLYseq barcoding in 10× single-cell RNA sequencing

(A) Barcode reads in three tagged HLO samples using distinct ssDNA barcodes and individually run on the 10× single-cell platform. (B) Percentage of cells aligned to each of the three barcodes within each sample with targeting accuracy (inset). (C) HLO lineages identified by gene expression and respective barcoded populations contained within each expressed population for hepatocytes (HNF4 α , ASGR1, CEBPA, RBP4), stellate (COL1A2, SPARC, TAGLN), and biliary (KRT7, TACSTD2). (D) Barcode reads within biliary, hepatocyte, and stellate populations for samples E2, E3, and E4.

a high density of primary and secondary amines (POLY2, POLY3) most readily bound and retained ssDNA barcodes under physiological pH, whereas respective linear vectors showed a reduction in binding ability (Figure 1D). The success of ssDNA binding is therefore a combination of branching architecture and cap type.

Quantification of cell targeting was achieved using flow cytometry to track fluorescently labeled vectors in a model anterior/posterior gut boundary fusion system (McMahon and Boucrot, 2011). Cellular labeling was dependent on vector formulation. Based on ssDNA binding efficiency and cell targeting performance, POLY2 was further

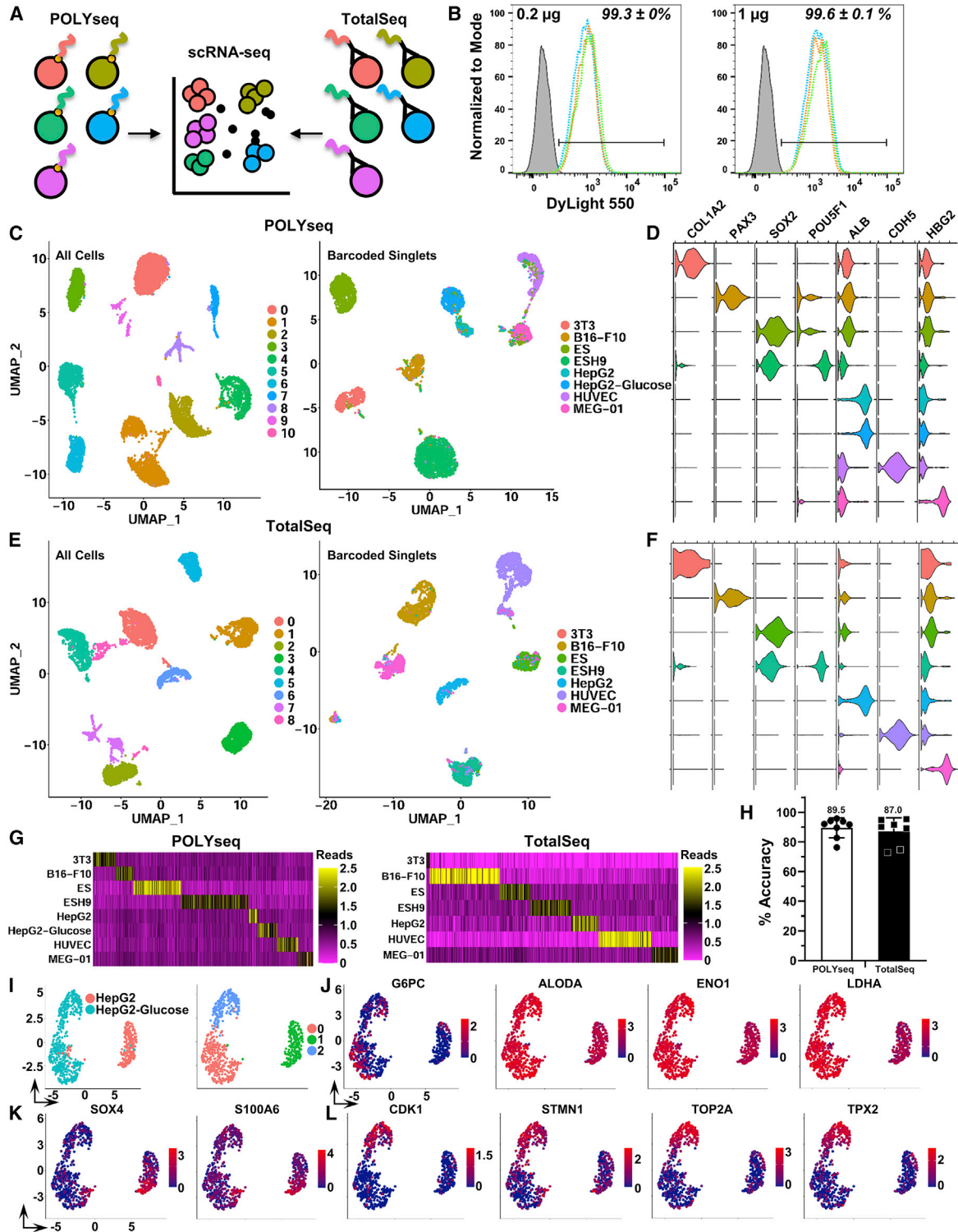


Figure 4. NGS multiplexing using POLYseq

(A–C) (A) POLYseq and TotalSeq pooling schematic. (B) Flow cytometry analysis of ESH9 cells tagged with POLYseq vectors in 5 min at 37°C. (C) UMAP clustering of POLYseq tagged *in vitro* cultures showing all cells (automatically clustered in Seurat) and barcoded singlets (cell type defined by barcode reads).

(legend continued on next page)



investigated for its potential in single-cell barcoding applications of HLO cultures. FACS analysis revealed rapid integration of POLY2 in 1 h in nearly all cells from HLO cultures in HCM with no appreciable double labeling and minimal toxicity. Distinctive labeling 24 h after barcoding suggests the ability to pool organoid samples. Confocal analysis of fluorescent conjugated POLYseq revealed formulation-dependent colocalization within lysosomes. As lysosomal sequestration is generally associated with maturation or fusion of late endosomes from early endosomes trafficked from clathrin-dependent, dynamin-dependent endocytosis or macropinocytosis, it suggests that cellular association of vector POLY2 and POLY3 readily occurs prior to this time point (Luzio et al., 2007; Mayor and Pagano, 2007; McMahon and Boucrot, 2011). Although the internalization mechanism is molecularly unknown, this selective association provides investigative opportunities into time-dependent endosomal/lysosomal organelle trafficking. Direct application of POLYseq barcoding to other single sequencing techniques such as single nuclear RNA sequencing (snRNA-seq) is an important topic of study. Through its present application, POLYseq may not interact with the nuclear membrane unless applied directly to nuclear isolates.

POLY2 was found to have the most attractive qualities of barcode binding, cellular labeling, and minimal toxicity. Moreover, POLY2 possessed the capacity to deliver readable barcodes and rapidly bind single-cell solutions by FACS analysis. This labeling speed provided the opportunity to rapidly barcode single-cell suspensions of diverse cell types and allowed for successful identification of directly pooled distinct human and mouse cell lines solely by barcode reads, providing a highly competitive hashing strategy compared with antibody-based methods without the restriction of specific target antigens. Moreover, labeling time mirrored another antigen-independent method, Multiplexing using lipid-tagged indicies (MULTI-seq, (McGinnis et al., 2019)), and was a vast improvement over labeling

time using traditional lipofection (Shin et al., 2019). Non-significant difference in labeling accuracy was observed for POLYseq hashing compared with TotalSeq; POLYseq allowed for the direct exclusion of doublets, an important feature requiring barcode integration, and the investigation into heterogenic gene expression patterns within pooled cells with highly similar transcriptomes. The direct juxtaposition of HepG2 under a high-glucose culture condition with low glucose was easily achieved by sub-clustering on barcode reads and revealed specific upregulation of glycolytic markers by differential gene expression analysis of significant upregulation of the glycolysis pathway by GAGE analysis. Differential expression of SOX4, a positive regulator of apoptosis, S100A6, glucose-6-phosphatase, CDK1, and Stathmin1 (STMN1) was revealed. Together, differential expression of DNA topoisomerase II alpha (TOP2A), associated with poor prognosis (Wong et al., 2009), and TPX2, involved in microtubule assembly, suggests a prominent, heterogenic metabolic response *in vitro*; this is an important implication when selecting HepG2 for metabolic and physiological studies during drug screening.

The cost of synthesizing POLYseq vector is 3 cents/mg with 1–10 μ g used in total per barcoded pool, and this achieved comparable labeling accuracy in a pooled hashing experiment juxtaposed with commonly used antibody-based methods without reliance on specific target antigen expression. Importantly, POLYseq labeled cells within 5 min and enabled the pooling, direct identification, and subsequent differential expression analysis of the same cell type (HepG2) under varying conditions. With the ability for efficient fluorescent conjugation, specific intracellular vesicle sequestration, and rapid delivery of NGS-readable ssDNA barcodes into cells without covalent conjugation, the POLYseq system provides the opportunity to inexpensively generate custom hashed pools for multiplex applications, reducing sequencing cost without reliance on specific target antigens.

(D) Respective gene expression for identified cell types from POLYseq barcoding: 3T3 (COL1A2), B16-F10 (PAX3), ES (SOX2), ESH9 (POU5F1), HepG2 (ALB), HUVEC (CDH5), MEG-01 (HBG2).

(E) UMAP clustering of TotalSeq tagged *in vitro* cultures showing all cells (automatically clustered in Seurat) and barcoded singlets (cell type defined by barcode reads).

(F) Respective gene expression for identified cell types from TotalSeq barcoding.

(G) Gene heatmaps for identified cells from POLYseq and TotalSeq libraries.

(H) Labeling accuracy within UMAP clusters for POLYseq ($89.5\% \pm 6.7\%$, mean \pm SD, $n = 8$ independent replicates) and TotalSeq (mean \pm SD, $87.0\% \pm 9.3\%$, $n = 7$ independent replicates).

(I) UMAP clustering of HepG2 + glucose versus HepG2.

(J) Metabolic genes are differentially upregulated in the high-glucose condition compared with control.

(K) Heterogenic gene expression among control clusters.

(L) Heterogenic gene expression is differentially upregulated in cluster 2 compared with cluster 0 within the high-glucose condition associated with DNA replication and cell proliferation.



EXPERIMENTAL PROCEDURES

Complete methods are included in the supplemental section.

Synthesis: POLYseq vectors were synthesized through Michael addition in two-step processes. NMR was performed on a Bruker Ascend 600 MHz spectrometer. An aliquot of 5 mg of either acrylate-terminated or capped vectors was directly dissolved in deuterated DMSO-d₆ (Sigma-Aldrich, United States).

Cell culture: 3T3 and B16-F10 were maintained in DMEM + GlutaMAX + 10% FBS. HUVECs (human umbilical vein endothelial cells) were maintained on gelatin in EGM-2 (Lonza). Hep G2 was maintained either in MEM or DMEM + 10% FBS. MEG-01 was maintained in RPMI + 10% FBS. Mouse ES cells were maintained feeder free on gelatin in serum-free ES medium + 2i + LIF (Neurobasal; 250 mL, Gibco), DMEM/F-12 (250 mL Gibco), N-2 supplement (2.5 mL, Gibco), B27 + Retinoic acid (5 mL, Gibco), 7.5% KnockOut Serum Replacement (Gibco), GlutaMAX (1× final, Gibco), monothioglycerol (6.3 μL, Sigma), LIF (1000 U/mL final, ESGRO), CHIR99021 (3 μM final, Tocris), PD0325901 (1 μM final, Tocris). Stem cells were maintained as previously described with slight modifications (Koike et al., 2019). All stem cells were maintained in feeder-free conditions using mTeSR (StemCell Technologies, Vancouver, Canada) at 37°C in 5% CO₂.

Flow cytometry: Cultures for spheroid labeling were tagged by DyLight conjugated POLYseq vectors overnight at a concentration of 20 μg/mL. Spheroids were allowed to form overnight. Anterior and posterior spheroids were mixed, fused over 24 h, and dissociated into single cells for analysis by multi-color flow cytometry.

Immunofluorescence: HLOs were incubated with DyLight conjugated POLYseq vectors diluted in HCM. F-actin was stained using SiR-Actin (Cytoskeleton, Inc., United States). Mitochondria were stained using tetramethylrhodamine, methyl ester (TMRM; Thermo Fisher Scientific). Lysosomes were stained with LysoTracker Blue DND-22 (Thermo Fisher Scientific).

Cell barcoding: POLY2 was mixed with 10× compatible DNA barcoding oligomers based on the CITE-seq (cellular indexing of transcriptomes and epitopes by sequencing) cell hashing oligomer structure (Table S2) at a mass ratio of 10 μg vector/1 μg oligo in 100 μL of HCM. HLOs were barcoded at 37°C for 1 h. Cells for pooling were stained with 0.2 μg of barcode in 100,000 cells/100 μL of OptiMEM for 5 min at 37°C. Prepared scRNA-seq libraries were run on the NovaSeq 6000 system.

Statistics

Data is reported as the mean ± standard deviation (SD). Statistical significance was determined with an alpha cutoff of 0.05. *p* values less than 0.05 were considered statistically significant. The Student's *t* test assuming unequal variance was used when comparing two means. Statistical analysis on viability (Figure 2F) was performed using one-way ANOVA. Pathway and gene enrichment analysis was performed through GAGE v2.38.3.

Data and code availability

The accession number for the HLO data sets reported in this paper are GEO: GSM4992600, GEO: GSM4992603, and GEO: GSM4992605.

The accession number for the POLYseq data reported in this paper is GEO: GSM4992607. The accession number for the TotalSeq data reported in this paper is GEO: GSM4992608.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.stemcr.2021.07.020>.

AUTHOR CONTRIBUTIONS

A.D. and T.T. conceived the study. A.D., Y.C., K.I., and M.K. performed the experiments. T.T. supervised the findings of this work. All authors discussed the results and contributed to the final article.

CONFLICT OF INTERESTS

A.D. and T.T. are listed as inventors for the POLYseq-related intellectual property.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to Drs. Jim Wells, Jorge Bezerra, Aaron Zorn, and Takebe laboratory members for their support and excellent technical assistance, and Wendy Thompson and Mary Koch for kind administrative/technical assistance. Illustrations were created using [Biorender.com](https://biorender.com). This work was supported by Cincinnati Children's Research Foundation grant, Ohio TVSF grant, CCHMC Innovation Acceleration fund, and NIH Director's New Innovator Award (DP2 DK128799-01). This work was also supported by an NIH grant UH3 DK119982, Cincinnati Center for Autoimmune Liver Disease Fellowship Award, PHS grant P30 DK078392 (Integrative Morphology Core and Pluripotent Stem Cell and Organoid Core) of the Digestive Disease Research Core Center in Cincinnati, the Falk Catalyst Research Awards Program, JST Moonshot R&D grant number JPMJMS2033 and JPMJMS2022, Takeda Science Foundation award, Mitsubishi Foundation award and AMED CREST (20gm1210012h0001), JP19fk0210037, JP19bm0704025, JP19fk0210060, JP19bm0404045, and JSPS JP18H02800, 19K22416. T.T. is a New York Stem Cell Foundation Robertson Investigator. All single-cell datasets were analyzed in Loupe v4.0 and Seurat v3.2.2. Gene set enrichment analysis was performed using GAGE v2.38.3. Pathways were visualized using Pathview v1.28.1 (Luo and Brouwer, 2013). POLYseq formulations presented in this article have been filed for provisional patent 62/855,448. The *t* tests were calculated as unpaired assuming unequal variance.

Received: December 21, 2020

Revised: July 30, 2021

Accepted: July 30, 2021

Published: August 26, 2021

REFERENCES

Buettner, F., Natarajan, K.N., Casale, F.P., Proserpio, V., Scialdone, A., Theis, F.J., Teichmann, S.A., Marioni, J.C., and Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in



- single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* **33**, 155.
- Dunn, A.W., Kalinichenko, V.V., and Shi, D. (2018). Highly efficient in vivo targeting of the pulmonary endothelium using novel modifications of polyethylenimine: an importance of charge. *Adv. Healthc. Mater.* **7**, 1800876.
- Guo, C., Kong, W., Kamimoto, K., Rivera-Gonzalez, G.C., Yang, X., Kirita, Y., and Morris, S.A. (2019). CellTag indexing: genetic barcode-based sample multiplexing for single-cell genomics. *Genome Biol.* **20**, 90.
- Kang, H.M., Subramaniam, M., Targ, S., Nguyen, M., Maliskova, L., McCarthy, E., Wan, E., Wong, S., Byrnes, L., and Lanata, C.M. (2018). Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* **36**, 89.
- Kebschull, J.M., da Silva, P.G., Reid, A.P., Peikon, I.D., Albeanu, D.F., and Zador, A.M. (2016). High-throughput mapping of single-neuron projections by sequencing of barcoded RNA. *Neuron* **91**, 975–987.
- Kester, L., and van Oudenaarden, A. (2018). Single-cell transcriptomics meets lineage tracing. *Cell Stem Cell* **23**, 166–179.
- Koike, H., Iwasawa, K., Ouchi, R., Maezawa, M., Giesbrecht, K., Saiki, N., Ferguson, A., Kimura, M., Thompson, W.L., Wells, J.M., and Takebe, T. (2019). Modelling human hepato-biliary-pancreatic organogenesis from the foregut-midgut boundary. *Nature* **574**, 112–116.
- Koike, H., Iwasawa, K., Ouchi, R., Maezawa, M., Kimura, M., Kodaka, A., Nishii, S., Thompson, W.L., and Takebe, T. (2021). Engineering human hepato-biliary-pancreatic organoids from pluripotent stem cells. *Nat. Protoc.* **16**, 919–936.
- Luo, W., and Brouwer, C. (2013). Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* **29**, 1830–1831.
- Luzio, J.P., Pryor, P.R., and Bright, N.A. (2007). Lysosomes: fusion and function. *Nat. Rev. Mol. Cell Biol.* **8**, 622–632.
- MacParland, S.A., Liu, J.C., Ma, X., Innes, B.T., Bartczak, A.M., Gage, B.K., Manuel, J., Khuu, N., Echeverri, J., and Linares, I. (2018). Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat. Commun.* **9**, 1–21.
- Mayor, S., and Pagano, R.E. (2007). Pathways of clathrin-independent endocytosis. *Nat. Rev. Mol. Cell Biol.* **8**, 603.
- McGinnis, C.S., Patterson, D.M., Winkler, J., Conrad, D.N., Hein, M.Y., Srivastava, V., Hu, J.L., Murrow, L.M., Weissman, J.S., and Werb, Z. (2019). MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat. Methods* **16**, 619.
- McMahon, H.T., and Boucrot, E. (2011). Molecular mechanism and physiological functions of clathrin-mediated endocytosis. *Nat. Rev. Mol. Cell Biol.* **12**, 517.
- Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., Cook, K., Stepansky, A., Levy, D., and Esposito, D. (2011). Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90.
- Ouchi, R., Togo, S., Kimura, M., Shinozawa, T., Koido, M., Koike, H., Thompson, W., Karns, R.A., Mayhew, C.N., McGrath, P.S., et al. (2019). Modeling steatohepatitis in humans with pluripotent stem cell-derived organoids. *Cell Metab.* **30**, 374–384.e6.
- Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098.
- Shapiro, E., Biezuner, T., and Linnarsson, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* **14**, 618–630.
- Shin, D., Lee, W., Lee, J.H., and Bang, D. (2019). Multiplexed single-cell RNA-seq via transient barcoding for simultaneous expression profiling of various drug perturbations. *Sci. Adv.* **5**, eaav2249.
- Shinozawa, T., Kimura, M., Cai, Y., Saiki, N., Yoneyama, Y., Ouchi, R., Koike, H., Maezawa, M., Zhang, R., and Dunn, A. (2021). High-fidelity drug-induced liver injury screen using human pluripotent stem cell-derived organoids. *Gastroenterology* **160**, 831–846.e10.
- Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P.K., Swerdlow, H., Satija, R., and Smibert, P. (2017). Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865.
- Stoeckius, M., Zheng, S., Houck-Loomis, B., Hao, S., Yeung, B.Z., Mauck, W.M., Smibert, P., and Satija, R. (2018). Cell hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 1–12.
- Weber, R.J., Liang, S.I., Selden, N.S., Desai, T.A., and Gartner, Z.J. (2014). Efficient targeting of fatty-acid modified oligonucleotides to live cell membranes through stepwise assembly. *Bio-macromolecules* **15**, 4621–4626.
- Wong, N., Yeo, W., Wong, W., Wong, N.L., Chan, K.Y., Mo, F.K., Koh, J., Chan, S.L., Chan, A.T., and Lai, P.B. (2009). TOP2A overexpression in hepatocellular carcinoma correlates with early age onset, shorter patients survival and chemoresistance. *Int. J. Cancer* **124**, 644–652.
- Zheng, G.X., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., and Zhu, J. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 1–12.