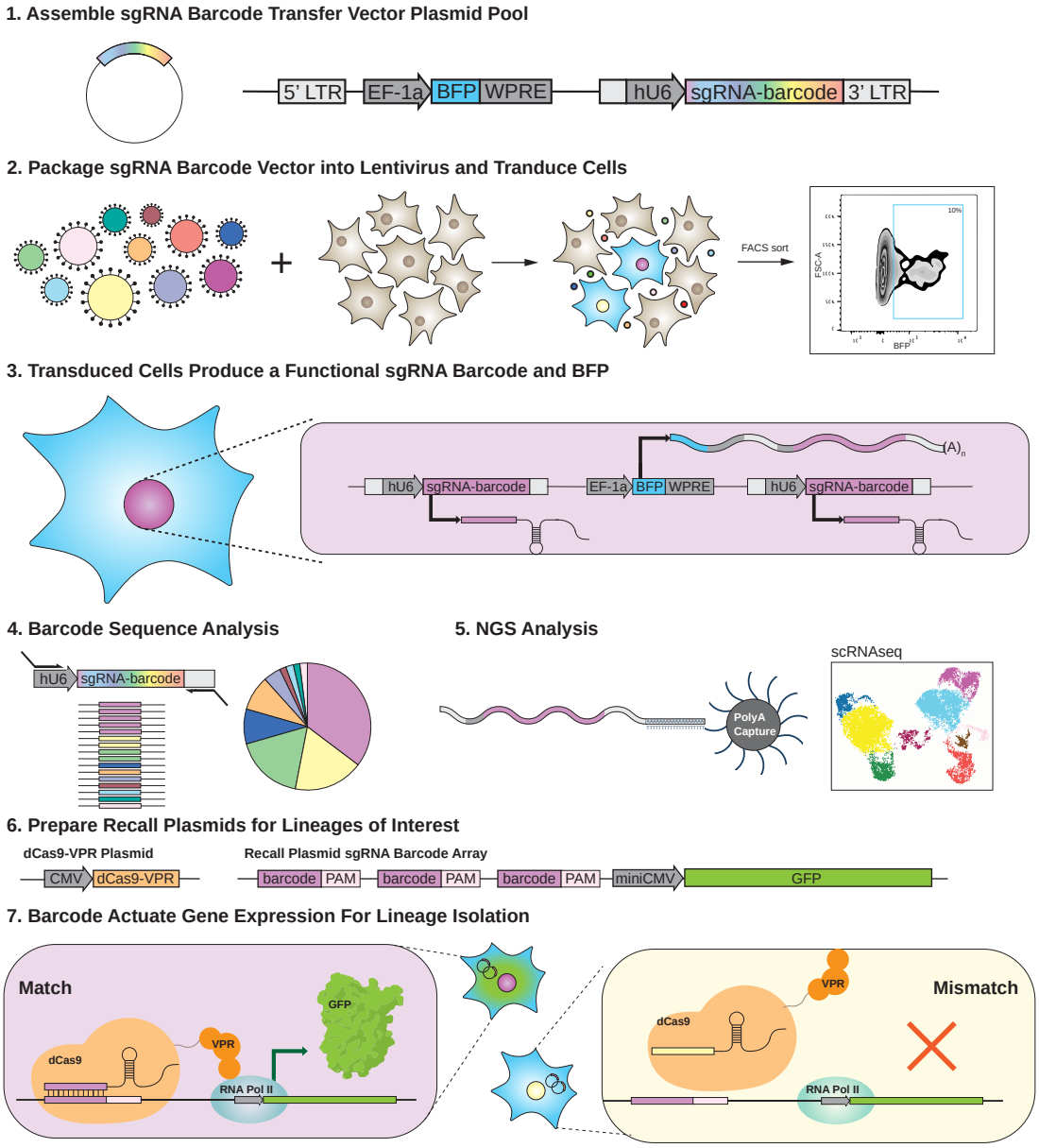
ClonMapper Barcoding System

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

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 ClonMapper 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.

ClonMapper Workflow



Contents

Introduction

Insight into the clonal composition of a population of cells during key events such as development, infection, tumor progression, or treatment response, is critical to understanding the nature of the interaction between cells and the selective forces shaping them. 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 (Blundell and Levy 2014; Kebschull and Zador 2018). 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/clones of cells cannot be easily isolated from the bulk population for clonally pure analysis. Here, we describe a detailed protocol for **ClonMapper**, a workflow that enables precise identification and isolation of populations of interest from heterogeneous mammalian cells (Al'Khafaji, Deatherage, and Brock 2018). ClonMapper 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 (Chavez et al. 2015). This protocol describes the use of ClonMapper for lineage-specific activation of Green Fluorescent Protein, enabling isolation of clonal cells from a heterogeneous population.

The sgRNA barcode is engineered using the CROPseq method (Datlinger et al. 2017) 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.

Cells are first transduced with lentivirus containing a ClonMapper sgRNA barcode vector at a low multiplicity of infection (MOI) to minimize the integration of multiple barcodes per cell. 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 a super-folding 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.

This protocol was originally developed by Aziz Al'Khafaji in the Brock Lab at the University of Texas at Austin. It was written and published in Methods in Molecular Biology by Andrea Gardner and Daylin Morgan. This version has been updated with the Brock Lab's current best practices.

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 [1]

Plasmids

1. CROPseq gRNA expression transfer vector, Cropseq-BFP-WPRE-TS-hU6-BsmbI (Addgene #137993;Brock Lab AA112)
2. Lentiviral packaging plasmid, VSV-G (Addgene #14888)
3. Lentiviral packaging plasmid, psPAX2 (Addgene #12260)
4. dCas9-VPR (Addgene #63798)
5. Recall-miniCMV-sfGFP (Addgene #137995;Brock Lab AA158)

Primers

see **Oligonucleotides**

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)

[1] Make sure cells are transducible with lentivirus. Timing of lentiviral exposure and detectable expression of transgene will vary across cell types.

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

Computational

1. Linux Computing Environment (Such as University HPC)
2. Python >=3.8
3. Cell Ranger (for 10X analysis)
4. Pycashier

see also **Recommended Reagents**

Methods

ClonMapper Barcode Plasmid Library Assembly

In this step we will generate a high-diversity barcode plasmid library. At this stage there is an opportunity to customize the gRNA design. To maximize diversity you should order a forward oligonucleotide containing an N20 sequence. However, it's also possible to insert known sequences at either the 5'/3' end or to require alternating strong weak bases. This protocol as written should produce approximately 20 (50 mL) bacteria cell pellets.

1. Perform a 4X extension reaction to generate the double-stranded gRNA insert. Mix the below reagents to create a 50 µL reaction.[2]

Reagent	volume (µL)
5X Q5 Reaction Buffer	10
10 mM dNTPs	1
100 µM CROPseq-PrimeF-BgL-BsmBI	2
100 µM CROPseq-RevExt-BgL-BsmBI	1
Q5 Polymerase	0.5
nuclease-free water	to 50

2. Run the extension reaction on a thermocycler using the following settings, repeating steps 2-3 for 10 cycles:

Step	Temp (°C)	Time
1	98	2 min
2	65	30 sec
3	72	10 sec
4	72	2 min
5	4	hold

3. Clean and concentrate double-stranded gRNA insert PCR product and elute in 30 µL nuclease-free water. Confirm dsDNA assembly on 2% agarose gel by running single stranded DNA against PCR product.
4. Digest 5-10 µg of CROPseq vector backbone in a reaction containing 20 µL Digestion Buffer 3.1, 8 µL BsmBI, and nuclease-free water to 200 µL for 4 hours at 55°C
5. Run the digested backbone on a 1-1.5% low melting point agarose gel, then follow the instructions on a DNA gel purification kit to extract and purify the linearized plasmid band.

[2] Always use filtered pipette tips when working with DNA to prevent cross-contamination.

6. Ligate double stranded gRNA insert into linearized transfer vector backbone at a molar ratio of 10:1 in a 50X Golden Gate assembly reaction by mixing the below reagents[3]:

Reagent	volume (µL)
1.25 pmol linearized backbone	variable
12.5 pmol gRNA insert	variable
T4 Ligase Buffer	50
T7 Ligase	25
BsmBI	25
nuclease-free water	to 500

7. Run the Golden Gate assembly reaction on a thermocycler overnight using the following settings, repeating steps 1-2 for 99 cycles:

Step	Temp (°C)	Time
1	42	2 min
2	16	5 min
4	55	30 min
5	4	hold

8. Clean barcoding library plasmid pool using a DNA clean and concentrator kit and elute in 22 µL warm, nuclease-free water.[4]
9. 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.[5]
10. 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.[6]
11. Transform *E. coli* by electroporating with 1 pulse at 2.5 kV.[7]
12. Add 2 mL Recovery Media and gently pipet up and down immediately after electroporation, and transfer to a sterile 50 mL conical tube.
13. Repeat steps 10-12 three times
14. Allow cells to recover for 30 min at 37 °C with shaking at 250 rpm.

[3] A 1X Golden Gate assembly reaction is setup by mixing 25 fmol digested gRNA transfer vector backbone, 250 fmol double stranded gRNA barcode DNA, 1 µL T4 ligase buffer, 0.5 µL T7 ligase, 0.5 µL BsmBI, 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.

15. Pre-warm 2xYT agar plates with 100 µg/mL carbenicillin.
16. After recovery, perform dilution plating 1:10^4, 1:10^5, 1:10^6 on carbenicillin agar plates.
17. Incubate plates overnight at 37 °C.
18. Put the remaining transformant mixture into 500 mL 2xYT with 100 µg/mL carbenicillin in a 2 L flasks.
19. Incubate flasks at 30 °C overnight with shaking at 250 rpm.
20. The culture can be pelleted or midi/maxi prepped for usage.
21. Calculate transformation efficiency from dilution plating.[8]

ClonMapper Barcode Sampling

The diversity of the initial plasmid pool should be assessed to ensure a sufficiently high diversity. To do this, a two-stage PCR is performed first with primers flanking the gRNA, followed by a second reaction with primers containing Illumina indices/adapters.

For primer sequences see **Oligonucleotides**. For pre-generated stage 2 sequences containing i5/i7 adapters see docs.brocklab.com/clonmapper/sequences.

1. Midi-prep one tube of transformed *E. coli* from step *ClonMapper Barcode Plasmid Library Assembly* according to manufacturer's instructions.
2. Generate the phasing primer mixture 'CM-FWD-S1-PAS' by mixing equimolar amounts of CM-FWD-S1-PASx0, CM-FWD-S1-PASx4, CM-FWD-S1-PASx7, and CM-FWD-S1-PASx8.[9]
3. Prepare *stage 1* PCR reaction to amplify barcodes by mixing the following reagents:

Reagent	volume (µL)
5X Q5 Reaction Buffer	10
10 mM dNTPs	1
CM-FWD-S1-PAS	2.5
CM-REV-S1	2.5
Q5 Polymerase	0.5
100 ng DNA	variable
nuclease-free water	to 50

4. Amplify barcodes by running 50 µL reaction on a thermocycler using the following settings[10], repeating steps 2-4 for 10 cycles[11]:

[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] Pre-heat thermocycler to 98 °C before adding tubes to heat block.

[11] The number of cycles will depend on the starting template amount.

Step	Temp (°C)	Time
1	95	5 min
2	98	10 sec
3	63	30 sec
4	72	15 sec
5	72	2 min
6	15	hold

5. Clean *stage 1* reaction as described in *Appendix: AmpureXP Bead PCR Cleanup*.
6. Prepare *stage 2* PCR reaction to attach index sequences and Illumina adapters by mixing the following reagents:

Reagent	volume (µL)
5X Q5 Reaction Buffer	10
10 mM dNTPs	1
CM-FWD-S2-i5	2.5
CM-REV-S2-i7	2.5
Q5 Polymerase	0.5
4 ng *stage 1* amplicon	variable
nuclease-free water	to 50

7. Amplify the barcodes by running the 50 µL reaction on a thermocycler using the above cycling parameters from *stage 1*, repeating steps 2-4 for 8 cycles [12].
8. Clean *stage 2* reaction as described in *Appendix: AmpureXP Bead PCR Cleanup*.

ClonMapper Lentivirus Production

In this step, we will generate the lentivirus used to integrate barcode gRNA sequences into our cells of interest. It's important that the necessary precautions are taken when handling live lentivirus. You should consult your local intuitions instructions and viral-handling protocols. Given batch to batch variation, and the need to control viral titer for individual cell lines, it is helpful to make large concentrated batches of virus to give you sufficient material to establish barcoded cell lines.

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.

[12] The number of cycles will depend on the starting template amount.

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™ 3000, p3000™, VSV-G, psPAX, and ClonMapper barcode library plasmid to room temperature.[13,14]
4. Per well of a 6 well plate, prepare "Tube A" containing 125 µL Opti-MEM™ and 7 µL Lipofectamine™ 3000.[15]
5. Incubate "Tube A" at room temperature for 5 minutes.
6. Per well of a 6 well plate, prepare "Tube B" containing 125 µL Opti-MEM™, 1.5 µg psPax, 0.4 µg VSV-G, 3-5 µg ClonMapper barcode library plasmid and p3000™ (µL/µg DNA).
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 250 µL of the transfection mix slowly and dropwise to each well of HEK-293T cells.
10. 16-18 hours post-transfection, carefully remove and dispose of media containing Lipofectamine™ 3000 complexes and slowly replenish with DMEM supplemented with 20% FBS **without antibiotics**.[16,17]
11. 48 hours post-transfection, harvest viral containing supernatant and store in a 50 mL conical tube at 4 °C.[18,19,20]
12. Spin down collected viral containing supernatant at 500 x g for 10 min at 4 °C to remove residual HEK-293T cells.
13. Remove plunger from 20 mL syringe and attach to a 0.45 µm PES syringe filter.
14. Transfer viral supernatant to the 20 mL syringe.
15. Filter viral supernatant through 0.45 µm PES syringe filter into a fresh 50 mL conical tube to remove any remaining cell debris.
16. 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.[21]
17. Aliquot 25-50 µL of concentrated virus in threaded cryovials and store at -80 °C.[22,23]
18. After freezing use a small amount of virus to determine viral titer on your cell line of interest (see *Appendix: Determine Viral Titer*).

[13] 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.

[14] 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.

[15] Slowly dilute Lipofectamine™ complexes dropwise with Opti-MEM™ media with occasional flicking of the tube.

[16] You are working with live virus at this stage and beyond. Stringently adhere to all biosafety procedures. Bleach and UV all media and containers exposed to live virus and virus producing reagents.

[17] Cells exposed to lentivirus are fragile and extra care must be taken in removing and adding media.

[18] Virus should be stored in labeled secondary containment.

[19] Virus-producing HEK-293T cells should be bleached and UV irradiated in culture for at least 30 minutes to inactivate remaining virus before disposal.

[20] Never use a vacuum line to disposal of virus waste as this may produce aerosols.

[21] 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).

[22] Even just a single freeze-thaw cycle can drastically alter viral titer, be sure to minimize freeze-thaw cycles.

[23] Virus should be completed frozen and then thawed before calculating viral titer.

Integrating ClonMapper Barcodes in Cells

Using the concentrated lentivirus generated in the previous section you should transduce your cells of interest. See *Appendix: Determine Viral Titer* for typical forward and reverse transduction/viral titer procedures.

It is critical that you infect your cells with a low (~0.1) multiplicity of infection (MOI), in order to limit the chances of a multiple integration event. To control the MOI you should titer every batch of concentrated virus on your specific cell line of interest.

Once you have ascertained the viral titer, you should transduce your cells and separate blue fluorescent protein (BFP) positive cells using FACS. When sorting live cells you should take all necessary efforts to maximize cell viability for your cell line of interest. So long as you have a low MOI you can control the starting diversity of your barcoded cell library using the total number of sorted cells as a proxy. Note that you are likely to recover fewer barcodes than cells you sort due to stochastic outgrowth and death following sort. The total number of barcodes recovered is typically half of the initial number of cells sorted.

The actual number of barcodes should be confirmed as soon as the population has sufficiently outgrown and archives have been prepared. Cell libraries should also be frequently sampled prior to and following any experiments to monitor changes in barcode diversity through the course of routine cell culture maintenance.

ClonMapper Barcode Sampling of Cells

Barcodes are amplified from cellular genomes similar to the plasmid library as described above. Barcoded cell libraries should be assessed to ensure a sufficiently high diversity at time of initial archival. Crucially, any experiments involving barcoded cells should be carried out diligently to prevent population skewing over time. The below protocol assumes a high amount of starting material (2 µg) in order to sample ~300,000 cells. However, it's possible to amplify from a smaller starting amount of gDNA by increasing the total number of cycles from 20.

For primer sequences see **Oligonucleotides**. For pre-generated stage 2 sequences containing i5/i7 adapters see docs.brocklab.com/clonmapper/sequences.

Preparing Samples for Sequencing

1. To assess cell barcode diversity harvest cells from culture and collect into cell pellet.[24]
2. Isolate genomic DNA from cell pellet using kit or standard protocol and proceed to PCR amplification.
3. Generate the phasing primer mixture 'CM-FWD-S1-PAS' by mixing equimolar amounts of CM-FWD-S1-PASx0, CM-FWD-S1-PASx4, CM-FWD-S1-PASx7, and CM-FWD-S1-PASx8.[25]
4. Prepare *stage 1* PCR reaction to amplify barcodes by mixing the following reagents[26]:

[24] It is important to ensure that you have enough cells to sufficiently sample your population depending upon the initial barcode diversity.

[25] Universal phase amplicon sequencing primers are used to add more diversity to the sequencing reads which helps prevents sequencing errors.

[26] 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.

Reagent	volume (µL)
5X Q5 Reaction Buffer	10
10 mM dNTPs	1
CM-FWD-S1-PAS	2.5
CM-REV-S1	2.5
Q5 Polymerase	0.5
2 µg DNA	variable
nuclease-free water	to 50

5. Amplify barcodes by running 50 µL reaction on a thermocycler using the following settings[27], repeating steps 2-4 for 20 cycles[28]:

Step	Temp (°C)	Time
1	95	5 min
2	98	10 sec
3	63	30 sec
4	72	15 sec
5	72	2 min
6	15	hold

6. Clean *stage 1* reaction as described in *Appendix: AmpureXP Bead PCR Cleanup*.
7. Prepare *stage 2* PCR reaction to attach index sequences and Illumina adapters by mixing the following reagents:

Reagent	volume (µL)
5X Q5 Reaction Buffer	10
10 mM dNTPs	1
CM-FWD-S2-i5	2.5
CM-REV-S2-i7	2.5
Q5 Polymerase	0.5
4 ng *stage 1* amplicon	variable
nuclease-free water	to 50

[27] Pre-heat thermocycler to 98 °C before adding tubes to heat block.
[28] The number of cycles will depend on the starting template amount.

8. Amplify the barcodes by running the 50 µL reaction on a thermocycler using the above cycling parameters from *stage 1*, repeating steps 2-4 for 8 cycles. [29]
9. Clean *stage 2* reaction as described in *Appendix: AmpureXP Bead PCR Cleanup*.

Processing Barcode Sequencing Data

See pycashier for more info about how to get started processing `fastq` data to extract barcode information.

Recall Plasmid Assembly

Once you have identified a barcode of interest within your cell library, you can generate a "Recall" plasmid to drive expression of green fluorescent protein (GFP) to isolate or track your clone of interest.

1. 3 pairs of overlapping oligos containing the barcode sequence of interest flanked by overlapping sequences should be ordered according to **Table 1**.[30]
2. In separate tubes, mix each of the 100 µM oligo pairs together:

- Tube AB: 10 µL Bg-AB-fwd + 10 µL Bg-AB-rev
- Tube BC: 10 µL Bg-BC-fwd + 10 µL Bg-BC-rev
- Tube CD: 10 µL Bg-CD-fwd + 10 µL Bg-CD-rev

3. Heat each to 80 °C and let cool to create DNA blocks containing a barcode, a PAM site, and overhang sequences.[31]
4. Ligate DNA blocks together creating the **3X-barcode array** by mixing the following reagents:

Reagent	volume (µL)
Tube "AB"	10
Tube "BC"	10
Tube "CD"	10
10 mM dNTPs	5
10x T4 PNK Buffer	5
T4 PNK	1
nuclease-free water	9

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 170 bp.

[29] The number of cycles will depend on the starting template amount.
[30] 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'.
[31] This process anneals the single stranded DNA oligos together, creating short double stranded DNA blocks that will be ligated together in the next step.

8. Ligate the **3X-barcode-array** into the recall plasmid backbone at a molar ratio of 10:1 in a Golden Gate assembly reaction by mixing the following reagents:

Reagent	volume (µL)
Recall-miniCMV-sfGFP	25 fmol
3X-barcode-array	250 fmol
T4 ligase buffer	1 µL
T7 ligase	0.5 µL
BbsI	0.5 µL
nuclease-free water	to 10 µL

9. Run the Golden Gate assembly reaction on a thermocycler using the following settings, repeating steps 1-2 for 35 cycles:

Step	Temp (°C)	Time
1	42	2 min
2	16	5 min
3	55	30 min
4	4	hold

10. Transform bacteria with golden gate product. See Addgene for standard protocol.
11. Verify insertion of barcode array into Recall-miniCMV-sfGFP backbone via Sanger sequencing.

Recall and Isolation of Barcoded Cells

With a barcode-specific recall vector you can isolate a clonal sub-population from your barcoded cell library. See below for the general procedure, but note that it may be necessary to optimize transfection and live-cell sorting for your specific cells of interest.

See notes [32,33]

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.[34]

[32] Lipofectamine effiency can vary significantly between cell lines. It's recommended you optimize transfection with a plasmid containing a constitutively promoter.

[33] 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.

[34] Slowly dilute Lipofectamine™ complexes dropwise with Opti-MEM™ media with occasional flicking of the tube.

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, 275 ng dCas9-VPR plasmid and 2 µL/µg DNA of p3000.
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.[35]
10. 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.
11. 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.[36]
12. Set sort gate on GFP and BFP double positive gate indicative of a recalled cell.[37]
13. Sort cells in GFP and BFP double positive gate.[38]
14. Maintain sorted cells in culture with complete growth medium.

[35] EDTA and FBS in FACS buffer help to prevent cell clumping. For extra-sticky cells, use 5 mM EDTA in FACS buffer.

[36] 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.

[37] When sorting for recalled cells, use stringent gating. Ensure that 0% of negative control and single positive samples appear in the sorting gate.

[38] Single cell sorting can be performed for isolation and growth of clonal populations.

Appendix

See below for general purpose procedures related to ClonMapper.

AmpureXP Bead PCR Cleanup

1. Transfer 50 µL PCR amplification product to a nuclease-free microcentrifuge tube
2. Allow AmpureXP beads to come to room temperature.
3. Add 35 µL (0.7X) AmpureXP beads and mix well with vortexing or pipetting up and down 10 times.
4. Incubate at room temperature for 5 minutes.
5. Place the tube on a magnetic rack and allow solution to clear (5-10 minutes).
6. While the tube is on the rack transfer the clear supernatant to a new tube without disturbing the bead pellet.
7. Add 45 µL (1.6-0.7x) AmpureXP beads to the supernatant from step 10 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. With the tube still in the rack, aspirate the clear supernatant.
11. With the tube still in the rack, add 180 µL of 80% ethanol and allow it to sit for 30 seconds.[39]
12. With the tube still in the rack, aspirate the ethanol and repeat step 11.
13. Remove supernatant and allow bead to dry for no **more** than 5 minutes.[40]
14. Remove tube from the magnetic rack and elute DNA by adding 42 µL of nuclease-free water.
15. Incubate at room temperature for 10 minutes.
16. Transfer tube to magnetic rack and collect 40 µL of purified PCR product after solution has cleared (5-10 minutes).[41]
17. Quantify DNA yield with a high sensitivity fluorometry kit ensuring yield between 0.5-10 ng/µL.

Determine Viral Titer

See [42,43]

Titering on Adherent Cells (Forward Procedure)

[44]

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.

[39] 80% ethanol should be prepared fresh for each PCR cleanup.
[40] Do not over dry the beads, this can result in a loss of yield and quality.
[41] 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.
[42] Viral titer will vary between cell type and with each new virus preparation.
[43] 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.
[44] 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.

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.[45,46]
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 to get 10 µg/mL).[47]
4. Place 600 µL of hexadimethrine bromide containing medium into separate microcentrifuge tubes.
5. Add virus in increasing amounts to each tube.
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.[48,49]
8. Incubate for an additional 32 hrs at 37 °C, then remove medium and wash each well gently with PBS.[50]
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. Perform this step three times to ensure removal of trace virus before flow cytometry.
11. Resuspend cells in chilled FACS Buffer.[51]
12. Keep cells on ice and continue to step *Flow Cytometry to Determine Viral Titer*

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).[52]
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.[53]

[45] It is very important to know the number of cells at the time of transduction. This number is used to calculate viral titer.
[46] 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.
[47] Hexadimethrine bromide is a cationic solution that assists in viral adsorption to cells (Davis, Morgan, and Yarmush 2002). 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.
[48] 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.
[49] Lentiviral exposure times range between 12-48 hours. Lentiviral exposure time should be minimized to reduce the occurrence of multiple viral integrations.
[50] Lentivirus transduced cells are very fragile and should be handled with added care.
[51] EDTA and FBS in FACS buffer help to prevent cell clumping. For extra-sticky cells, use 5 mM EDTA in FACS buffer.
[52] Hexadimethrine bromide is a cationic solution that assists in viral adsorption to cells (Davis, Morgan, and Yarmush 2002). 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.
[53] 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.

6. Incubate for 16 hrs at 37 °C.[54,55]
7. Transfer cell suspensions to sterile 1.7 mL microcentrifuge tubes and spin down at 500 x g for 5 minutes at 4 °C.[56]
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.[57]
10. Wash cell pellets with PBS and repeat spin.[58]
11. Resuspend cells in chilled FACS Buffer.[59]
12. Keep cells on ice and continue to step *Flow Cytometry to Determine Viral Titer*

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.[60]
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.[61]
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.[62]
5. 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.[63]

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

ClonMapper Viral 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

[54] 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.
[55] Lentiviral exposure times range between 12-48 hours. Lentiviral exposure time should be minimized to reduce the occurrence of multiple viral integrations.
[56] Use a pipette to remove lentivirus containing supernatant and dispose of in bleach. Do not vacuum aspirate, vacuums can cause dangerous viral aerosols.
[57] Lentivirus transduced cells are very fragile and should be handled with added care when pipetting.
[58] Perform this step three times to ensure removal of trace virus before flow cytometry.
[59] EDTA and FBS in FACS buffer help to prevent cell clumping. For extra-sticky cells, use 5 mM EDTA in FACS buffer.
[60] Ensure proper controls for flow. Minimally have a positive control expressing BFP and a negative control expressing no fluorescent proteins.
[61] 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.
[62] 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).
[63] 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

0.1 (**Equation 2**) to minimize the occurrence of multiple barcode integrations.[64,65]

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.[66]
3. Set sort gate on BFP-positive cells indicative of a productive sgRNA barcode.[67]
4. Sort cells on BFP-positive gate via FACS.
5. Maintain sorted cells in culture with complete growth medium.

$$\text{MOI [TU/cell]} = \frac{(\text{Volume of Virus needed [mL]}) \times (\text{Titer of Virus [TU/mL]})}{(\text{Number of cells exposed to virus})} = 0.1$$

[64] Example: If your viral titer was 6.0 x 10^6 TU/mL and you wanted to infect 3.0 x 10^6 cells at an MOI of 0.1, you would need to subject the 3.0 x 10^6 cells to 50 µL of virus.

[65] 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 (Fehse et al. 2004; Kustikova et al. 2003).

[66] Ensure proper controls for flow. Minimally have a positive control expressing BFP and a negative control expressing no fluorescent proteins.

[67] 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.

Tables

Table 14: Oligonucleotides

Name	Sequence (5' to 3')	Notes
CROPseq-PrimeF-BgL-BsmBI	GAGCCTCGTCTCCCACCGNNNNNNNNNNNNNNNNNN-NNGTTTTGAGACGCATGCTGCA	The N20 sequence is a random string of oligonucleotides
CROPseq-RevExt-BgL-BsmBI	TGCAGCATGCGTCTCAAAAC	
CM-FWD-S1-PASx0	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTT-GTGGAAAGGACGAAACAC	
CM-FWD-S1-PASx4	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCA-ACTTGTGGAAAGGACGAAACAC	
CM-FWD-S1-PASx7	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGC-CACCCTTGTGGAAAGGACGAAACAC	
CM-FWD-S1-PASx8	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTAG-TGAATCTTGTGGAAAGGACGAAACAC	
CM-REV-S1	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGG-ACTAGCCTTATTTTAACTTGCTATTTCTAGCTC	
CM-FWD-S2-i5	AATGATACGGCGACCACCGAGATCTACACNNNNNNNN-NTCGTCGGCAGCGTC	The N8 sequence is where the i5 Illumina index should be placed
CM-REV-S2-i7	CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTCT-CGTGGGCTCGG	The N8 sequence is where the i7 Illumina index should be placed
BgN20-AB-fwd	TACTCGACCAAGAACCGCANNNNNNNNNNNNNNNNN-NNNAGGTGGATTAGTTCTCT	Insert barcode in place of N20
BgN20-AB-rev	AAGCAGAGAACTAATCCACCTNNNNNNNNNNNNNNNN-NNNNNTGCGGTTCTTGGTCG	Insert reverse-complement barcode in place of N20
BgN20-BC-fwd	GCTTGTCCTGCGGTTACCCNNNNNNNNNNNNNNNN-NNNAGGCTGTAATCCAGCTG	Insert barcode in place of N20
BgN20-BC-rev	AGCGCAGCTGGATTACAGCCTNNNNNNNNNNNNNNNN-NNNNNGGGTAACCGCAGGAC	Insert reverse-complement barcode in place of N20
BgN20-CD-fwd	CGCTGTGGTATCACTCGTCNNNNNNNNNNNNNNNN-NNNAGGCTCAGCTAAGGTGC	Insert barcode in place of N20

Table 14: Oligonucleotides

Name	Sequence (5' to 3')	Notes
BgN20-CD-rev	CATTGCACCTTAGCTGAGCCTNNNNNNNNNNNNNNN-NNNNNGACGAGTGATACCAC	Insert reverse-complement barcode in place of N20

Table 15: Recommended Reagents

Name	Vendor	Catalog No.
AMPure XP Reagent	Beckman Coulter	A63880
BbsI-HF	NEB	R3539S
BsmBI-v2	NEB	R0739S
NEBuffer r3.1	NEB	B6003S
Q5 Hot Start Polymerase	NEB	M0493S
T4 Ligase	NEB	M0202S
T4 PNK	NEB	M0201S
T7 Ligase	NEB	M0318S
Genomic DNA Mini Kit	ThermoFisher	K182001
Lipofectamine 2000	ThermoFisher	11668030
Lipofectamine 3000	ThermoFisher	L3000001
OptiMEM	ThermoFisher	31985070
Typan Blue Stain	ThermoFisher	T10282
2xYT medium	Millpore Sigma	Y2377-250G
Carbenicillin	Millpore Sigma	C1389-250MG
DMEM	Millpore Sigma	D5671
Hexadimethrine bromide	Millpore Sigma	TR-1003-G
Plasmid Plus Kit (Midi)	Qiagen	12941
DNA Clean and Concentrator-5	Zymo	D4013

Acknowledgements

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

References

Al'Khafaji, Aziz M., Daniel Deatherage, and Amy Brock. 2018. "Control of Lineage-Specific Gene Expression by Functionalized gRNA Barcodes." *ACS Synthetic Biology* 7 (10): 2468–74. https://doi.org/10.1021/acssynbio.8b00105.

Blundell, Jamie R., and Sasha F. Levy. 2014. "Beyond Genome Sequencing: Lineage Tracking with Barcodes to Study the Dynamics of Evolution, Infection, and Cancer." *Genomics* 104 (6): 417–30. https://doi.org/10.1016/j.ygeno.2014.09.005.

Chavez, Alejandro, Jonathan Scheiman, Suhani Vora, Benjamin W Pruitt, Marcelle Tuttle, Eswar P R Iyer, Shuailiang Lin, et al. 2015. "Highly Efficient Cas9-mediated Transcriptional Programming." *Nature Methods* 12 (4): 326–28. https://doi.org/10.1038/nmeth.3312.

Datlinger, Paul, André F Rendeiro, Christian Schmidl, Thomas Krausgruber, Peter Traxler, Johanna Klughammer, Linda C Schuster, Amelie Kuchler, Donat Alpar, and Christoph Bock. 2017. "Pooled CRISPR Screening with Single-Cell Transcriptome Readout." *Nature Methods* 14 (3): 297–301. https://doi.org/10.1038/nmeth.4177.

Davis, Howard E., Jeffrey R. Morgan, and Martin L. Yarmush. 2002. "Polybrene Increases Retrovirus Gene Transfer Efficiency by Enhancing Receptor-Independent Virus Adsorption on Target Cell Membranes." *Biophysical Chemistry* 97 (2-3): 159–72. https://doi.org/10.1016/S0301-4622(02)00057-1.

Fehse, B, O S Kustikova, M Bubenheim, and C Baum. 2004. "Pois(s)on – It's a Question of Dose…." *Gene Therapy* 11 (11): 879–81. https://doi.org/10.1038/sj.gt.3302270.

Kebschull, Justus M., and Anthony M. Zador. 2018. "Cellular Barcoding: Lineage Tracing, Screening and Beyond." *Nature Methods* 15 (11): 871–79. https://doi.org/10.1038/s41592-018-0185-x.

Kustikova, Olga S., Anke Wahlers, Klaus Kühlcke, Birgit Stähle, Axel R. Zander, Christopher Baum, and Boris Fehse. 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 (12): 3934–37. https://doi.org/10.1182/blood-2003-05-1424.