Chapter Title:

Functionalized lineage tracing for the study and manipulation of heterogeneous cell populations

Authors:

Andrea Gardner[*,1] , Daylin Morgan[*,1], Aziz Al'Khafaji[2], Amy Brock[1,2]

Affiliations:

1. Department of Biomedical Engineering, The University of Texas at Austin, Austin TX
2. Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin TX

* *equal contribution*

Corresponding Author:

Amy Brock, Ph.D.; amy.brock@utexas.edu

Running Head (short version of chapter title):

Functionalized lineage tracing

i. Summary/Abstract

The ability to track and isolate unique cell lineages from large heterogeneous populations increases the resolution at which cellular processes can be understood under normal and pathogenic states beyond snapshots obtained from single cell RNA sequencing (scRNA-seq). Here, we describe the Control of Lineages by Barcode Enabled Recombinant Transcription (COLBERT) method in which unique single guide RNA (sgRNA) barcodes are used as functional tags to identify and recall specific lineages of interest. A sgRNA barcode is stably integrated and actively transcribed, such that all cellular progeny will contain the parental barcode and produce a functional sgRNA. The sgRNA barcode has all the benefits of a DNA barcode and added functionalities. Once a barcode pertaining to a lineage of interest is identified, the lineage of interest can be isolated using an activator variant of Cas9 (such as dCas9-VPR) and a barcode-matched sequence upstream of a fluorescent reporter gene. CRISPR activation of the fluorescent reporter will only occur in cells producing the matched sgRNA barcode, allowing precise identification and isolation of lineages of interest from heterogeneous populations.

ii. Key Words:

lineage tracking, single cell, barcoding, clonal dynamics, RNA-seq, scRNA-seq, CRISPR, lineage isolation

1. Introduction

Insight into the clonal composition of a cells during key events – such as development, infection, tumor progression, or treatment response – is critical to understanding the nature of the

interaction between the population of cells and the selective forces shaping it. While advances in genomics and transcriptomics and the advent of single-cell RNA sequencing (scRNA-seq) have vastly increased the resolution at which we can understand cellular processes, they lack the ability to directly assign clonal relationships. To meet this need, lineage tracing technologies, such as DNA barcoding, have been developed to label and track individual cells and their progeny ***(1)(2)***. In DNA barcoding, each individual cell in a population is labeled with a unique random string of nucleotides that is integrated into the genome and heritable by its daughter cells. The ensemble of all DNA barcodes in the cell population can be quantified by next-generation sequencing (NGS) to determine how clonal abundance changes over time.

While highly informative, DNA barcoding and other lineage tracing techniques are still limited in that interesting lineages of cells cannot be easily isolated from the bulk population for clonally pure analysis. Here, we describe a detailed protocol for the Control of Lineages by Barcode Enabled Recombinant Transcription (COLBERT), a workflow that enables precise identification and isolation of populations of interest from heterogeneous mammalian cells ***(3)***. An overview of COLBERT is shown in **Fig. 1**. COLBERT is a functionalized variant of DNA barcoding in which the DNA barcode is a CRISPR-Cas9 compatible single-guide RNA (sgRNA). The sgRNA-barcode has multiple functionalities: (1) It is an integrated DNA barcode, (2) It is transcribed and captured in scRNA-seq workflows, and (3) It can be used to actuate lineage-specific genes of interest using an activator variant of Cas9 ***(4)***. This protocol describes the use of COLBERT for lineage-specific activation of GFP, enabling isolation of clonal cells from a heterogeneous population.

This protocol describes two variants of the COLBERT system, one compatible with single-cell RNA sequencing workflows that use polyA capture and another with specific compatibility with 10X Genomics systems. In the polyA capture version, the sgRNA barcode is engineered using

the CROPseq method ***(5)*** such that the sgRNA barcode is transcribed by both RNA polymerase III and RNA polymerase II, creating a functional sgRNA barcode transcript and a polyadenylated transcript containing the barcode, respectively. In the 10X Genomics version, the sgRNA is engineered to contain a capture sequence that allows targeted capture by the Chromium Single Cell 3' v3 Gel Beads ***(6)***.

Cells are first transduced with lentivirus containing either the CROPseq sgRNA barcoding vector or the 10X Capture sgRNA barcoding vector at a low multiplicity of infection (MOI) to minimize the integration of multiple barcodes per cell. In both versions of the vector, the sgRNA barcode is co-expressed with blue fluorescent protein (BFP) for easy identification and collection of barcoded cells via flow cytometry and fluorescence-activated cell sorting (FACS). Once established, the barcoded cell population is available for experimental manipulation. Clonal dynamics may be measured by NGS analysis and gene expression signatures of clonal populations may be resolved by scRNA-Seq. Once a barcode of interest is identified from NGS or scRNA-seq, the barcode identifier can be exploited for isolation of the clone. This is achieved by transfecting the cell population with a plasmid containing an activator variant of Cas9 (dCas9-VPR) and a second plasmid containing the Cas9-homing PAM sites adjacent to the identified barcode upstream of green fluorescent protein (sfGFP) reporter. Expression of sfGFP will occur only in cells that are producing the matching sgRNA barcode, allowing precise identification and FACS isolation of cells from lineages of interest.

[INSERT FIGURE 1 NEAR HERE]

2. Materials

Equipment

1. Electroporator
2. Mammalian cell incubator
3. Bacterial cell incubator with shaking
4. Thermocycler
5. Gel electrophoresis box
6. Bioanalyzer
7. Illumina sequencer
8. Flow cytometer with filters for BFP (Ex: 380/20, Em: 460/40)

Disposables

1. Sterile filtered pipette tips
2. 1.5 mL microcentrifuge tubes (sterile)
3. 1.8 mL Screw top cryovials (sterile)
4. 20 mL Luer-tapered syringe (sterile)
5. 0.45 µm polyethersulfone (PES) syringe filter
6. 30,000 molecular weight cutoff (MWCO) PES concentrator capable of processing 20 mL

Biologics

1. Electrocompetent *e. coli* suitable for unstable DNA (restriction minus, endonuclease deficient, and recombination deficient)
2. Cells of interest (*see* **Note 1**)

Plasmids

1. CROPseq gRNA expression transfer vector, Cropseq-BFP-WPRE-TS-hU6-BsmbI {Addgene # Pending / Brock Lab AA112}

2. 10X Capture gRNA expression transfer vector, pLKV2-hU6-BbsI-PGK-Puro-TagBFP-WPRE {Addgene # Pending / Brock Lab AA174}
3. Lentiviral packaging plasmid, VSV-G (Addgene #14888)
4. Lentiviral packaging plasmid, psPAX2 (Addgene #12260)
5. dCas9-VPR (Addgene #63798)
6. Recall-miniCMV-sfGFP (Addgene # Pending/ Brock Lab #AA158)

Primers

(*see* **Table 1**)

Buffers

1. Buffer 3.1: 100 mM NaCl , 50 mM Tris-HCl, 10 mM $MgCl_2$, 100 µg/ml BSA , pH 7.9 at 25°C
2. NEB 5X Q5 Reaction Buffer
3. 10X T4 PNK Buffer
4. 10 mM dNTPs
5. 1X Tris-acetate-EDTA (TAE)
6. FACS Buffer: 5% FBS, 1-5 mM EDTA, 95% Phosphate-Buffered Saline

Enzymes

1. BsmBI (10,000 U/mL)
2. BbsI (10,000 U/mL)
3. NEB Q5 polymerase
4. T4 ligase (400,000 U/mL)
5. T7 ligase (3,000,000 U/mL)
6. PNK (10,000 U/mL)

Other Reagents

1. Lipofectamine™ 2000
2. Lipofectamine™ 3000
3. Nuclease-free water
4. Agarose
5. DNA Clean and Concentrator kit
6. 2xYT microbial growth medium
7. Dulbecco's Modified Eagle Medium (DMEM)
8. OptiMEM™ reduced serum medium
9. Fetal Bovine Serum (FBS)
10. Carbenicillin
11. Solid Phase Reversible Immobilization (SPRI) paramagnetic beads for PCR cleanup
12. 70% molecular biology grade ethanol in nuclease-free water
13. 10 mg/mL hexadimethrine bromide
14. 0.05% Trypan blue
15. Plasmid Midi-Prep Kit

Computational

1. Python 3.7
2. Cell Ranger (for 10X analysis)
3. Samtools
4. Cashier (https://github.com/russelldurrett/cashier)

3. Methods

3.1 sgRNA Barcode Library Plasmid Pool Assembly

1. [CROPseq] Perform an extension reaction to generate double-stranded insert gRNA barcode DNA. Mix 10 µL NEB 5X Q5 Reaction Buffer, 1 µL of 10 mM dNTPs, 2 µL 100µM CROPseq-PrimeF-BgL-BsmBI, 1 µL 100µM CROPseq-RevExt-BgL-BsmBI , and 0.5 µL to create a 50 µL reaction (*see* **Note 2**).

   [10X Capture] Perform an extension reaction to generate double-stranded insert gRNA barcode DNA. Mix 10 µL NEB 5X Q5 Reaction Buffer, 1 µL of 10 mM dNTPs, 2 µL 100µM 10X-PrimeF-BgL-BbsI, 1 µL 100µM 10X-RevExt-BgL-BbsI and 0.5 µL to create a 50 µL reaction (*see* **Note 2**).
2. Run the extension reaction on a thermocycler using the following settings: (1) 98 °C for 2 min, (2) 65 °C for 30 sec, (3) 72 °C for 10 sec, (4) Repeat steps 2-3 for 10 cycles, (5) 72 °C for 1 min, (6) 4 °C hold
3. Confirm dsDNA assembly on 2% agarose gel by running single stranded DNA against PCR product.
4. [CROPseq] Digest 5-10 µg of CROPseq vector backbone in a reaction containing 20 µL Digestion Buffer 3.1, 8 µL BsmBI, and nuclease-fee water to 200 µL.

   [10X Capture] Digest 5-10 µg of 10X Capture vector backbone in a reaction containing 20 µL Digestion Buffer 3.1, 8 µL BbsI, and nuclease-fee water to 200 µL.
5. Ligate double stranded gRNA barcode DNA into digested gRNA transfer vector at a molar ratio of 10:1 in a 50X Golden Gate assembly reaction by mixing 1.25 pmol digested gRNA transfer vector (*from step 3.1.4*), 12.5 pmol double stranded gRNA

barcode DNA (*from step 3.1.1*), 50 µL T4 ligase buffer, 25 µL T7 ligase, 25 µL BsmBI (CROPseq) or BbsI (10X Capture), and nuclease-free water to 500 µL (*see* **Note 3**).

6. Run the Golden Gate assembly reaction on a thermocycler using the following settings: (1) 42 °C for 2 min, (2) 16 °C for 5 min, (3) Repeat steps 1-2 for 99 cycles, (4) 55 °C for 30 min, (6) 4 °C hold
7. Clean barcoding library plasmid pool using a DNA clean and concentrator kit and elute in 22 µL warm, nuclease-free water (*see* **Note 4**).
8. Prepare for *e. coli* electroporation by pre-warming recovery media to room temperature, thawing electrocompetent *e. coli* on ice, and pre-chilling 2 mm electroporation cuvettes on ice (*see* **Note 5**).
9. Aliquot 100 µL of E.coli into the chilled 0.2 cm electroporation cuvette, add 5 µL of purified assembled plasmid, and stir with pipet tip (*see* **Note 6**).
10. Transform e. coli by electroporating with 1 pulse at 2.5 kV (*see* **Note 7**).
11. Add 2 mL Recovery Media and gently pipet up and down immediately after electroporation, and transfer to a sterile 50ml conical tube.
12. Repeat steps 9-11 three times
13. Allow cells to recover for 30 min at 37 °C with shaking at 250 rpm.
14. Pre-warm 2xYT agar plates with 100 µg/mL carbenicillin.
15. After recovery, perform dilution plating 1:10^4, 1:10^5, 1:10^6 on carbenicillin agar plates.
16. Incubate plates overnight at 37 °C.
17. Put the remaining transformant mixture into 500 mL 2xYT with 100 µg/mL carbenicillin in a 2 L flasks.
18. Incubate flasks at 30 °C overnight with shaking at 250 rpm.
19. The culture can be pelleted or midi/maxi prepped for usage.
20. Calculate transformation efficiency from dilution plating (*see* **Note 8**).

3.2 sgRNA Barcode Sampling

The diversity of the initial plasmid pool should be assessed to ensure a high diversity library. To do this, PCR is performed with primers containing Illumina indices that anneal to regions flanking the barcodes.

1. Midi-prep one tube of transformed e. coli from step 3.1.1 according to manufacturer's instructions.
2. Create the phasing primer mixture 'PrimeF PAS-hU6_F' by mixing equimolar amounts of UniF PASx0 1 stage, UniF PASx1 1 stage, UniF PASx2 1 stage, UniF PASx3 1 stage, UniF PASx4 1 stage, UniF PASx6 1 stage (*see* **Table 1**)(*see* **Note 9**).
3. [CROPseq] Prepare the PCR reaction to amplify barcodes and add Illumina indices (*see* **Note 10**) by mixing 10 uL 5X Q5 Reaction Buffer, 1 µL 10 mM dNTPs, 2.5 µL PrimeF PAS-hU6_F (from step 3.2.2), 2.5 µL PrimeR_cropseq (*see* **Table 1**), 0.5 µL Q5 polymerase, 50 ng plasmid DNA, and nuclease-free water to 50 µL.
   [10X Capture] Prepare the PCR reaction to amplify barcodes and add Illumina indices (*see* **Note 10**) by mixing 10 uL 5X Q5 Reaction Buffer, 1 µL 10 mM dNTPs, 2.5 µL PrimeF PAS-hU6_F (from step 3.2.2), 2.5 µL PrimeR_10x (*see* **Table 1**), 0.5 µL Q5 polymerase, 50 ng plasmid DNA, and nuclease-free water to 50 µL.
4. Amplify the barcodes by running the 50 µL reaction on a thermocycler using the following settings (*see* **Note 11**): (1) 98 °C for 2 min, (2) 98 °C for 10 sec, (3) 63 °C for 30 sec, (4) 72 °C for 15 sec, (4) Repeat steps 2-4 for 7 cycles (*see* **Note 12**), (5) 72 °C for 2 min, (6) 4 °C hold
5. Transfer 50 µL PCR amplification product to a nuclease-free microcentrifuge tube
6. Allow SPRI beads to come to room temperature.

7. Add 80 µL (1.6X) paramagnetic SPRI beads and mix well with vortexing or pipetting up and down 10 times.
8. Incubate at room temperature for 5 minutes.
9. Place the tube on a magnetic rack and allow solution to clear (5-10 minutes).
10. Remove supernatant.
11. Remove tube from magnetic rack and wash beads with 200 µL of freshly prepared 80% ethanol.
12. Return the tube to magnetic rack and allow the solution to clear (5-10 minutes).
13. Remove supernatant, repeat wash step, and remove supernatant.
14. Allow the pelleted beads to dry until they have obtained a matte finish and minimal cracking in the bead pellet (*see* **Note 13**).
15. Remove tube from the magnetic rack and elute DNA by adding 42 µL of nuclease-free water.
16. Incubate at room temperature for 10 minutes.
17. Transfer tube to magnetic rack and collect 40 µL of purified PCR product after solution has cleared (5-10 minutes) (*see* **Note 14**).
18. Quantify DNA yield with a high sensitivity fluorometry kit ensuring yield between 1-10 ng/µL.
19. Load sample on to BioAnalyzer chip according to the manufacturer's protocol and ensure a clear peak around 225 bp (*see* **Note 15**).
20. Submit sample for Illumina sequencing.
21. See 3.7.2 for processing barcode sequence data.

3.3 sgRNA Barcoding Lentivirus Production

1. 48 hours before transfection, plate 0.22-0.25 x 10^6 low-passage HEK-293T cells in DMEM supplemented with 10% FBS without antibiotics in each well of a sterile 6-well tissue culture treated plate such that cells will be 70-80% confluent at the time of transfection.
2. On the morning of transfection, replace media on HEK-293T cells with 2 mL of fresh Opti-MEM™ (or your cells growth medium) supplemented with 10% FBS without antibiotics.
3. In the afternoon, warm Opti-MEM™, Lipofectamine™ 2000, and VSV-G, psPAX, and sgRNA barcoding plasmid to room temperature (*see* **Note 16**, **Note 17**).
4. Per well of a 6 well plate, prepare "Tube A" containing 150 µL Opti-MEM™ and 9 µL Lipofectamine™ 2000 (*see* **Note 18**).
5. Incubate "Tube A" at room temperature for 5 minutes.
6. Per well of a 6 well plate, prepare "Tube B" containing 150 µL Opti-MEM™,1.5 µg psPax, 0.4 µg VSV-G, 3-5 µg sgRNA barcoding plasmid.
7. Slowly add "Tube B" dropwise to "Tube A" and carefully mix by gently inverting 10 times
8. Incubate at room temperature for 20 minutes.
9. Add 300 µL of the transfection mix slowly and dropwise to each well of HEK-293T cells.
10. 16-18 hours post-transfection, carefully remove media containing Lipofectamine™ 2000 complexes and slowly replenish with DMEM supplemented with 20% FBS without antibiotics (*see* **Note 19**, **Note 20**).
11. 48 hours post-transfection, harvest viral containing supernatant and store in a 50 mL conical tube at 4 °C. Replenish media dropwise on virus producing cells (*see* **Note 21**)
12. 72 hours post-transfection, harvest viral containing supernatant and combine with 48 hour viral collection (*see* **Note 22**, **Note 23**).

13. Spin down collected viral containing supernatant at 500 x g for 10 min at 4 °C to remove residual HEK-293T cells.
14. Remove plunger from 20 mL syringe and attach to a 0.45 µm PES syringe filter.
15. Transfer viral supernatant to a 20 mL syringe.
16. Filter viral supernatant through 0.45 µm PES syringe filter into a fresh 50 mL conical tube to remove any remaining cell debris.
17. Concentrate virus 20X in 30,000 MWCO PES ultrafiltration centrifugal concentrator by loading 20 mL of filtered viral supernatant into concentrator chamber and spinning at 4000 x g for 60-75 minutes at 4 °C until ~1 mL of media remains in filter (*see* **Note 24**).
18. Aliquot 25-50 µL of concentrated virus in threaded cryovials and store at -80 °C (*see* **Note 25**, **Note 26**).

3.4 Determine sgRNA Viral Titer

(*see* **Note 27**, **Note 28**)

3.4.1 Titering on Adherent Cells (Forward Procedure)

(*see* **Note 29**)

1. 24-48 hours before performing viral transduction seed your cell line of interest in a 12-well plate such that it is near 60-70% confluent at time of transduction.
2. Prior to transduction, one well of the replicate 12 wells should be dissociated and counted using trypan blue exclusion on a hemocytometer to know approximate number of live cells at time of transduction (*see* **Note 30**, **Note 31**).
3. Create stock of media containing your cells' standard growth medium supplemented with 20% FBS containing 0-10 µg/mL hexadimethrine bromide (1:1000 dilution from hexadimethrine bromide stock for 10 µg/mL) (*see* **Note 32**).

4. Place 600 µL of hexadimethrine bromide containing medium separate microcentrifuge tubes.
5. Add virus in increasing amounts to each tube (*see* **Note 33**).
6. Replace media on cells of interest with virus and hexadimethrine bromide containing dilutions.
7. Incubate for 16 hrs at 37 °C, then carefully remove viral containing supernatant and replace with complete growth medium. (*see* **Note 34**, **Note 35**).
8. Incubate for an additional 32 hrs at 37 °C, then remove medium and wash each well gently with PBS (*see* **Note 36**).
9. Dissociate the cells from the plate and centrifuge at 300 x g for 5 minutes at 4 °C.
10. Wash cell pellets with PBS and repeat spin.
    a. Perform this step three times to ensure removal of trace virus before flow cytometry.
11. Resuspend cells in chilled FACS Buffer (*see* **Note 37**).
12. Keep cells on ice and continue to step 3.4.3

3.4.2 Titering on Suspension Cells

1. Count your cells of interest using a hemocytometer.
2. Create stock of media containing your cells' standard growth medium supplemented with 20% FBS containing 0-10 µg/mL hexadimethrine bromide (1:1000 dilution from hexadimethrine bromide stock for 10 µg/mL) (*see* **Note 32**).
3. Resuspend 1.20 x 10^6 cells in 7.2 mL of containing hexadimethrine bromide media such that the final solution contains 1 x 10^5 cells in 600 µL.
4. Plate 600 µL of cell solution in 10 wells of a tissue culture treated 12-well plate
5. Add virus in increasing amounts to each well and mix well (*see* **Note 33**).
6. Incubate for 16 hrs at 37 °C (*see* **Note 34**, **Note 35**).

7. Transfer cell suspensions to sterile 1.7 mL microcentrifuge tubes and spin down at 500 x g for 5 minutes at 4 °C (*see* **Note 38**).
8. Resuspend each cell pellet with complete growth medium and transfer to fresh 12-well plate.
9. Incubate for an additional 32 hrs at 37 °C, then transfer wells to microcentrifuge tubes and spin down at 400 x g for 5 minutes at 4 °C (*see* **Note 39**).
10. Wash cell pellets with PBS and repeat spin (*see* **Note 40**).
11. Resuspend cells in chilled FACS Buffer (*see* **Note 37**).
12. Keep cells on ice and continue to step 3.4.3

3.4.3 Flow Cytometry to Determine Viral Titer

1. Pass cells resuspended in FACS buffer through a 35 µm nylon mesh strainer into a 5 mL flow cytometry test tube (*see* **Note 41**).
2. Use control samples to set laser voltages on FSC-A, SSC-A, and BFP such that nearly all cells are seen within FSC-A vs. SSC-A plot and both negative and positive populations can be seen and distinguished on the BFP channel (*see* **Note 42**).
3. After setting voltages with control samples, run transduced samples from lowest viral to highest. Set the cytometer to record at least 10,000 events for each sample. Record %BFP-positive for each titration.
4. Create a plot showing volume of virus on the x-axis and %BFP-positive on the y-axis (*see* **Note 43**).

   Calculate viral titer in titering units (TU) per mL using **Equation 1** using a pair of values within the linear region of the titer curve (*see* **Note 44**).

$$\frac{\text{TU}}{\text{mL}} = \frac{(\text{Number of cells at time of transduction}) \times (\text{Fraction of Positive Cells})}{(\text{Volume of virus [mL]})} \quad \textbf{(Eq. 1)}$$

3.5 sgRNA Barcode Transduction

1. After calculating the viral titer (TU/mL) on your cell line of interest, determine the final number of cells you require for your experiment using and transduce cells at a multiplicity of infection (MOI) of 0.1 (**Equation 2**) to minimize the occurrence of multiple barcode integrations (*see* **Note 45**, **Note 46**).
2. Use control samples to set laser voltages on FSC-A, SSC-A, and BFP such that nearly all cells are seen within FSC-A vs. SSC-A plot and both negative and positive .populations can be seen and distinguished on the BFP channel (*see* **Note 41**).
3. Set sort gate on BFP-positive cells indicative of a productive sgRNA barcode (*see* **Note 47**).
4. Sort cells on BFP-positive gate via FACS.
5. Maintain sorted cells in culture with complete growth medium.

$$\text{MOI [TU/cell]} = \frac{(\text{Volume of Virus needed [mL]}) \times (\text{Titer of Virus [TU/mL]})}{(\text{Number of cells exposed to virus})} = \textbf{(Eq. 2)}$$

3.7 Targeted sgRNA Barcode Sampling of Cells

3.7.1 Preparing Samples for Sequencing

1. To assess cell barcode diversity harvest cells from culture and collect into cell pellet. (See **Note 48**).
2. Isolate genomic DNA from cell pellet using kit or standard protocol and proceed to PCR amplification.

3. [CROPseq] Prepare the PCR reaction to amplify barcodes and add Illumina indices (*see* **Note 10**) by mixing 10 uL 5X Q5 Reaction Buffer, 1 µL 10 mM dNTPs, 2.5 µL PrimeF PAS-hU6_F (from step 3.2.2), 2.5 µL PrimeR_cropseq (*see* **Table 1**), 0.5 µL Q5 polymerase, up to 2 µg genomic DNA (*see* **Note 49**), and nuclease-free water to 50 µL. [10X Capture] Prepare the PCR reaction to amplify barcodes and add Illumina indices (*see* **Note 10**) by mixing 10 uL 5X Q5 Reaction Buffer, 1 µL 10 mM dNTPs, 2.5 µL PrimeF PAS-hU6_F (from step 3.2.2), 2.5 µL PrimeR_10x (*see* **Table 1**), 0.5 µL Q5 polymerase, up to 2 µg genomic DNA (*see* **Note 49**), and nuclease-free water to 50 µL.
4. Amplify the barcodes by running the 50 µL reaction on a thermocycler using the following settings (*see* **Note 11**): (1) 98 °C for 2 min, (2) 98 °C for 10 sec, (3) 63 °C for 30 sec, (4) 72 °C for 15 sec, (4) Repeat steps 2-4 for 20 cycles, (5) 72 °C for 2 min, (6) 4 °C hold
5. Finish preparing barcodes as described in 3.2 steps 4-20.

3.7.2 Processing Barcode Sequencing Data

1. Processing barcode data is accomplished through several bash scripts which can be found at https://github.com/russelldurrett/cashier (*see* **Note 50**).
2. Concatenate all fastq files for a given sample into a single fastq file using the following command: `cat sampe1*.fastq > sample1.fastq` (*see* **Note 51**).
3. Extract the barcodes from each illumina read with a minimum Phred quality of 30 using the following command: `cashier_extract -i sample1.fastq -q 30` (*see* **Note 52**).
4. This will generate multiple output files. Extracted barcodes can be found in the tsv file with the naming format *.barcodes.q30.tsv.

5. Remove any barcode sequences that do not occur at least twice using the following command: *cat sample1.barcodes.q30.tsv | cut -f 2 | sort | uniq -D > sample1.barcodes.q30.raw2.tsv*
6. To correct for sequencing and PCR errors within the sequence we implement message-passing clustering by wrapping starcode ***(7)*** with the *cluster_columns* command: *cluster_columns -i sample1.barcodes.q30.raw2.tsv -r 3 -d 1 -c 1.* (*see* **Note 53**).
7. This will create a file containing three columns the original sequence, the centroid, and the size of the centroid. Since we are only interested in centroids and the number of occurrences we remove the original sequences with the following command: *sample1.barcodes.q30.raw2.c1d1r3.tsv | cut -f 2,3 -d ' ' | sort -k 1,1 | uniq > sample1.barcodes.q30.raw2.c1d1r3.uniq.tsv*
8. We recommend eliminating the lowest abundant and rare sequences that are not likely to be present in the population. This can be accomplished with the *cat* and *awk* command to, for instance remove sequences occurring less than 20 times: *cat sample1.barcodes.q30.raw2.c1d1r3.uniq.tsv | awk '$2>19 {print;}' > sample1.barcodes.q30.raw2.c1d1r3.uniq.min20.tsv*

3.7.3 Processing CROP-seq Barcodes from 10X Cell Ranger Output

1. To obtain barcode data from cells with 10X Capture sgRNA barcodes (*see* **Note 54**) run 10X Cell Ranger on fastq samples obtained from Illumina sequencing.
2. To process barcodes from 10X data, run *samtools* to convert the unmapped read bam file into a sam file. You will first take the unmapped reads bam file and convert to a sam file using: *samtools view possorted.bam > potsorted.sam*
3. Next pipe the bead- and umi-tagged reads through cutadapt to identify and trim barcodes, then translate to a tsv: *python*

$cashier/scripts/sam_to_name_labeled_fastq.py possorted.sam | cutadapt -g CTTGTGGAAAGGACGAAACACCG -a GTTTTAGAGCTAGAA -n 2 - | python $cashier/scripts/fastq_tagged_to_tsv.py - > readname_umi_cellbarcode_lineagebarcode.tsv

3.8 Recall Plasmid Assembly

1. 3 pairs of overlapping oligos containing the barcode sequence of interest flanked by overlapping sequences should be ordered according to **Table 1** (*see* **Note 55**).
2. In separate tubes, mix each of the 100 µM oligo pairs together

   “Tube AC”: 10 µL Barcode_PAM A-C extraction + 10 µL Barcode_PAM A-C reversed

   “Tube BE”: 10 µL Barcode_PAM B-E extraction + 10 µL Barcode_PAM B-E reversed

   “Tube EF”: 10 µL Barcode_PAM E-F extraction + 10 µL Barcode_PAM E-F reversed
3. Heat each to 80 °C and let cool to create DNA blocks containing a barcode, a PAM site, and overhang sequences. (*see* **Note 56**).
4. Ligate DNA blocks together creating the barcode array by mixing 10 µL “Tube AC”, 10 µL “Tube BE”, 10 µL “Tube EF”, 5 µL 10 mM dNTPs, 5 µL 10X T4 PNK buffer, 1 µL T4 PNK, and 9 µL nuclease-free water (50 µL reaction volume).
5. Incubate at 37 °C for 45 minutes.
6. Add 2 µL T7 DNA ligase to the 50 µL mixture and incubate at room temperature overnight.
7. Run ligation product in a 2% agarose gel and gel purify band from approximately 200 bp
8. Ligate the barcode array into the recall plasmid backbone at a molar ratio of 10:1 in a Golden Gate assembly reaction by mixing 25 fmol Recall-miniCMV-sfGFP, 250 fmol assembled barcode array (*from step 3.8.5*), 1 µL T4 ligase buffer, 0.5 µL T7 ligase, 0.5 µL BbsI, and nuclease-free water to 10 µL.

9. Run the Golden Gate assembly reaction on a thermocycler using the following settings: (1) 42 °C for 2 min, (2) 16 °C for 5 min, (3) Repeat steps 1-2 for 35 cycles, (4) 55 °C for 30 min, (6) 4 °C hold
10. Verify insertion of barcode array into Recall-miniCMV-sfGFP backbone via Sanger sequencing.

3.9 Recall and Isolation of Barcoded Lineages

(*see* **Note 57**)

1. 24-48 hours before performing recall transfection seed your cell line of interest in growth medium in a 6-well plate such that it is near 60-80% confluent at time of transfection.
2. Per well of a 6 well plate, prepare “Tube A” containing 100 µL Opti-MEM™ and 9 µL Lipofectamine™ 3000 (*see* **Note 18**).
3. Incubate “Tube A” at room temperature for 5 minutes.
4. Per well of a 6 well plate, prepare “Tube B” containing 125 µL Opti-MEM™, 225 ng Recall plasmid (*from Section 3.8*), and 275 ng dCas9-VPR plasmid.
5. Slowly add “Tube B” dropwise to “Tube A” and carefully mix by gently inverting 10 times
6. Incubate at room temperature for 20 minutes.
7. Add 225 µL of the transfection mix slowly and dropwise to each well of adherent cells.
8. 16-18 hours post-transfection, carefully remove media containing Lipofectamine™ 3000/DNA complexes and slowly replenish with growth medium supplemented with 20% FBS without antibiotics.
9. 48-72 hours post-transfection, dissociate cells from the plate and wash cells with PBS twice at 300 x g for 5 minutes at 4 °C before resuspending in chilled FACS buffer (*see* **Note 37**).

5. Pass cells resuspended in FACS buffer through a 35 μm nylon mesh strainer into a 5 mL flow cytometry test tube and keep on ice.
6. Use control samples to set laser voltages on FSC-A, SSC-A, BFP, and GFP on FACS sorter such that nearly all cells are seen within FSC-A vs. SSC-A plot and both negative and positive populations can be seen and distinguished on the BFP and the GFP channel. Set compensations based on single positive populations (*see* **Note 58**).
10. Set sort gate on GFP and BFP double positive gate indicative of a recalled cell (*see* **Note 59**).
11. Sort cells in GFP and BFP double positive gate (*see* **Note 60**).
12. Maintain sorted cells in culture with complete growth medium.

4. Notes

1. Make sure cells are transducible with lentivirus. Timing of lentiviral exposure and detectable expression of transgene will vary across cell types.
2. Always use filtered pipette tips when working with DNA to prevent cross-contamination.
3. A 1X Golden Gate assembly reaction is setup by mixing 25 fmol digested gRNA transfer vector (*from step 3.1.4*), 250 fmol double stranded gRNA barcode DNA (*from step 3.1.1*), 1 μL T4 ligase buffer, 0.5 μL T7 ligase, 0.5 μL BsmBI (CROPseq) or BbsI (10X Capture), and nuclease-free water to 10 μL.
4. Letting the water sit on the column for 3-5 minutes before elution increases yield. Re-run elution product through column 3 times to maximize yield.
5. Make sure to use *E. coli* suitable for use with unstable DNA.

6. Do not pipet up and down. Ensure bubbles are not added to the mix which can cause electrical arcing and cell death during electroporation).
7. Optimal time constants should be between 4.2-5.4 ms. This protocol was optimized with the EC2 setting on the Bio-Rad MicroPulser™ Electroporator.
8. Transformation efficiency (TE) is defined as the number of colonies produced with transformation with 1 µg of plasmid DNA. To calculate TE, count the number of colonies formed on the plate, calculate the amount of DNA used in µg, and determine your dilution factor. With those variable, TE = Colonies/µg/Dilution.
9. Universal phase amplicon sequencing primers are used to add more diversity to the sequencing reads which helps prevents sequencing errors.
10. Choose Illumina i7 index NNNNNN based on Illumina sequencer to be used and avoid using the same i7 index for different samples.
11. Pre-heat thermocycler to 98 °C before adding tubes to heat block.
12. The number of cycles will depend on the starting template amount (7-23 cycles). A nested PCR reaction may have to be performed to enhance barcode specificity.
13. Do not over dry the beads, this can result in a loss of yield and quality.
14. Beads may become trapped within the meniscus of the water. Pipetting slowly will keep the beads against the wall of the tube and leave them in the remaining 2 µL of water.
15. If there are considerable peaks at 120 bp or less, SPRI bead cleanup can be repeated with 1.1X beads to further purify PCR sample, but this will greatly reduce yield.
16. Lentivirus can promiscuously infect cells, including your skin! Use a cuffed-sleeve lab coat and double-glove (one glove under sleeve cuffs, one glove over) at every step involving use of virus.
17. Ethanol does not kill lentivirus. Always keep a working stock of 100% bleach in the BSL-2 culture hood in which virus is being handled. Soak pipet tips, serological pipets, and other disposables that come in contact with virus in 100% bleach and irradiate with UV

for at least 30 minutes before disposal as biohazardous waste. Wipe down virus containing tissue culture plates with disinfecting wipes certified to kill HIV such as CaviCide before removing from culture hood.

18. Slowly dilute Lipofectamine™ 2000 complexes dropwise with Opti-MEM™ media with occasional flicking of the tube.
19. You are working with live virus at this stage and beyond. Stringently adhere to all biosafety procedures.
20. Cells exposed to lentivirus are fragile and extra care must be taken in removing and adding media.
21. Virus should be stored in labeled secondary containment.
22. Virus-producing HEK-293T cells should be bleached and UV irradiated in culture for at least 30 minutes to inactivate remaining virus before disposal.
23. Never use a vacuum line to disposal of virus waste as this may produce aerosols.
24. Spin times will vary based on centrifuge angle. spinning at 4 °C will increase the amount of time it takes for media to pass through filter (We have noted that 22 mL takes about 75 minutes).
25. Even just a single freeze-thaw cycle can drastically alter viral titer, be sure to minimize freeze-thaw cycles.
26. Virus should be completed frozen and then thawed before calculating viral titer.
27. Viral titer will vary between cell type and with each new virus preparation.
28. Lentivirus susceptibility and timing should first be determined on your cells of interest using a control plasmid such as a constitutively active GFP. Some cells will require longer or shorter incubation times with the virus and some cells will take longer to produce the transgenic reporter protein.
29. To perform reverse titer on adherent cells, follow the steps for titering on suspension cells through step 3.4.2.5, then return to the adherent protocol at step 3.4.1.6.

30. It is very important to know the number of cells at the time of transduction. This number is used to calculate viral titer.
31. Trypan blue exclusion is performed by mixing equal parts 0.05% Trypan blue with your cell suspension, usually 10 µL of each, then load 10 µL of the stained suspension into the hemocytometer.
32. Hexadimethrine bromide is a cationic solution that assists in viral adsorption to cells ***(8)***. Hexadimethrine bromide can be toxic to some cells. Hexadimethrine bromide sensitivity should be assessed via serial dilution to determine maximum tolerable hexadimethrine bromide dose before determining viral titer. Most cells respond well to 6-8 µg/mL hexadimethrine bromide.
33. Ensure one well is kept uninfected as a negative control. A range of 0.5-200 µL is usually sufficient to find viral titer, e.g. 0, 0.5, 1, 5, 10, 25, 50, 100, 150, 200 µL.
34. Lentiviral exposure time will vary across cell type dependent on growth dynamics and properties intrinsic to the cells. Optimize lentiviral exposure time with constitutively active GFP virus before transduction with sgRNA barcoding library virus. The protocol as described was optimized for MDA-MB-231 (ATCC HTB-26) breast cancer cells.
35. Lentiviral exposure times range between 12-48 hours. Lentiviral exposure time should be minimized to reduce the occurrence of multiple viral integrations.
36. Lentivirus transduced cells are very fragile and should be handled with added care.
37. EDTA and FBS in FACS buffer help to prevent cell clumping. For extra-sticky cells, use 5 mM EDTA in FACS buffer.
38. Use a pipette to remove lentivirus containing supernatant and dispose of in bleach. Do not vacuum aspirate -- vacuums can cause dangerous viral aerosols.
39. Lentivirus transduced cells are very fragile and should be handled with added care when pipetting.
40. Perform this step three times to ensure removal of trace virus before flow cytometry.

41. Ensure proper controls for flow. Minimally have a positive control expressing BFP and a negative control expressing no fluorescent proteins.
42. BFP populations will be normally distributed. For titer calculations, it is useful to set tight gates such that 99.98% of the negative control cells are captured in the negative gate.
43. Plot will appear logarithmic. Only values within the linear region of the plot should be used to calculate viral titer (usually between 10-40% BFP-positive).
44. Example: If 5 µL of virus added to 100,000 cells resulted in 30% BFP-positive cells within the linear region of the titer curve, then the viral titer would be (100,000 x 0.30) / (0.005 mL) = 6.0×10^6 TU/mL
45. Example: If your viral titer was 6.0×10^6 TU/mL and you wanted to infect 3.0×10^6 cells at an MOI of 0.1, you would need to subject the 3.0×10^6 cells to 50 µL of virus.
46. A low MOI of 0.1 or below helps prevent occurrence of multiple barcode integrations. In order to uniquely recall cell lineages it is important to maximize the probability that there is one or zero barcodes per cell at the time of transduction. The probability of barcode integration can be modeled as a Poisson distribution ***(9)(10)***.
47. When sorting for sgRNA barcoded cells, use more stringent gating than used for titer determination. Ensure that 0% of negative control samples appear in the sorting gate.
48. It is important to ensure that you have enough cells to sufficiently sample your population depending upon the initial barcode diversity.
49. DNA amount used will be dependent on the nature of the cell population and desired sampling depth. To capture rare events, a maximum of 2 µg of DNA per reaction can be used and multiple reactions can be done. Given that a single diploid human genome is estimated at ~6.6 pg, 2 µg of genomic DNA represents that of ~300,000 cells. To capture only highly represented clonal populations, less DNA can be used.
50. We assume basic familiarity working with bash in a UNIX based environment

51. If your amplicons were analyzed on a paired end run you should first merge the paired reads from each lane before proceeding.
52. When processing barcodes from 10X vector it is necessary to specify the downstream adapter with *-a GTTTAAGAGCTAAGCTGG.*
53. This command requires an input file, a clustering ratio, Levenshtein distance and column to cluster on.
54. For the 10X compatible sgRNA processing refer to the 10X genomics documentation on their feature barcoding technology.
55. The barcode sequence should be ordered to match the extracted barcode for the fragments labeled as ‘extraction’ and in reverse-complement for oligos labeled as ‘reversed’.
56. This process anneals the single stranded DNA oligos together, creating short double stranded DNA blocks that will be ligated together in the next step.
57. This protocol is optimized for adherent cell lines. If using suspension lines, electroporation can be done to introduce the plasmids to your cells. Be sure to optimize electroporation parameters on your cells for maximized plasmid expression and minimized cell death before recall electroporation. If electroporating, total plasmid load per cell may vary by cell type. Example: CD8 T cells respond well to 2.5 µg of each plasmid (5 µg total DNA load) per 5 x 10^5 cells.
58. Ensure proper controls for flow. Minimally have a positive control singularly positive for BFP, a positive control singularly positive for GFP, and a negative control expressing no fluorescent proteins.
59. When sorting for recalled cells, use stringent gating. Ensure that 0% of negative control and single positive samples appear in the sorting gate.
60. Single cell sorting can be performed for isolation and growth of clonal populations.

[Insert Table 1 near here]

5 References

1. Blundell, J.R., and Levy, S.F. (2014). Beyond genome sequencing: Lineage tracking with barcodes to study the dynamics of evolution, infection, and cancer. Genomics **104**, 417–430.
2. Kebschull, J.M., and Zador, A.M. (2018). Cellular barcoding: lineage tracing, screening and beyond. Nature Methods **15**, 871–879.
3. Al'Khafaji, A.M., Deatherage, D., and Brock, A. (2018). Control of Lineage-Specific Gene Expression by Functionalized gRNA Barcodes. ACS Synth Biol **7**, 2468–2474.
4. Chavez, A., Scheiman, J., Vora, S., et al. (2015). Highly efficient Cas9-mediated transcriptional programming. Nature Methods **12**, 326–328.
5. Datlinger, P., Rendeiro, A.F., Schmidl, C., et al. (2017). Pooled CRISPR screening with single-cell transcriptome readout. Nat. Methods **14**, 297–301.
6. 10xGenomics. Guide RNA Specifications Compatible with Feature Barcoding technology for CRISPR screening. 10x Genomics. Technical Note. CG000197 RevA.
7. Zorita, E., Cuscó, P., and Filion, G.J. (2015). Starcode: sequence clustering based on all-pairs search. Bioinformatics **31**, 1913–1919.
8. Davis, H.E., Morgan, J.R., and Yarmush, M.L. (2002). Polybrene increases retrovirus gene transfer efficiency by enhancing receptor-independent virus adsorption on target cell membranes. Biophys. Chem. **97**, 159–172.
9. Fehse, B., Kustikova, O.S., Bubenheim, M., et al. (2004). Pois(s)on--it's a question of dose.. Gene Ther. **11**, 879–881.
10. Kustikova, O.S., Wahlers, A., Kuhlcke, K., et al. (2003). Dose finding with retroviral vectors: correlation of retroviral vector copy numbers in single cells with gene transfer efficiency in a cell population. Blood **102**, 3934–3937.

Figure Captions

Figure 1. Overview of COLBERT workflow

Table Captions

Table 1. Oligonucleotides

Step	Name	Sequence (5' -> 3')	Notes
3.1	CROPseq-PrimeF-BgL-BsmBI	GAGCCTCGTCTCCCACCGNNNNNNNNNNNNNNNNNNNGTTTTGAGACGCATGCTGCA	The N_{20} sequence is a random string of oligonucleotides
3.1	CROPseq-RevExt-BgL-BsmBI	TGCAGCATGCGTCTCAAAAC	
3.1	10X PrimeF-BgL-BbsI	GCCTGAAGACCTCACCGNNNNNNNNNNNNNNNNNNNGTTTTAGTCTTCCATGCTGC	The N_{20} sequence is a random string of oligonucleotides
3.1	10X-RevExt-BgL-BbsI	TGCAGCATGGAAGACTAAAAC	
3.2, 3.7	PrimeR_cropseq	CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTC	The N_6 sequence is where the i7 Illumina index should be placed
3.2, 3.7	PrimeR_10x	CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGGACCGGCCTTAAAGC	The N_6 sequence is where the i7 Illumina index should be placed
3.2, 3.7	PrimeF PAS-hU6_F	equimolar mixture of UniF PAS primers	
3.2, 3.7	UniF PASx0 1 stage	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGCTTTATATATCTTGTGGAAAGGACGAAAC	
3.2, 3.7	UniF PASx1 1 stage	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCGCTTTATATATCTTGTGGAAAGGACGAAAC	
3.2, 3.7	UniF PASx2 1 stage	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGCGCTTTATATATCTTGTGGAAAGGACGAAAC	
3.2, 3.7	UniF PASx3 1 stage	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTAGCGCTTTATATATCTTGTGGAAAGGACGAAAC	

3.2, 3.7	UniF PASx4 1 stage	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCAACGCTTTATATATCTTGTGGAAAGGACGAAAC	
3.2, 3.7	UniF PASx6 1 stage	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTGCACCGCTTTATATATCTTGTGGAAAGGACGAAAC	
3.8	Barcode_PAM A-C - Fragment 2 (reversed)	CATTGCACCTTAGCTGAGCCTNNNNNNNNNNNNNNNNNNNNGACGAGTGATACCAC	Insert reverse-complement barcode in place of N_{20}
3.8	Barcode_PAM A-C - Fragment 2 extraction	GGAGGTGGTATCACTCGTCNNNNNNNNNNNNNNNNNNNNAGGCTCAGCTAAGGTGC	Insert barcode in place of N_{20}
3.8	Barcode_PAM B-E - Fragment 2 (reversed)	AAGCAGAGAACTAATCCACCTNNNNNNNNNNNNNNNNNNNNTGCGGTTCTTGGTCG	Insert reverse-complement barcode in place of N_{20}
3.8	Barcode_PAM B-E - Fragment 2 extraction	TACTCGACCAAGAACCGCANNNNNNNNNNNNNNNNNNNNAGGTGGATTAGTTCTCT	Insert barcode in place of N_{20}
3.8	Barcode_PAM E-F - Fragment 2 (reversed)	AGCGCAGCTGGATTACAGCCTNNNNNNNNNNNNNNNNNNNNGGGTAACCGCAGGAC	Insert reverse-complement barcode in place of N_{20}
3.8	Barcode_PAM E-F - Fragment 2 extraction	GCTTGTCCTGCGGTTACCCNNNNNNNNNNNNNNNNNNNNAGGCTGTAATCCAGCTG	Insert barcode in place of N_{20}
3.8	F-A Shunt by BbsI - Fragment 2 (reversed)	CTCCAGCTAATGACCTTCCATAGC	
3.8	F-A Shunt by BbsI - Fragment 2 extraction	CGCTGCTATGGAAGGTCATTAGCT	

Acknowledgements

This work has been supported by funding through the NIH (R21CA212928 to AB).