Nov 21, 2025

Single Cell Transcriptional Perturbome in Pluripotent Stem Cell Models: Protocols 1-4

DOI
https://dx.doi.org/10.17504/protocols.io.e6nvw4q29lmk/v1

elisa.balmas[1]

[1]University of Torino

elisa.balmas

DOI: https://dx.doi.org/10.17504/protocols.io.e6nvw4q29lmk/v1

Protocol Citation: elisa.balmas 2025. Single Cell Transcriptional Perturbome in Pluripotent Stem Cell Models: Protocols 1-4. **protocols.io** https://dx.doi.org/10.17504/protocols.io.e6nvw4q29lmk/v1

Protocol status: Working
We use this protocol and it's working

Created: October 18, 2025

Last Modified: November 21, 2025

Protocol Integer ID: 230162

Keywords: 10X genomics, iPS2-10X-seq, iPS2-CITE-seq, iPS2-MULTI-seq, seq from 10x genomic, conjunction with the appropriate 10x genomics protocol, appropriate 10x genomics protocol, use of cell multiplexing oligo, cell multiplexing oligo, parallel to the standard gene expression library, cell transcriptome, pluripotent stem cell model, 10x genomic, seq relies on microfluidics partitioning, gene expression kit, standard gene expression library, 10x genomics single cell multiome atac, single cell transcriptional perturbome, gene expression library, unique cell, cell doublet, shrna barcode library library, absence of additional multiplexing barcode, cell number, microfluidics chip, single microfluidics channel, microfluidics partitioning, additional multiplexing barcode, shared unique cell, multiplexed sample, manufacturer for barcoded sample pool, treated cell, seq rely, silico matching of uci, barcoded sample pool, based scrna, gene expression

Funders Acknowledgements:
Giovanni Armenise-Harvard Foundation
Grant ID: Career Development Award 2021
ERC starting grant
Grant ID: 101076026

Disclaimer

Contact elisa.balmas@unito.it and or alessandro.bertero@unito.it for further tips and assistance

Abstract

iPS2-10X-seq relies on microfluidics partitioning-based scRNA-seq from 10X Genomics. Despite the higher costs, the ease and reproducibility of this commercial method have led to its widespread adoption. The procedure integrates the manufacturer's protocol to generate an additional NGS library containing UCI-BCs. This is sequenced in parallel to the standard gene expression library to enable in silico matching of UCI-BCs and single-cell transcriptomes through their shared unique cell barcodes. Of note, iPS2-10X-seq provides built-in support for the cost-saving strategy of reasonably overloading the microfluidics chip, since cell doublets are readily identified by the co-expression of two different UCI-BCs and filtered out as part of the standard catcheR_10Xcatch pipeline. Thus, it is possible to load the cell number recommended by the manufacturer for barcoded sample pools even in the absence of additional multiplexing barcodes. Other cost-saving strategies include the use of cell multiplexing oligos (CMOs) or barcoded antibodies against housekeeping cell surface antigens to multiplex tet-untreated and tet-treated cells in a single microfluidics channel, enabling robust control versus knockdown experimental design without having to perform two separate reactions.

This Protocol describes the following procedures:

1. Prepare the gene expression library
2. Prepare the shRNA barcode library
3. Library pooling and NGS

This protocol was optimized using the *Chromium Next GEM Single Cell 3′ Reagent Kits v3.1* (Dual Index) from 10X Genomics and has been successfully validated with the *next-generation GEM-X Single Cell 3′ Reagent Kits v4*, without requiring changes to primer sequences. It must be used in conjunction with the appropriate 10X Genomics protocols: **CG00315 Rev E** (single sample), **CG000388 Rev B** (multiplexed samples), or **CG000419 Rev D** (high-throughput) for v3.1;, and **CG000731 Rev A** for v4. Notably, v4 no longer supports cell multiplexing oligo (CMO)-based on sample multiplexing but remains fully compatible with antibody-derived tag (ADT)–based strategies.

A first variation, iPS2-CITE-seq, enables surface protein profiling and ADT-based multiplexing using barcoded TotalSeq-B antibodies from BioLegend. When using antibodies, refer to protocol **CG000149 Rev D** (antibody labeling) in combination with **CG000317 Rev E** for v3.1 or **CG000731 Rev A** for v4.

A second variation, iPS2-multi-seq, adapts the protocol for use with the 10X Genomics Single Cell Multiome ATAC & Gene Expression kit. It should be paired with **CG000365 Rev C** (nuclei isolation) and **CG000338 Rev F** (library construction).

Image Attribution

Figure SP3.1

Materials

- NEBNext High-Fidelity 2X PCR Master Mix
- iPS2-10X-seq_true5 (10 µM)
- iPS2-10X-seq_inner (10 µM)
- Preamplified 10X Genomics cDNA (50 ng)
- Nuclease-free water
- Dual Index Kit TT Set A primers
- Qubit dsDNA High Sensitivity kit

Protocol 1: Generation of iPS2-seq plasmids and hiPSCs

1 The iPS2-seq pipeline starts with a two-step cloning procedure for a barcoded pool of shRNAs into an AAVS1 locus targeting plasmid to generate all-in-one tet inducible cassette.
This begins with designing single stranded DNA (ssDNA) oligonucleotides each carrying an shRNA with a polythymidine (pT) Pol III terminator sequence, a uniquely matched barcode (BC), and a random sequence serving as a unique clonal identifier (UCI).
The shRNA and UCI-BC sequences are separated by a short multicloning site (SalI, SwaI, AscI) that is needed at a later stage, and are flanked by short regions of homology to the destination vector to enable assembly.
The oligos are synthesized, pooled, and converted to double stranded DNA (dsDNA) through a single round of isothermal primer extension from the 3′ end using a mesophilic DNA polymerase with strong strand displacement activity.
This is essential to overcome the secondary structure of shRNAs in the oligos and minimize representation biases due to differential hairpin strength.
dsDNA fragments are then cloned through Gibson assembly into an AAVS1 targeting plasmid downstream of an inducible H1 Pol III promoter and upstream of a pA signal.
The vector contains a puromycin (puro) resistance gene trap and is based on our previously reported plasmid, but was modified to remove repetitive sequences that led to recombination errors in the next cloning step.
The plasmid also includes a novel thymidine kinase (TK) cassette outside of the AAVS1 homology arms. This allows subsequent negative selection of cells carrying random, off-target integrations of the whole plasmid as they become sensitive to fialuridine (FIAU), a thymidine analog preferable to ganciclovir.
The resulting intermediate plasmids are then modified in pool to insert a CAG-OPtettR cDNA cassette between the shRNA and UCI-BC sequence.
This second cloning step reconstitutes the tet-ON system while marking the 3′ UTR of the OPtettR with the UCI-BC sequence matched to a given shRNA.
This is performed via restriction digestion and ligation, as the CAG promoter is too GC-rich for efficient and reliable PCR amplification.
An optional strategy built into the system allows depletion of the intermediate plasmid from the final population through restriction digestion with SwaI.
The final plasmids are utilized to engineer pools of hPSCs enriched for clones expressing single inducible shRNAs.
Plasmids are transfected alongside vectors encoding for obligate heterodimer zinc finger nucleases (ZFNs) to induce double-strand breaks specifically at the AAVS1 locus and facilitate homology-directed repair (HDR) of the inducible barcoded shRNA cassette.
A second AAVS1 targeting plasmid is also co-delivered: this is a filler vector that carries no relevant cargo but has a neomycin (neo) resistance gene trap that allows enrichment

of cells that integrated an shRNA on one AAVS1 allele and the filler sequence on the other allele; these biallelic edited cells are co-selected with puro and neo.
Additional negative selection with FIAU kills clones that carry randomly integrated shRNAs.
Collectively, this genome editing strategy is designed to reduce the frequency of double shRNA integrations and maximise the fraction of cells expressing a single shRNA from the AAVS1 locus.
Of note, the first cloning step can result in a low rate of swap between the shRNA and its expected BC based on oligonucleotide design and synthesis.
This issue likely arises from DNA polymerase template switching during dsDNA synthesis and/or Gibson assembly.
The problem can be efficiently countered by next generation sequencing (NGS) analysis of Illumina-compatible libraries spanning the shRNA and UCI-BC region from the intermediate plasmid pool; this enables re-assignment of swapped shRNAs to the observed barcode, leveraging the fact that each plasmid also carries a UCI (also essential for tracking hPSC clones downstream).

2 **Plan the screening**

Before initiating a screening experiment, it is essential to define three key elements: screening objectives, available budget, and overall feasibility.

The flow chart below summarizes key parameters for optimizing molecular cloning, genome editing, and sequencing efficiency, followed by a worked example: a screen targeting 18 genes with 5 shRNAs per gene plus 10% control constructs.

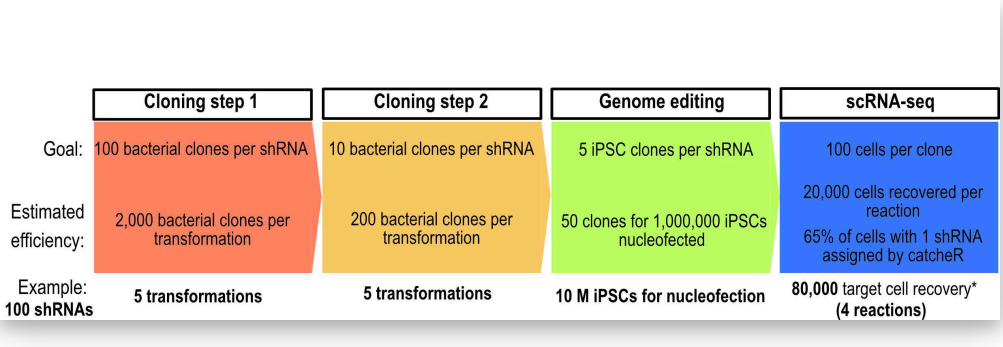


Figure SP1.1. Screening design example
Flow chart illustrating how to estimate the number of cells required for sufficient iPS2-seq coverage, accounting for dropout rates from cloning to transfection and single-cell sequencing. The top section defines the screening goal and provides an example estimation based on our protocol; the lower section shows a worked example with 100 shRNAs. scRNA-seq calculations are based the maximum loading capacity per reaction for either the Chromium Next GEM Single Cell 3' HT GEM Kit v3.1 or the Chromium GEM-X Universal 3' v4 Kit.

2.1 Effective screening design should begin with:

Target gene selection — Define the genes to perturb and decide how many shRNAs per gene will be used. We recommend a minimum of 3 shRNAs per gene, and ideally 5, to buffer against off-target effects or poorly performing constructs. Include at least 10% non-targeting control shRNAs.

2.2 **Cell editing scale** — Estimate the number of cells to genome-edit in order to achieve robust clonal representation for downstream iPS2-10X-seq analysis. An example calculation is provided in Figure SP1.1, accounting for both bacterial clone yield and the number of hiPSC clones required for sequencing.

2.3 **Sequencing depth** — Determine how many cells need to be profiled to capture biological variability and achieve statistical power. Depending on experimental goals and budget, controls may include non-targeting shRNAs (e.g., B2M, SCR), tet-untreated cells, or both. We strongly recommend including tet-untreated cells to control for clonal variability.

3 **Design shRNA oligonucleotides**

ssDNA oligonucleotides for iPS2-seq can be designed using catcheR_design as described in Supplemental Protocol 4 - Oligonucleotides design. This function automatically performs the following steps, reducing the chance of manual errors when designing multiple oligos:

1. Identify validated shRNAs or design appropriate sequences. In our lab we use the Broad Institute TRC library, available at the GPP Web Portal. When using this resource take the ~58 bp forward oligos and remove the sequences required for cloning in pLKO vectors (typically the first 4 nucleotides and the last 6 nucleotides).
2. If an shRNA does not start with A or a G, add an initial G (required for efficient transcription by the H1 Pol III promoter).
3. Construct the cloning oligos by adding the following sequences to each shRNA, in order:

- At the 5′, the Gibson homology region 5′-AGTTCCCTATCAGTGATAGAGATCCC-3′
- At the 3′, the polythymidine (pT) Pol III terminator sequence 5′-TTTTTTTT-3′
- At the 3′, SalI, SwaI, and AscI restriction sites 5′-GTCGACATTTAAATGGCGCGCC-3′
- At the 3′, a random sequence (UCI; i.e., 5′-NNNNNN-3′)
- At the 3′, an shRNA-specific barcode (BC; i.e., 5′-CAGTTCCA-3′)
- At the 3′, the Gibson homology region 5′-GTAGCTGCGTGATCAGC-3′

The resulting oligos should look like Figure SP1.2

Note: These sequences are passed as arguments to catcheR_design. The recommended minimum length for UCIs and BCs is 6 and 8 bp, respectively. Tools such as **BARCOSEL** can be used to generate and select BCs with optimal nucleotide distribution.

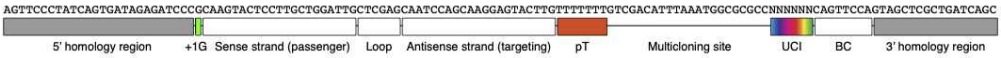


Figure SP1.2. Example of iPS2-seq ssDNA oligo
Annotated SMAD2.1 shRNA sequence extracted from TRC ID:TRCN0000010477 and modified by
catcheR_design.

1. Modify or replace any cloning oligo that contains BglII (5′-AGATCT-3′) or MluI (5′-ACGCGT-3′) restriction sites that would interfere with subsequent cloning steps
2. Synthesize oligos as high quality ssDNA. In our lab we use the oPools Oligo Pools service from Integrated DNA Technologies

Pooled cloning step 1: obtain the intermediate plasmid pool

4 This section outlines the pooled conversion of barcoded shRNA-containing ssDNA oligos into dsDNA by second-strand synthesis (SSS), and the insertion of such fragments in pAAV-Puro_siKD2.0.

1. Prepare the oligo pool:

4.1 (a) Spin down the lyophilized ssDNA and resuspend it at 50 µM in nuclease-free water

4.2 (b) Dilute the 50 µM stock 1:10 in nuclease-free water to make a 5 µM working stock

5 2. Convert ssDNA to dsDNA using Bst3.0 DNA polymerase

5.1 (a) Denature and snap cool the ssDNA oligo pool
i. Mix the following for 8 reactions in a single tube on ice:

	A	B	C	D
	Reagent	one reaction (µL)	eight reactions (µL)	Final conc.
	ssDNA oligo pool (5µM)	5	40	0.5nM
	siKDbc_REV Primer (10µM)	2.5	20	0.5µM
	Deoxynucleotide (dNTP) Solution Mix (10mM)	2	16	400µM
	Nuclease-free water	15.5	124	

	A	B	C	D
	Total volume	25	200	

Table SP1.1. SSS reaction mix 1

ii. Distribute 25 µL of the reaction mix into 8 PCR tubes
iii. Place PCR tubes in a thermocycler and run the following cycling conditions:

	A	B	C	D
	Steps	Temperature	Time	Cycles
	Denaturation	98°C	5 min	1
	Hold (rapid ramp down)	4°C	Hold	

Table SP1.2. Cycling conditions for ssDNA denaturation and snap cooling

5.2 (b) Perform SSS
i. Mix the following for 8 reactions in a single tube on ice:

	A	B	C	D
	Reagent	One reaction (µL)	Eight reactions (µL)	Final conc.
	MgSO4 (100mM)	1	8	2mM
	Isothermal Amplification Buffer II (10X)	5	40	1X
	Bst 3.0 DNA Polymerase (8 U/µL)	2	16	0.32 U/µL
	Nuclease-free water	17	136	
	Reaction from previous step	25	200	
	Total volume	25	200	

Table SP1.3. SSS reaction mix 2

ii. Add 25 µL of this reaction mix to each of the 8 PCR tubes from step 2(a)iii

iii. Place PCR tubes in a thermocycler and run the following cycling conditions:

	A	B	C	D
	Steps	Temperature	Time	Cycles
	Annealing	55°C	5 min	1
	Extension	72°C	15 min	1
	Inactivation	80°C	5 min	1
	Slow ramp down	75°C	-0.1deg/s	1
	Hold	75°C	4 min	1
	Slow ramp down	70°C	-0.1deg/s	1
	Hold	70°C	4 min	1
	Slow ramp down	10°C	-0.1deg/s	1
	Hold	10°C	Hold	

Table SP1.4. Cycling conditions for SSS

5.3 (c) Remove residual ssDNA
i. Add 4 µL of exonuclease I (20 U/µL) to each PCR tube and pipette mix
ii. Place PCR tubes in a thermocycler and run the following cycling conditions:

	A	B	C	D
	Steps	Temperature	Time	Cycles
	Digestion	37°C	30 min	1
	Inactivation	80°C	20 min	1
	Slow ramp down	75°C	-0.1deg/s	1
	Hold	75°C	4 min	1
	Slow ramp down	70°C	-0.1deg/s	1
	Hold	70°C	4 min	1
	Slow ramp down	10°C	-0.1deg/s	1

	A	B	C	D
	Hold	10°C	Hold	

Table SP1.5. Cycling conditions for ssDNA removal

5.4 (d) Clean up and concentrate the dsDNA
i. Pool all 8 reactions into a single tube
ii. Column purify using QIAquick PCR Purification Kit, following its manual
iii. Elute in 30 µL elution buffer
Note: the dsDNA shRNA pool can be stored at -20 °C for at least one week

6 Prepare the plasmid backbone

6.1 (a) Prepare a digestion mix and an undigested control mix following the table below:

	A	B	C	D
	Reagent	Digestion reaction (µL)	Undigested control (µL)	Final conc.
	pAAV-Puro_siKD2.0 plasmid	Variable	Variable	5µg
	FastDigest Green Buffer (10X)	15	1.5	1X
	FastDigest MluI	5	0	
	FastDigest BglII	5	0	
	FastAP (1 U/µL)	5	0.5	0.03 U/µL
	Nuclease-free water	Top up to 150	Top up to 15	
	Total volume	150	15	

Table SP1.6. pAAV-Puro_siKD2.0 restriction digestion mix

6.2 (b) Incubate at 37 °C overnight (~16 h), followed by 10 min at 80 °C for heat inactivation

7 Perform electrophoretic separation of the insert (step 2(d)iii) and backbone (step 3b) using 2% and 0.8% (w/v) agarose gels in TBE, respectively. Cast gels including SYBR

Safe DNA gel stain to visualize DNA, preferably using a blue light transilluminator (avoid UV light if possible). Figure SP1.3 shows exemplary electrophoreses

7.1 (a) Use 50 bp and 1 kbp DNA ladders for the insert and backbone, respectively

7.2 (b) Run 10% of each sample in a separate well and use this to monitor the electrophoretic separation, so as to avoid unnecessary exposure to high-energy light of the gel extracted DNA (particularly important if using UV light by necessity)

7.3 (c) Use FastDigest Green Buffer (10X) as loading dye for the insert: commonly used DNA loading dyes containing bromophenol blue and/or xylene cyanol FF can interfere with visualization of the small shRNA pool fragment

7.4 (d) If a smear is present in the insert, continue electrophoresis until the 135 bp band is clearly separated

7.5 (e) Run 90% of the digested empty plasmid (~140 µL) in a single well as this will facilitate gel extraction: merge multiple wells using autoclave tape if needed

7.6 (f) Load the plasmid undigested control next to 10% of the digestion reaction and continue the electrophoretic separation until the 8715 bp backbone band is completely separated from the undigested control (this can take several hours: use a low voltage and/or change buffer to prevent gel overheating). Incomplete separation can result in contamination of undigested or partially digested plasmid, which can dramatically increase transformation background

7.7

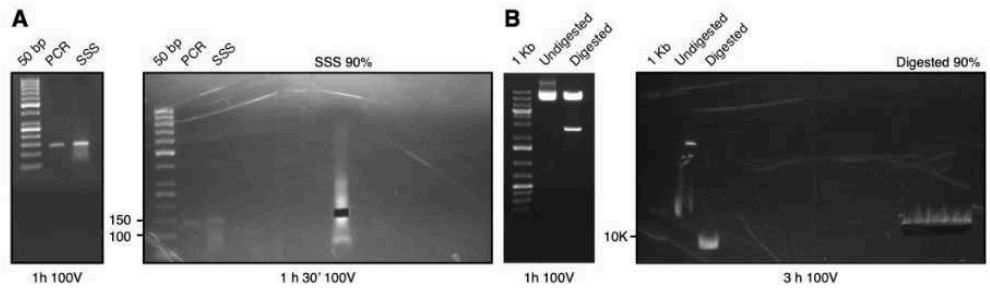


Figure SP1.3. Electrophoretic separation and gel extraction during cloning step 1 (A) Separation of shRNA pool insert (SSS); a PCR-amplified fragment indicates the expected molecular weight. (B) Separation of digested plasmid backbone.

8 Use a clean scalpel to excise the DNA bands corresponding to the shRNA pool insert (135 bp) and plasmid backbone (~8.7 kbp) and place them in a microcentrifuge tube

9 Proceed with gel extraction using the QIAEX II Gel Extraction Kit following the manual eluting in 20 µL EB; quantify the DNA concentration using NanoDrop

10 Setup the desired assembly alongside control reactions following the table below:

	A	B	C	D	E
	Reagent	Reaction (µL)	BB only CTRL (µL)	No Assembly CTRL (µL)	Final conc.
	Gel extracted shRNA fragment (135bp)	Variable	0	0	14.34ng (172.4 fmol)
	Gel extracted iPS2-seq backbone (8715bp)	Variable	Variable	Variable	50ng (9.32 fmol)
	NEBuilder HiFi DNA Assembly Master Mix (2X)	10	10	0	1X
	Nuclease-free water	To 20	To 20	To 20	
	Total volume	20	20	20	

Table SP1.7. Pooled cloning step 1 assembly mix

10.1 Incubate the reactions at 50 °C for 15 min, then transfer to ice
Note: avoid longer incubation times as they can increase background

11 Transform 2 µL of assembly or control reactions in 50 µL NEB 5-alpha competent *E. coli* (high-efficiency), following the manufacturer's instructions

12 Plate transformed bacteria

12.1 Pre-warm LB-agar plates containing 100 µg/mL ampicillin to 37 °C

12.2 To facilitate estimation of colony number and cloning efficiency, plate 10% of transformed bacteria on one dish, and the remaining 90% on a second dish

12.3 Incubate the plates at 37 °C overnight

13 Estimate the number of bacterial clones by counting the colonies in the plate containing 10% of the transformation of the desired assembly (>200 are to be expected)

Notes:
The number of colonies in the matching backbone-only control and no assembly control plates should be at least 10-fold and 100-fold lower; it is advisable to repeat the procedure should a substantially higher background be observed
Bacterial plates can be stored at 4 °C for up to one week to allow for the subsequent optional quality controls

14 **OPTIONAL.** Confirm cloning efficiency by bacterial colony PCR

14.1 (a) For each colony to be screened, aliquot in PCR plates a 15 µL PCR mix based on nuclease-free water and containing: iPS2-seq_step1QC_F Primer (200 nM), iPS2seq_seq_QC_R Primer (200 nM), dNTPs (200 µM), MgCl2 (2 mM), Colorless GoTaq Flexi buffer (1X), and GoTaq DNA polymerase (0.025 U/µL)

14.2 (b) Pick an individual bacterial colony from the 10% plate, mix it by pipetting in 10 µL nuclease-free water aliquoted in a PCR plate, and transfer 5 µL of the suspension to the PCR mix for a final volume of 20 µL. The remaining 5 µL can be transferred to a PCR plate containing LB with 100 µg/mL ampicillin and stored at 4 °C for later use. Repeat for all colonies

14.3 (c) Place the PCR plate in a thermocycler and run following cycling conditions: (1) 95 °C for 5 min; (2) 95 °C for 30 sec; (3) 60 °C for 30 sec; (4) 72 °C for 1 min; (5) go to step 2 for 34 times; (6) 72 °C for 2 min; (7) hold at 10 °C

14.4 (d) Aliquot 5 µL of each colony PCR in a new plate, add 1 µL of 6X DNA loading dye, and perform an electrophoretic run on a 1.5% (w/v) agarose gel in TBE
Note: clones with the correct insert deliver a PCR product of 543 bp.

15 **OPTIONAL.** Confirm cloning accuracy by Sanger sequencing

15.1 Aliquot 5 µL of each colony PCR in a new plate and add 1.5 µL of a mastermix containing 0.5 µL of exonuclease 1 (20 U/µL) and 1 µL of shrimp alkaline phosphatase (1 U/µL)

15.2 Incubate the reaction at 37 °C for 30 min, followed by heat inactivation for 15 min at 80 °C

15.3 Add 5 µL of iPS2-seq_QC_R primer (10 µM) to each well, and run Sanger sequencing.
Note: alignment of each electropherogram to the common sequence 5′-TTTTTTTGTCGACATTTAAATGGCGCGCCNNNNNNNNNNNNNNGTAGCTCGCTGATCAGC

-3′
simplifies theextraction of the BC (last 8 Ns); the electropherogram can then be aligned to the expected oligo based on the BC, to verify the correct sequence of the associated shRNA

16 Collect all bacterial colonies from the 90% plate by washing and scraping the surface of LB-agar plates with 5 mL of LB broth with 100 µg/mL ampicillin

17 Inoculate the bacterial suspension in a clean Erlenmeyer flask with 45 mL LB broth containing 100 µg/mL ampicillin, and grow the bacterial cultures at 37 °C at 225 rpm for 16 h
Note: inoculate bacteria in a sterile environment and use a sterile flask to minimize the chances of contamination. Do not grow bacteria for more than 16 h

18 Isolate the intermediate plasmid pool from bacterial cultures using the QIAGEN plasmid midiprep kit following the manual, and quantify the DNA concentration using NanoDrop

Pooled cloning step 2: obtain the final plasmid pool

19 This section outlines the generation of the final plasmid pool by re-inserting the CAG-OPTtetR into the intermediate plasmid pool.

20 Prepare the following restriction digestion reaction of the empty pAAV-Puro_siKD2.0 plasmid with SalI and MluI to extract the CAG-OPTtetR insert:

	A	B	C	D
	Reagent	Digestion reaction (µL)	Undigested control (µL)	Final conc.
	pAAV-Puro_siKD2.0 plasmid	Variable	Variable	5µg
	FastDigest Green Buffer (10X)	15	1.5	1X
	FastDigest MluI	5	0	
	FastDigest SalI	5	0	
	Nuclease-free water	Top up to 100	Top up to 10	
	Total volume	100	10	

Table SP1.8. pAAV-Puro_siKD2.0 second restriction digestion mix

21 Prepare the following restriction digestion reaction of the intermediate plasmid pool from step 15 of the previous section with SalI and AscI to obtain the backbone:

	A	B	C	D
	Reagent	Digestion reaction (µL)	Undigested control (µL)	Final conc.
	Intermediate iPS2-seq plasmid	Variable	Variable	5µg
	FastDigest Green Buffer (10X)	15	1.5	1X
	FastDigest SalI	5	0	
	FastDigest SgsI (AscI)	5	0	
	FastAP (1 U/µL)	5	0.5	0.03 U/µL
	Nuclease-free water	To 150	To 15	
	Total volume	150	15	

Table SP1.9. Intermediate plasmid pool restriction digestion mix

22 Incubate at 37 °C overnight (~16 h), followed by 10 min at 80 °C for heat inactivation.

23 Perform electrophoretic separation of all fragments using 0.8% (w/v) agarose gel in TBE supplemented with SYBR Safe. Figure SP1.4 shows an exemplary electrophoresis.

Note: follow the same general recommendations described for step 7 of the previous section, paying particular attention to separating the backbone from the undigested plasmid.

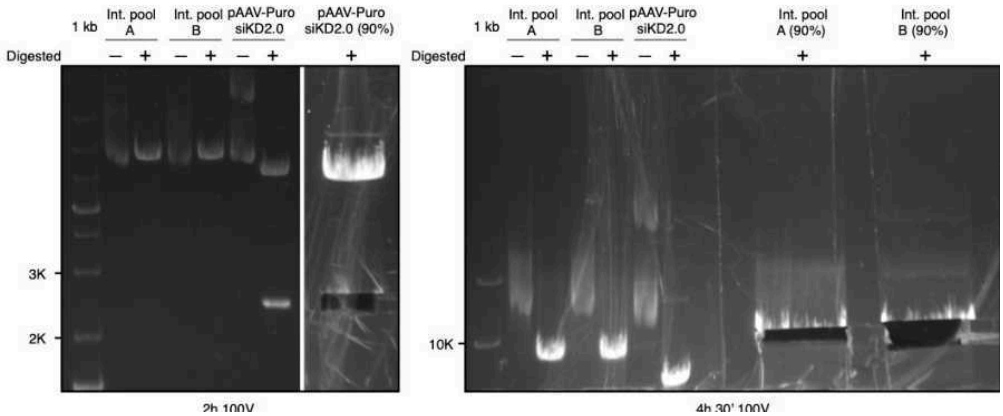


Figure SP1.4. Electrophoretic separation and gel extraction during cloning step 2. Separation of the CAG-OPTtetR insert and two digested intermediate plasmid pool backbones.

24 Use a clean scalpel to excise the ~8.8 kbp backbone from the digested intermediate plasmid pool and the ~2.4 kbp CAG-OPTtetR insert from pAAV-Puro_siKD2.0.

25 Proceed with gel extraction using the QIAEX II Gel Extraction Kit following the manual and eluting in 20 µL EB; quantify the DNA concentrations using NanoDrop.

26 Set up desired ligation alongside a control reaction following the table below:

	A	B	C	D
	Reagent	Reaction (µL)	BB only CTRL (µL)	Final conc.
	Gel extracted CAG-OPTtetR fragment (2392bp)	Variable	0	81.59ng (55.38 fmol)
	Gel extracted intermediate iPS2-seq pool (8795bp)	Variable	1	100ng (18.46 fmol)
	Rapid Ligation Buffer (5X)	4	4	1X
	T4 DNA Ligase (5 U/µL)	1	1	0.25 U/µL
	Nuclease-free water	To 20	To 20	
	Total volume	20	20	

Table SP1.10. Pooled cloning step 2 ligation mix

26.1 Incubate the reaction at RT for 1 h.

27 Transform 5 µL of ligation and control reaction in 50 µL NEB 5-alpha competent E. coli (high-efficiency), following the manufacturer's instructions.

28 Plate and grow overnight transformed bacteria according to the general recommendations described for step 9 of the previous section, particularly regarding splitting the bacteria on two plates (10% and 90%).

29 Estimate the number of bacterial clones by counting the colonies in the plate containing 10% of the transformation of the desired assembly (3e20 are to be expected).
Notes:
(a) the number of colonies in the backbone-only control should be at least 5-fold lower.
(b) bacterial plates can be stored at 4 °C for up to one week.

30 **OPTIONAL**. Confirm cloning efficiency by bacterial colony PCR

30.1 Perform colony PCRs and electrophoresis as described for step 14 of the previous section, except using primers iPS2-seq*step2QC*F and iPS2-seq*QC*R
Notes:
i. Clones with the correct insert deliver a PCR product of 312 bp; intermediate plasmids are not amplified, while the parental pAAV-Puro_siKD2.0 plasmid gives a 295 bp band (Figure S1C)
ii. Sanger sequencing can also be performed as described for step 15 of the previous section, to confirm the distribution of shRNA BCs; however, at this final stage, the shRNA and barcode cannot be co-amplified

31 Collect all bacterial colonies and inoculate them as described for steps 16-17 of the previous section. Be extra careful to avoid contamination at this stage.

32 Isolate the final plasmid pool from bacterial cultures using the QIAGEN plasmid midiprep kit with endotoxin-free buffers, and quantify the DNA concentration using Qubit.

33 Perform a diagnostic restriction digestion using EcoRI, MluI, and SwaI: the molecular weight of the expected bands is 5669 bp and 5518 bp
Note: additional bands at 2002 bp or 4343 bp would indicate contamination with intermediate plasmids or pAAV-Puro_siKD2.0, respectively. These contaminants can be removed by digestion with SwaI or BglII and MluI, respectively, followed by re-transformation

OPTIONAL: prepare NGS libraries for quality control of plasmid pools

34 This section outlines the optional generation of Illumina-compatible next-generation sequencing (NGS) libraries to determine the composition of intermediate and final plasmid pools (Figure SP1.5A, Figure SP1.6A). Data analysis relies on catcheR*step1QC and catcheR*step2QC, respectively, as described in Supplemental Protocol 4 - Pooled cloning step 1 QC and Pooled cloning step 2 and hiPSC genome editing QC.

QC n1: Intermediate plasmid pool QC

35 **PCR #1:** amplify the shRNA and UCI-BC region from the intermediate plasmid pool while adding the TruSeq read 2 adapter, a diversity index (DI, a random 12 bp sequence that ensures adequate clustering during NGS and can be leveraged to partially demultiplex reads to eliminate duplicates from the subsequent PCR), and partial TruSeq read 1 adapter

35.1 (a) Prepare a PCR reaction according to the following calculations:

	A	B	C
	Reagent	Volume (µL)	Final conc.
	Intermediate iPS2-seq plasmid pool	Variable	150 ng
	ThermoPol Reaction Buffer (10X)	2.5	1X
	Deoxynucleotide (dNTP) Solution Mix (10mM)	0.5	200µM
	siKDbc_FW Primer (10µM)	1.5	0.5µM
	siKDbc_REV Primer (10µM)	1.5	0.5µM
	MgSO4 (100mM)	1	4mM
	Deep Vent DNA Polymerase (2 U/µL)	0.25	0.02 U/µL
	Nuclease-free water	To 25	
	Total volume	25	

Table SP1.11. Intermediate plasmid pool QC, PCR #1 mix

35.2 (b) Place the PCR tube in a thermocycler and run the following cycling conditions:

	A	B	C	D
	Steps	Temperature	Time	Cycles
	Initial Denaturation	95°C	2 min	1
	Denaturation	95°C	30 sec	3
	Annealing	54-59°C (increase by 2.5°C each cycle)	15 sec	
	Extension	72°C	30 sec	
	Denaturation	95°C	30 sec	9
	Annealing	63°C	15 sec	
	Extension	72°C	30 sec	
	Final extension	72°C	5 min	1
	Hold (rapid ramp down)	4°C	Hold	

Table SP1.12. Cycling conditions for PCR #1, intermediate plasmid pool QC

Note: a "touch-up" strategy minimizes the annealing of shRNA hairpins during the second part of the reaction: once a sufficient number of molecules have integrated the TruSeq adapters, they anneal to the primers at higher temperatures

35.3 (c) Add 25 µL of nuclease-free water to the reaction and perform a double-sided size selection using 0.6X & 1.2X volumes of SPRIselect beads, following the manufacturer's instructions and eluting in 20 µL of elution buffer

35.4 (d) Run 2 µL of the undiluted PCR product on a TapeStation High Sensitivity D1000 Screen Tape assay, following the manual. Expect a single peak of ~203 bp. An exemplary result is reported in (Figure SP1.5B)

35.5 (e) Quantify the DNA concentration in the range 100-400 bp, and dilute it to 1 ng/µL

36 **PCR #2:** prepare dual-indexed Illumina TruSeq sequencing libraries

36.1 (a) Prepare a PCR reaction according to the following calculations:

	A	B	C
	Reagent	Volume (µL)	Final conc.
	PCR#1 product (1 ng/ul)	2	2 ng
	ThermoPol Reaction Buffer (10X)	2	1X
	Deoxynucleotide (dNTP) Solution Mix (10mM)	0.4	200µM
	Dual Index primers Kit TT Set A	4,00	
	MgSO4 (100mM)	0.8	4mM
	Deep Vent DNA Polymerase (2 U/µL)	0.2	0.02 U/µL
	Nuclease-free water	10.6	
	Total volume	20	

Table SP1.13. Intermediate plasmid pool QC, PCR #2 mix

36.2 (b) Place the PCR tube in a thermocycler and run the following cycling conditions:

	A	B	C	D
	Steps	Temperature	Time	Cycles
	Initial Denaturation	95°C	2 min	1
	Denaturation	95°C	30 sec	3
	Annealing	45-47°C (increase by 2.5°C each cycle)	15 sec	
	Extension	72°C	45 sec	
	Denaturation	95°C	30 sec	9
	Annealing	63°C	15 sec	
	Extension	72°C	30 sec	
	Final extension	72°C	5 min	1
	Hold (rapid ramp down)	4°C	Hold	

Table SP1.14. Cycling conditions for PCR #2, intermediate plasmid pool QC

36.3 Add 30 µL of nuclease-free water to the reaction and perform a double-sided size selection using 0.65X & 0.85X volumes of SPRIselect beads, following the manufacturer's instructions and eluting in 20 µL of elution buffer

36.4 Run 2 µL of the undiluted PCR product on a TapeStation High Sensitivity D1000 Screen Tape assay, following the manual. Expect a single peak of ~287 bp

37

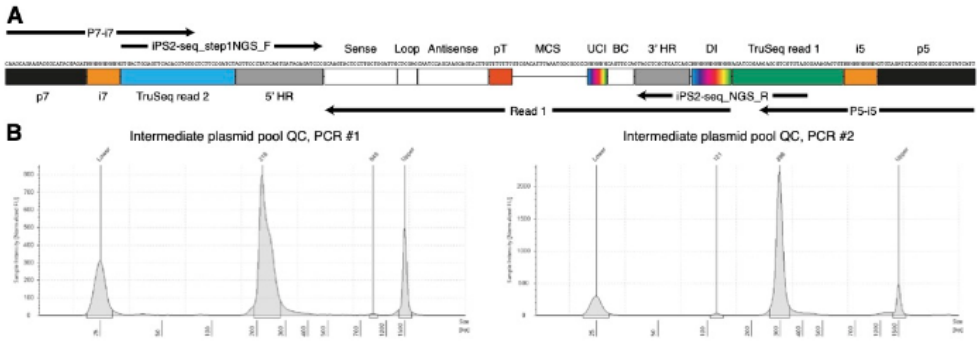


Figure SP1.5. NGS library generation from intermediate plasmid pools
(A) Library structure and primer sequences used for PCR #1 and PCR #2; read 1 is also indicated.
(B) Exemplary TapeStation High Sensitivity D1000 Screen Tape assays of size-selected DNA from PCRs #1 and #2

QC n2: Final plasmid pool QC

38 **PCR #1:** amplify UCI-BCs from the final plasmid pool, adding TruSeq adapters and DIs

39

	A	B	C
	Reagent	Volume (µL)	Final conc.
	Final iPS2-seq plasmid pool	Variable	150 ng
	Q5 Reaction Buffer (5X)	5	1X
	Deoxynucleotide (dNTP) Solution Mix (10mM)	0.5	200µM
	iPS2-seq_step2NGS_Primer (10 µM)	1.25	0.5µM
	iPS2-seq_NGS_R Primer (10 µM)	1.25	0.5µM

	A	B	C
	Q5 Hot Start High-Fidelity DNA Polymerase (2 U/µL)	0.25	0.02 U/µL
	Nuclease-free water	To 25	
	Total volume	25	

Table SP1.15. Final plasmid pool QC, PCR #1 mix

40

	A	B	C	D
	Steps	Temperature	Time	Cycles
	Initial Denaturation	98°C	30 sec	1
	Denaturation	98°C	10 sec	9 cycles
	Annealing	67°C	30 sec	
	Extension	72°C	20 sec	
	Final extension	72°C	2 min	1
	Hold (rapid ramp down)	4°C	Hold	

Table SP1.16. Cycling conditions for PCR #1

41 Purify, quantify, and dilute the PCR product following steps 35.2, 35.3, and 35.4 of the previous section, except for using 0.6X & 1.5X volumes of SPRIselect beads for double-sided size selection. Expect a single peak of ~144 bp (Figure SP1.6B)

42 **PCR #2:** prepare dual indexed Illumina TruSeq sequencing libraries

43

	A	B	C
	Reagent	Volume (µL)	Final conc.
	PCR#1 product (1 ng/ul)	1	1 ng
	Dual Index primers Kit TT Set A	4,00	

	A	B	C
	NEBNext high fidelity 2X PCR master mix	10	1X
	Nuclease-free water	5,00	
	Total volume	20	

Table SP1.17. Final plasmid pool QC, PCR #2 mix

44

	A	B	C	D
	Steps	Temperature	Time	Cycles
	Initial Denaturation	98°C	30 sec	1
	Denaturation	98°C	10 sec	8 cycles
	Annealing	58°C	20 sec	
	Extension	72°C	20 sec	
	Final extension	72°C	30 sec	1
	Hold (rapid ramp down)	4°C	Hold	

Table SP1.18. Cycling conditions for PCR #2

45 Purify and quantify the PCR product following steps 36.3 and 36.4 of the previous section, except for using 0.65X & 1.5X volumes of SPRIselect beads for double-sided size selection. Expect a single peak of ~228 bp (Figure SP1.6B)

46 Perform NGS using NextSeq 1000/2000 and 50 cycles reagents, with the following settings: read 1 - 68 cycles; index 1 - 10 cycles; index 2 - 10 cycles (>10,000 reads/shRNA)

47

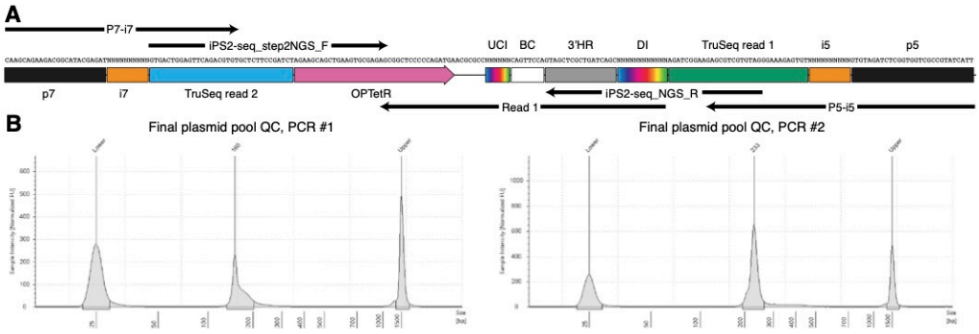


Figure SP1.6. NGS library generation from final plasmid pools
(A) Library structure and primer sequences used for PCR #1 and PCR #2; read 1 is also indicated.
(B) Exemplary TapeStation High Sensitivity D1000 Screen Tape assays of size-selected DNA from PCRs #1 and #2.

Genome editing: obtain hPSC clonal pools

48 This section outlines the generation of hPSCs genome edited at the *AAVS1* locus with the final iPS2-seq plasmid pool, and the optional genotyping of individual clones as quality control
(Figure SP1.1).

49 Preparatory steps:

49.1 Determine the number of nucleofections required to obtain the total target number of hPSC clones, based on the number of shRNAs, the target number of clones per shRNA, and the efficiency of genome editing
Notes:
i. The protocol describes plasmid delivery by nucleofection as this is most efficient, but if the necessary instrumentation is not available, plasmids can instead be transfected as described in the STAR METHODS
ii. While our standard protocol uses ZFNs to target the *AAVS1* locus, alternative programmable nucleases such as TALENs and CRISPR/Cas9 can also be used.
We opted for ZFNs because we previously demonstrated >95% on-target integration efficiency in hPSCs. That said, our targeting vector is fully compatible with other systems, including TALENs and widely used CRISPR/Cas9 reagents (e.g., Addgene #59025, #59026, #129726). Besides plasmid delivery, in side-by-side comparisons, we observed comparable integration efficiencies using 200,000 cells per nucleocuvette strip co-electroporated with the iPS2-seq vector and either 1 µg ZFN plasmids or Cas9 RNPs (1 µg Cas9 protein + 200 ng sgRNA)

iii. Perform a pilot genome editing experiment to empirically determine the efficiency of genome editing in an hPSC line used for iPS2-seq for the first time

49.2 Prepare the following plasmids using the QIAGEN plasmid midiprep kit with endotoxin-free buffers, diluting them to 1 µg/µL (2.5 µg/nucleofection):

i. **Pool of pAAV-Puro_siKD2.0** with the barcoded shRNAs (step 32 from Pooled cloning step 2: obtain the final plasmid pool)
ii. **pAAV-Neo_CAG**
iii. **pZFN-AAVS1_ELD**
iv. **pZFN-AAVS1_KKR**

49.3 Prepare Geltrex-coated 24-well culture dishes (1 well per nucleofection)

49.4 Culture hPSCs in Essential 8 at ~40% confluency (1 × 10^6/nucleofection)

50 16 h before nucleofection, refresh the hPSC culture media to Essential 8 supplemented with 2 µM Thiazovivin. Do not add antibiotics until step 11

51 Prepare hPSC Essential 8 supplemented with CEPT [50 nM Chroman 1; 5 µM Emricasan; 0.1% Polyamine Supplement; 0.7 µM trans-ISRIB] (1.5 mL/nucleofection)

52 Remove the coating solution from the pre-coated culture dish, add 1 mL of Essential 8 with CEPT per well, and place the dish in the incubator to acclimatize to 37 °C

53 Prepare a mix of equal mass of the four plasmids from step 1b (10 µg/transfection) and nucleofection buffers (18 µL P3 Supplement 26 82 µL Nucleofector Solution per reaction)

54 Obtain a single cell suspension by treating hPSCs with StemPro Accutase, following the manufacturer's instructions, aliquot 1 × 10^6^ hPSCs/nucleofection in a conical tube, and pellet the cells at 100 g for 5 min

55 Remove the supernatant, gently resuspend the pellet with the nucleofection mix from step 5, transfer 110 µL/nucleofection in a Nucleocuvette Vessel (avoiding bubbles), place the cuvette into the 4D-Nucleofector System and run program CA137

56 Add 500 µL of Essential 8 with CEPT to each cuvette and use the suction pipette to gently transfer the cell suspension to the 24-well plate from step 4. Place in the incubator

57 24 h post nucleofection, refresh media with Essential 8, and prepare Geltrex-coated 100 mm dishes (1 per nucleofection)

58 48 h post nucleofection, dissociate cells as small clumps with 0.5 mM EDTA in DPBS, and replate in Essential 8 with 2 µM Thiazovivin (using one 100 mm dish per well of 24-wp)

59 Refresh media daily with Essential 8. Antibiotics can be added from now on

60 When hPSCs are ~50% confluent (~5-6 days post-nucleofection), perform dual positive drug selection for 4 days by supplementing Essential 8 with 0.5 µg/mL puromycin and 25 µg/mL Geneticin, refreshing media daily; include 2 µM Thiazovivin for the first 2 days
Note: to optimize the procedure for a specific hPSC line, perform a kill curve with 0.25-2 µg/mL puromycin and 12.5-100 µg/mL Geneticin to identify the minimal concentration of each drug that in combination, eliminates all unedited hPSCs within 72 h

61 When hPSCs colonies are ready for passaging (~1-2 mm in diameter), count the number of colonies (clones), dissociate cells as small clumps with 0.5 mM EDTA in DPBS, and replate in Essential 8 with 2 µM Thiazovivin at a density of ~1 clones per cm^2^

62 From 24 h post passaging until 5 days post passaging, refresh media daily with Essential 8 supplemented with 200 nM fialuridine (FIAU) to perform the negative selection

63 hPSC clonal pools are now ready for screening experiments and/or cryopreservation

63.1 **Notes:**
(a) Minimize the number of passages before the induction of gene knockdown, to minimize the expansion of hPSC clones with a genetic/epigenetic growth advantage

63.2 (b) To account for the possible emergence of hPSC clones with neuroectodermal bias (“iPSC-neuro” clones), ensure sufficient clonal representation per perturbation. iPS2-seq enables robust retrospective control for this source of variability; however, it may still be helpful to screen for expression of early neuroectodermal markers such as ZIC1 or UNC5D. If a given pool contains an overrepresentation of iPSC-neuro clones, consider repeating the targeting or depleting biased clones using, for example, a UNC5D antibody for FACS or MACS

63.3 (c) For hPSC growth and differentiation protocols involving media containing FBS and/or BSA, batch-test these reagents for the absence of tetracycline contamination (i.e., by generating a clonal line expressing a validated inducible shRNA, to confirm that shRNA expression is not leaky in the absence of exogenously added tetracycline)

63.4 (d) To induce knockdown, add 1 ng/mL tetracycline, refreshing the media daily or every 48 h, when compatible for the specific differentiation protocol

hiPSC pool QC

64 This procedure aims to determine clonal diversity and shRNA representation in the genome-edited hiPSC pool by generating Illumina-compatible NGS libraries of UCI-BCs that can be analyzed using *catcheR_step2QC* as described in Supplemental Protocol 4 - Pooled cloning step 2 and hiPSC genome editing QC.

65 After hPSC passaging, collect the cells from a confluent 6-well plate, and pellet the cells at 100 g for 5 min.
Note: Cell pellets can be frozen at -20 °C if needed.

66 Obtain purified genomic DNA (gDNA) of the genome-edited hiPSCs with the Monarch Spin gDNA Extraction Kit.

67 Elute samples with 50-100 µL of elution buffer and assess DNA purity and quantity by NanoDrop and Qubit broad range dsDNA.

68 Use 500 ng of gDNA per PCR reaction, performing as many reactions as needed to achieve coverage of at least 100-fold the estimated number of clones.
Note: One nanogram of human gDNA contains approximately 152 genome equivalents; therefore, 500 ng corresponds to ~76,000 potential UCI-BC templates at the AAVS1 locus.

69 Using the same primers as the Final plasmid pool QC, perform a PCR to amplify UCI-BCs from the genomic DNA, adding TruSeq adapters and DI, according to the adjusted calculations and cycling conditions:

	A	B	C
	Reagent	Volume (µL)	Final conc.
	gDNA iPS2-seq pool	Variable	500 ng
	Q5 Reaction Buffer (5X)	10	1X
	Deoxynucleotide (dNTP) Solution Mix (10mM)	1	200µM
	iPS2-seq_step2NGS_Primer (10 µM)	2.5	0.5µM
	iPS2-seq_NGS_R Primer (10 µM)	2.5	0.5µM

	A	B	C
	Q5 Hot Start High-Fidelity DNA Polymerase (2 U/µL)	1,00	0.04 U/µL
	Nuclease-free water	To 50	
	Total volume	50	

Table SP1.19. PS2-seq transfected cells QC, PCR #1 mix

70

	A	B	C	D
	Steps	Temperature	Time	Cycles
	Initial Denaturation	98°C	30 sec	1
	Denaturation	98°C	10 sec	26 cycles
	Annealing	64°C	30 sec	
	Extension	72°C	20 sec	
	Final extension	72°C	2 min	1
	Hold (rapid ramp down)	10°C	Hold	

Table SP1.20. Cycling conditions for PCR #1

71 Purify the expected ~144 bp amplification product and purify it using SPRIselect beads with a double side selection using 0.6X & 1.5X bead ratios to remove primer dimers and small fragments.

72 Dilute the PCR product to 0.1 ng/µL in EB buffer and perform indexing PCR according to Table SP1.17 and Table SP1.18.

73 An indexed library product of ~228 bp is expected. Purify the reaction using a double-sided SPRIselect bead cleanup (0.65X & 1.5X ratios).

74 Run 2 µL of the undiluted and 1:10 final library on a TapeStation High Sensitivity D1000 Screen Tape assay, following the manual.
Note: This is the same amplicon described for the final plasmid QC (Figure SP1.6B), but adjusted to work on genomic DNA material.

75 Perform NGS using MiSeq or NextSeq 1000/2000 with the following settings: read 1 - 68 cycles; index 1 - 10 cycles; index 2 - 10 cycles (>10,000 reads/clone).

OPTIONAL: hiPSC clone genotyping

76 This optional procedure can be used to assess the efficiency of biallelic *AAVS1* genome editing in a subset of hPSC clones, or to isolate homozygously edited clones for validation studies. Genotyping leverages various genomic PCRs to determine on-site integration, copy number, and random integrations (Figure SP1.7)

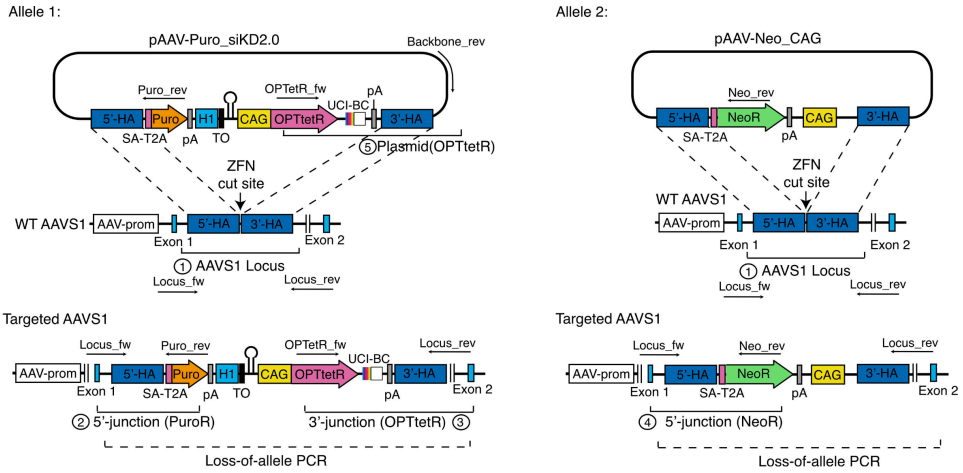


Figure SP1.7. iPS2-seq clone genotyping strategy
Schematic of the genotyping strategies to monitor integration of *pAAV-Puro_siKD2.0* in allele 1 and *pAAV-Neo_CAG* in allele 2. Primer binding sites and directions are shown on the plasmid or integrated cassettes. PCR types and products are numbered as in Table SP1.23.

77 After step 13 of Genome editing: obtain hPSC clonal pools, pick and expand individual hPSC clones.
Note: treat clones with 200 nM FIAU to test for random integrations.

78 Prepare a culture dish with 2-5 × 105 cells, and extract gDNA using the Monarch Spin gDNA Extraction Kit, diluting to a concentration of 50-200 ng/µL

79 Perform PCR reactions according to the following calculations and cycling conditions, employing the primer pairs listed in **Table SP1.23**. Include the following controls: wild type hPSCs; pAAV-Puro_siKD2.0; no template

80

	A	B	C
	Reagent	one reaction (µL)	Final conc.
	Genomic DNA	1	50-200ng
	LongAmp Taq Reaction Buffer (5X)	2	1X
	Deoxynucleotide (dNTP) Solution Mix (10mM)	0,3	300µM
	Forward Primer (10µM)	0,5	0.5µM
	Forward Primer (10µM)	0,5	0.5µM
	DMSO	0,2	2\%
	LongAmp Taq DNA Polymerase (2.5U/µL)	0,4	0.1 U/µL
	Nuclease-free water	To 10	
	Total volume	10	

Table SP1.21. Genotyping mix

81

	A	B	C	D
	Steps	Temperature	Time	Cycles
	Initial Denaturation	94°C	5 min	1
	Denaturation	94°C	15 sec	35 cycles
	Annealing	Tab step SP1.23	30 sec	
	Extension	65°C	Tab step SP1.23	
	Final extension	65°C	5 min	1
	Hold (rapid ramp down)	4°C	Hold	

Table SP1.22. Cycling conditions for genotyping

82 Run half the PCR products on 0.8% (w/v) agarose gels and determine the genotype based on Table SP1.24

83

PCR type NB	PCR type	Primer	Primer name	Primer location	Amplicon wild type[a]	Amplicon target[b]	Amplicon plasmid[c]	Annealing temp.[d]	Extension time[d]
1	AAVS1 locus	FW	Locus_fw	Genomic, 5' to 5' HA	1692 bp	No band	no band	65 °C	1' 30"
		REV	Locus_rev	Genomic, 3' to 3' HA					
2	5' junction (PuroR)	FW	Locus_fw	Genomic, 5' to 5' HA	no band	991 bp	no band	65 °C	1'
		REV	Puro_rev	Puro resistance					
3	3' junction (OPTetR)	FW	OPTtetR_fw	OPTtetR cDNA	no band	1463 bp	no band	60 °C	1' 30"
		REV	Locus_rev	Genomic, 3' to 3' HA					
4	5' junction (NeoR)	FW	Locus_fw	Genomic, 5' to 5' HA	no band	1032 bp	no band	60 °C	1'
		REV	Neo_rev	Neo resistance					
5	Plasmid (OPTtetR)	FW	OPTtetR_fw	OPTtetR cDNA	no band	no band	1819 bp	60 °C	2'
		REV	Backbone_rev	Backbone, 3' to 3' HA					

Table SP1.23. Genotyping strategies for iPS2-seq clone genotyping
aResult of PCR on wild type *AAVS1* allele
bResult of PCR on iPS2-seq-targeted *AAVS1* allele
cResult of PCR on pAAV-Puro_siKD2.0 (positive control for off-target plasmid integration)
dVariable parameter in PCR protocol (Table SP1.22)

84

AAVS1 locus	5' junction (PuroR)	3' junction (OPTtetR)	5' junction (NeoR)	Plasmid (OPTtetR)	*AAVS1* genotype	FIAU sensitive?	Number of shRNAs
-	+	+	+	-	Compound heterozygous	No	1
-	+	+	+	+	Compound heterozygous	No	1
+	+	+	-	-	Heterozygous iPS2-seq	No	1
+	+	+	-	+	Heterozygous iPS2-seq	No	1
-	+	+	+	+	Compound heterozygous	Yes	>1
+	+	+	-	+	Heterozygous iPS2-seq	Yes	>1
-	+	+	-	-	Homozygous iPS2-seq	No	2
-	+	+	-	+	Homozygous iPS2-seq	No	2
-	+	+	-	+	Homozygous iPS2-seq	Yes	>2
+	-	-	-	-	Wild type	No	0
+/-	-	+/-	+/-	-	Incorrect iPS2-seq targeting	No	0
+/-	+/-	-	+/-	-	Incorrect iPS2-seq targeting	No	0
+/-	-	+/-	+/-	+	Incorrect iPS2-seq targeting	No	0
+/-	+/-	-	+/-	+	Incorrect iPS2-seq targeting	No	0
+/-	-	+/-	+/-	+	Incorrect iPS2-seq targeting	Yes	1 or more
+/-	+/-	-	+/-	+	Incorrect iPS2-seq targeting	Yes	1 or more

Table SP1.24. Interpretation of iPS2-seq clone genotyping

Protocol 2: iPS2-sci-seq library preparation

85 Decoding single-cell transcriptional perturbomes in iPS2-seq relies on detecting the shRNA-associated UCI-BC within the 3′ UTR of the OPTtetR transgene (Figure 1A). This

can be in principle achieved with any scRNA-seq method that tags and counts the 3′ ends of mRNAs, usually as a result of pA-primed RT.
iPS2-sci-seq is a home-brew adaptation of the 2-level indexing sci-RNA-seq protocol19, that relies on 2 rounds of transcriptome barcoding through sequential pool-split of defined cell numbers so that the resulting combinatorial barcode in each cell is statistically most likely unique. This approach requires limited custom reagents and supports the analysis of 1,000-10,000 cells per experiment, a scale compatible with the everyday needs of most laboratories. Minor adaptations of this protocol that require additional reagents can enable the 3-level indexing variant of sci-RNA-seq (sci-RNA-seq3), which supports analysis of over ~1 million cells per experiment. Since transcript counts in sci-RNA-seq are notoriously low and zero-inflated (i.e., genes may not be counted at least once in cells that express them), it is key to incorporate a strategy that explicitly enriches for sequences containing the UCI-BC, thereby detecting it with higher sensitivity and specificity. Therefore, in the iPS2-sci-seq protocol, UCI-BCs are enriched at both transcriptome barcoding steps: first during RT, using a set of target-specific primers that share RT barcodes with standard pT primers, well by well, and secondly during PCR, using a set of target-specific primers that share i7 barcodes with standard P7 primers, once again well by well. The resulting multiplexed NGS library pools contain both single-cell transcriptomes and UCI-BCs, which share library structure and carry identical combinatorial indexes, cell by cell.

This Supplemental Protocol describes the following procedures:
1. Reagent setup
2. Nuclei preparation
3. Reverse transcription
4. Nuclei sorting
5. Second-strand synthesis
6. Tagmentation
7. Indexing PCR and NGS

86 **Reagent setup:** Prepare the RT and PCR oligos using nuclease-free water:

86.1 Obtain master stocks of the 96 indexed iPS2-sci-seq_pT primers (250 µM) and iPS2-sci-seq_tetR primers (100 µM). Use the RT_indexes indicated in Table SP2.1

86.2 Prepare a working stock of multiplexed RT primers by diluting in the same plate iPS2-sci-seq_pT primers (25 µM) and iPS2-sci-seq_tetR primers (2.5 µM), matching well positions so that pooled primers share the same RT index

86.3 Obtain master stocks of indexed iPS2-sci-seq_P5, iPS2-sci-seq_P7, and iPS2-sci-seq_P7A (all at 100 µM). Use the i5 and i7 indexes indicated in Table SP2.1.

Note: it is not necessary to obtain 96 indexed oligos of each type, as i5 and i7 can be combined to generate unique combinations.

86.4 Prepare a working stock of iPS2-sci-seq_P5 (10 µM).

86.5 Prepare a working stock of multiplexed P7 primers by diluting in the same plate iPS2-sci-seq*P7 (10 µM) and iPS2-sci-seqP7*A (5 µM), matching well positions so that pooled primers share the same i7 index.

87 Prepare the following buffers using nuclease-free water, and chill them on ice:

87.1 Nuclei buffer (store at 4 °C for up to 1 month, 5 mL/sample): 10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2.

87.2 Nuclei lysis buffer (made fresh, 1 mL/sample): nuclei buffer with 0.1-0.3% IGEPAL CA-630, 1% SUPERase In RNase Inhibitor (20 U/µL) and 1% BSA (20 mg/mL).
Note: the optimal concentration of IGEPAL CA-630 for a given cell type should be determined in a pilot experiment. For reference, 0.1% works for hPSCs, but not for hPSC-CMs which are lysed appropriately only at 0.3%.

87.3 Nuclei suspension buffer (make fresh, 3 mL/sample): nuclei buffer with 1% SUPERase In RNase Inhibitor (20 U/µL) and 1% BSA (20 mg/mL).

87.4 4% PFA (store at 4 °C for up to 1 month, 10 mL/sample). 4 °C

87.5 DPBS without Ca2+ and Mg2+ (store at 4 °C for up to 1 month, 10 mL/sample). 4 °C

88

	Well	RT_index (5'-3')	i7 for primer (5'-3')	i7 for sample sheet (5'-3')	i5 for primer & sample sheet, forward strand (5'-3')	i5 for sample sheet, reverse strand (5'-3')		Well	RT_index (5'-3')	i7 for primer (5'-3')	i7 for sample sheet (5'-3')	i5 for primer & sample sheet, forward strand (5'-3')	i5 for sample sheet, reverse strand (5'-3')

	A01	TTCTCGCATG	CCGAATCCGA	TCGGATTCGG	CTCCATCGAG	CTCGATGGAG		E01	AGAGAAGGTT	TTCGTTCCAT	ATGGAACGAA	AGGCGAGAGC	GCTCTCGCCT
	A02	TCCTACCAGT	ATAAGCCGGA	TCCGGCTTAT	TTGGTAGTCG	CGACTACCAA		E02	CATACTCCGA	TACCTAATCA	TGATTAGGTA	TCAAGATAGT	ACTATCTTGA
	A03	GCGTTGGAGC	CCGGCGGCGA	TCGCCGCCGG	GGCCGTCAAC	GTTGACGGCC		E03	GCTAACTTGC	AAGTAATATT	AATATTACTT	TAATTGACCT	AGGTCAATTA
	A04	GATCTTACGC	GGCTTGCCAA	TTGGCAAGCC	CCTAGACGAG	CTCGTCTAGG		E04	AATCCATCTT	AGCTAAGAAT	ATTCTTAGCT	CAGCCGGCTT	AAGCCGGCTG
	A05	CTGATGGTCA	CCGCTAGCTG	CAGCTAGCGG	TCGTTAGAGC	GCTCTAACGA		E05	GGCTGAGCTC	GTCGAGGTAT	ATACCTCGAC	AGAACCGGAG	CTCCGGTTCT
	A06	CCGAGAATCC	CTTATCCTAC	GTAGGATAAG	CGTTCTATCA	TGATAGAACG		E06	CCGATTCCTG	TTATTAGTAG	CTACTAATAA	GAGATGCATG	CATGCATCTC
	A07	GCCGCAACGA	TGAGCTACTT	AAGTAGCTCA	CGGAATCTAA	TTAGATTCCG		E07	ACCGCCAACC	TGCGAAGATC	GATCTTCGCA	GATTACCGGA	TCCGGTAATC
	A08	TGAGTCTGGC	TCAGGACTTA	TAAGTCCTGA	ATGACTGATC	GATCAGTCAT		E08	TGGCCTGAAG	AACTACGGCT	AGCCGTAGTT	TCGTAACGGT	ACCGTTACGA
	A09	TGCGGACCTA	CCGCAGCCGC	GCGGCTGCGG	TCAATATCGA	TCGATATTGA		E09	AACCTCATTC	AACGGAACGC	GCGTTCCGTT	TGGCGACGGA	TCCGTCGCCA
	A10	ACCTCGTTGA	TGCGCCTGGT	ACCAGGCGCA	GTAGACCTGG	CCAGGTCTAC		E10	ATAAGGAGCA	GATGCTACGA	TCGTAGCATC	AGTCATAGCC	GGCTATGACT
	A11	ACGGAGGCGG	AATCATACGG	CCGTATGATT	TTATTGACCA	TTGGTCATAA		E11	CGAACGCCGG	ATCTGCCAAT	ATTGGCAGAT	GTCAAGTCCA	TGGACTTGAC
	A12	TAGATCTACT	CGCCAATCAA	TTGATTGGCG	TTGGTCCGTT	AACGGACCAA		E12	GGTATGCTTG	ATCGTATCAA	TTGATACGAT	ATTCGGAAGT	ACTTCCGAAT

	B01	AATTAAGACT	CAAGGCTTAG	CTAAGCCTTG	GGTACGTTAA	TTAACGTACC		F01	AACCTGCGTA	AACGCCTCTA	TAGAGGCGTT	GTCGGTAGTT	AACTACCGAC
	B02	CCATTGCGTT	GCGCTCGACG	CGTCGAGCGC	CAATGAGTCC	GGACTCATTG		F02	GGCAGACGCC	ACGGCAACCA	TGGTTGCCGT	AGGACGGACG	CGTCCGTCCT
	B03	TTATTCATTC	TCCAGCAATA	TATTGCTGGA	GATGCAGTTC	GAACTGCATC		F03	TAGCCGTCAT	CAGGCTAAGA	TCTTAGCCTG	CTCCTGGACC	GGTCCAGGAG
	B04	ATCTCCGAAC	CATGAGAACT	AGTTCTCATG	CCATCGTTCC	GGAACGATGG		F04	CCTGGAAGAG	CGCAATATCA	TGATATTGCG	TAGCCTCGTT	AACGAGGCTA
	B05	TTGACTTCAG	AACGTAATCT	AGATTACGTT	TTGAGAGAGT	ACTCTCTCAA		F05	GGAGGTTCTA	TTCGATAACC	GGTTATCGAA	GGTTGAACGT	ACGTTCAACC
	B06	GGCAGGTATT	ATTCTCCTCT	AGAGGAGAAT	ACTGAGCGAC	GTCGCTCAGT		F06	CTAGTAGTCT	AACCTCAAGA	TCTTGAGGTT	AGGTCCTCGT	ACGAGGACCT
	B07	AGAGCTATAA	TCTGCGCGTT	AACGCGCAGA	TGAGGAATCA	TGATTCCTCA		F07	ATCATCAACG	CAGGCGCCAT	ATGGCGCCTG	GGAAGTTATA	TATAACTTCC
	B08	CTAAGAGAAG	GCTCATATGC	GCATATGAGC	CCTCCGACGG	CCGTCGGAGG		F08	ACGCGAGATT	AACTATTATA	TATAATAGTT	TGGTAATCCT	AGGATTACCA
	B09	ACTCAATAGG	AGCGGTAACG	CGTTACCGCT	CATTGACGCT	AGCGTCAATG		F09	GAAGAGGCAT	AAGTTACCTA	TAGGTAACTT	AAGCTAGGTT	AACCTAGCTT
	B10	CTTGCGCCGC	AATGAATAGT	ACTATTCATT	TCGTCCTTCG	CGAAGGACGA		F10	GGTATCCGCC	CGGCAGAGGA	TCCTCTGCCG	TCCGCGGACT	AGTCCGCGGA
	B11	AATCGTAGCG	CCGTATCTGG	CCAGATACGG	TGATACTCAA	TTGAGTATCA		F11	AACTAGGCGC	GCCTCAATAA	TTATTGAGGC	TGCGGATAGT	ACTATCCGCA

	B12	GGTACTGCCT	CCTTAGTCTG	CAGACTAAGG	TTCTACCTCA	TGAGGTAGAA		F12	TCGCTAAGCA	TTAACGCCGT	ACGGCGTTAA	TGGCAGCTCG	CGAGCTGCCA
	C01	TAGAATTAAC	ACCTAGTTAG	CTAACTAGGT	TCGTCGGAAC	GTTCCGACGA		G01	TATATACTAA	CATACGATGC	GCATCGTATG	TGCTACGGTC	GACCGTAGCA
	C02	GCCATTCTCC	ATAGGAGTAC	GTACTCCTAT	ATCGAGATGA	TCATCTCGAT		G02	ACTTGCTAGA	AAGCTGACCT	AGGTCAGCTT	GCGCAATGAC	GTCATTGCGC
	C03	TGCCGGCAGA	CTACGACGAG	CTCGTCGTAG	TAGACTAGTC	GACTAGTCTA		G03	AACCATTGGA	GAGTCCTTAT	ATAAGGACTC	CTTAATCTTG	CAAGATTAAG
	C04	TTACCGAGGC	AGTCGAGTTC	GAACTCGACT	GTCGAAGCAG	CTGCTTCGAC		G04	TCGCGGTTGG	CCTACGGCAA	TTGCCGTAGG	GGAGTTGCGT	ACGCAACTCC
	C05	ATCATATTAG	TGGTCCAGTC	GACTGGACCA	AGGCGCTAGG	CCTAGCGCCT		G05	CGTAGTTACC	AATATTCGAA	TTCGAATATT	ACTCGTATCA	TGATACGAGT
	C06	TGGTCAGCCA	ATCTAAGCAA	TTGCTTAGAT	AGATGCAACT	AGTTGCATCT		G06	TCCAATCATC	TTCAAGAATC	GATTCTTGAA	GGTAATAATG	CATTATTACC
	C07	ACTATGCAAT	CGAATTCGTT	AACGAATTCG	AAGCCTACGA	TCGTAGGCTT		G07	AATCGATAAT	ATGCTCGCAA	TTGCGAGCAT	TCCTTATAGA	TCTATAAGGA
	C08	CGACGCGACT	CAGCGATAGA	TCTATCGCTG	GTAGGCAATT	AATTGCCTAC		G08	CCATTATCTA	GGAGTAAGCC	GGCTTACTCC	CCGACTCCAA	TTGGAGTCGG
	C09	GATACGGAAC	GGTCGCTATG	CATAGCGACC	TGCCAGTTGC	GCAACTGGCA		G09	TCAACGTAAG	TTATCGTATT	AATACGATAA	GCCAAGCTTG	CAAGCTTGGC
	C10	TTATCCGGAT	ATCCGTTAGC	GCTAACGGAT	CTTAGGTATC	GATACCTAAG		G10	TCTAATAGTA	AAGTCTAATA	TATTAGACTT	CATATCCTAT	ATAGGATATG
	C11	TAGAGTAATA	TCGCAATTAG	CTAATTGCGA	GAGACCTACC	GGTAGGTCTC		G11	AACCGCTGGT	CGGCTTACTA	TAGTAAGCCG	ACCTACGCCA	TGGCGTAGGT

	C12	GCAGGTCCGT	GGCTGGCTAG	CTAGCCAGCC	ATTGACCGAG	CTCGGTCAAT		G12	GATCGCTTCT	GATATGGTCT	AGACCATATC	GGAATTCAGT	ACTGAATTCC
	D01	TCGGCCTTAC	ACGGTCTTGC	GCAAGACCGT	GGAGGCGGCG	CGCCGCCTCC		H01	CTAACTAGAT	TAGTCGTCCA	TGGACGACTA	TGGCGTAGAA	TTCTACGCCA
	D02	AGAACGTCTC	GCTCCATTCG	CGAATGGAGC	CCAGTACTTG	CAAGTACTGG		H02	GCTGGAACTT	TAGCTGCTAC	GTAGCAGCTA	ATTGCGGCCA	TGGCCGCAAT
	D03	CCAGTTCCAA	ACGATAAGCG	CGCTTATCGT	GGTCTCGCCG	CGGCGAGACC		H03	AGGTTAGTTC	CTCTTCAAGC	GCTTGAAGAG	TTCAGCTTGG	CCAAGCTGAA
	D04	GGCGTTAAGG	ACCATAGCGC	GCGCTATGGT	GGCGGAGGTC	GACCTCCGCC		H04	CATTCGACGG	ATGAACGCGC	GCGCGTTCAT	CCATCTGGCA	TGCCAGATGG
	D05	ACTTAACCTT	CTCTTAGCGG	CCGCTAAGAG	TAGTTCTAGA	TCTAGAACTA		H05	CATTCAATCA	GTCGACGGAA	TTCCGTCGAC	CTTATAAGTT	AACTTATAAG
	D06	CAACCGCTAA	TGATTCAACT	AGTTGAATCA	TTGGAGTTAG	CTAACTCCAA		H06	CGGATTAGAA	ACTAATTGAG	CTCAATTAGT	GATTAGATGA	TCATCTAATC
	D07	GACCTTGATA	TATGGCCGCG	CGCGGCCATA	AGATCTTGGT	ACCAAGATCT		H07	ATCGGCTATC	CTTGCATAAT	ATTATGCAAG	TATAGGATCT	AGATCCTATA
	D08	TCTGATACCA	AGAGGTCGCA	TGCGACCTCT	GTAATGATCG	CGATCATTAC		H08	CCTTGATCGT	TCCTTACCAA	TTGGTAAGGA	AGCTTATAGG	CCTATAAGCT
	D09	GAAGATCGAG	AGGAGATTGA	TCAATCTCCT	CAGAGAGGTC	GACCTCTCTG		H09	ACGAAGTCAA	TGCAGCCTAC	GTAGGCTGCA	GTCTGCAATC	GATTGCAGAC
	D10	AGGAGCGGTA	GGCTATATAG	CTATATAGCC	TTAATTAGCC	GGCTAATTAA		H10	TTACCTCGAC	GGAGCTGAGG	CCTCAGCTCC	CGCCTCTTAT	ATAAGAGGCG
	D11	AAGAAGCTAG	TCGCGTACTT	AAGTACGCGA	CTCTAACTCG	CGAGTTAGAG		H11	GGAGGATAGC	GCAGCGGACT	AGTCCGCTGC	GTTGGATCTT	AAGATCCAAC

	D 12	TCC GG CCT CG	AAT AAT AAT G	CAT TAT TAT T	TAC GAT CAT C	GAT GAT CGT A		H 12	GG CTC TCT AT	CAT CGC GCT C	GAG CGC GAT G	GCG ATT GCA G	CTG CAA TCG C

Nuclei preparation

89 This is an exemplary protocol for one sample containing 1-5 × 10^6 starting cells. Perform all procedures on ice using chilled reagents and using a refrigerated swing bucket centrifuge.

90 Obtain a single-cell suspension using a dissociation protocol that preserves >95% cell viability. Eliminate any cell clump using a 40-100 µm cell strainer.

91 Extract and fix nuclei:

91.1 Centrifuge cells in a 15 mL conical at 150 g for 5 min, discard the supernatant, resuspend the pellet in 1 mL nuclei lysis buffer, and incubate for 5 min.

91.2 Centrifuge nuclei at 300 g for 3 min, discard the supernatant, re-suspend the pellet in 100 µL nuclei suspension buffer, add 10 mL 4% PFA, mix and incubate for 15 min.

91.3 Centrifuge fixed nuclei at 300 g for 5 min, discard the supernatant, and wash the pellet three times with 1 mL nuclei suspension buffer by repeating this step.

91.4 After the final wash, resuspend the pellet in 300 µL nuclei suspension buffer.

92 Count nuclei with a hemocytometer with YOYO-1 staining (Figure SP2.1), dilute them to 2,500 nuclei/µL, prepare single-use aliquots, and either proceed with the next section or snap-freeze nuclei in liquid nitrogen and store them at -80 °C for up to one month.

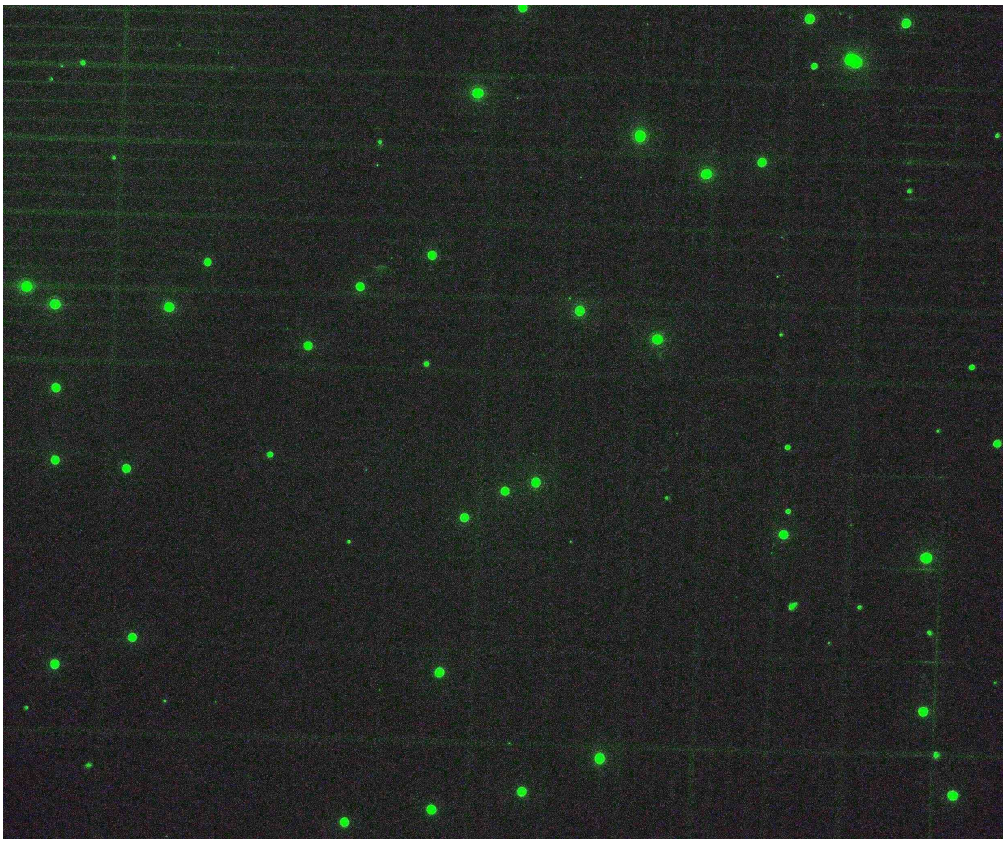


Figure SP2.1. Nuclear extraction QC
Exemplary picture of PFA-fixed nuclei stained with YOYO-1 visualized on a hemocytometer. Nuclei are round, not clumped, and there are minimal cell debris.

Reverse transcription

93 Clean the work area and instrumentation to prevent RNase contamination. Unless stated otherwise, keep samples on ice, perform centrifugations at 4 °C, use DNA LoBind 1.5 mL tubes and low-retention tips, and set the thermocycler heated lid at 105 °C to prevent evaporation.

93.1 Prepare fresh stop solution (40 mM EDTA and 1 mM spermidine; 600 µL/plate)

93.2 Thaw PFA-fixed nuclei in hand or at 37 °C thermal block, mix by flicking, then move to ice

93.3 Prepare a mix by combining nuclei suspension (2 µL, 5,000 nuclei, per well) and 10 mM dNTP (0.25 µL per well), and distribute 2.25 µL into each well of a 96-well LoBind plate

93.4 Add 1 µL of multiplexed RT primers to each well (step 1b of Reagent setup). Cap the plate and centrifuge it for 10 seconds at 100 g

93.5 Incubate the plate at 55 °C for 5 minutes and immediately place it on ice. After 3-5 minutes on ice, centrifuge the plate for 10 seconds at 100 g

93.6 Prepare the following RT mix, distribute 1.75 µL to each well without mixing, cap the plate, and centrifuge it for 10 seconds at 100 g

93.7

	A	B	C
		one well (ul)	one plate
	5X Superscript IV First-strand buffer	1 ul	105 ul
	100mM DTT	0.25 ul	26.25 ul
	Superscript IV reverse transcriptase	0.25 ul	26.25 ul
	RNAseOUT Recombinant Ribonuclease Inhibitor	0.25 ul	26.25 ul

Table SP2.2. iPS2-sci-seq RT mix
Reagent

93.8 Incubate the plate at 55 °C for 10 minutes, immediately place it on ice, and add 5 µL of stop solution into each well to stop the reaction. After 3-5 minutes on ice, centrifuge the plate for 10 seconds at 100 g.

93.9 Gently pool all wells into a FACS tube using wide bore tips, passing through a 40 µm cell strainer, and immediately proceed with the next procedure.

Nuclei sorting

94 Prepare collection plate(s) by adding 5 µL of elution buffer to each well one or more 96-well LoBind
plates, cover, and place on ice

95 Aliquot 50 µL of nuclei suspension (8 from Reverse transcription) in a separate tube to be used as
unstained control; add 3 µM DAPI to the remaining volume to stain nuclei

96 Setup gating for single nuclei (Fig. SP2.2), and sort 25 nuclei into each wells of collection plate
(assuming 96 indexes in the RT and an accepted rate of ~10% index collisions)
Note: use a cell sorter equipped with a 100 µm nozzle and minimize sample pressure, chill both
sample and collection plate, and mix the sample while sorting

97 Centrifuge the plates for 1 minute at 900 g, and either proceed with the next step or store plates at
-80 °C for up to one week

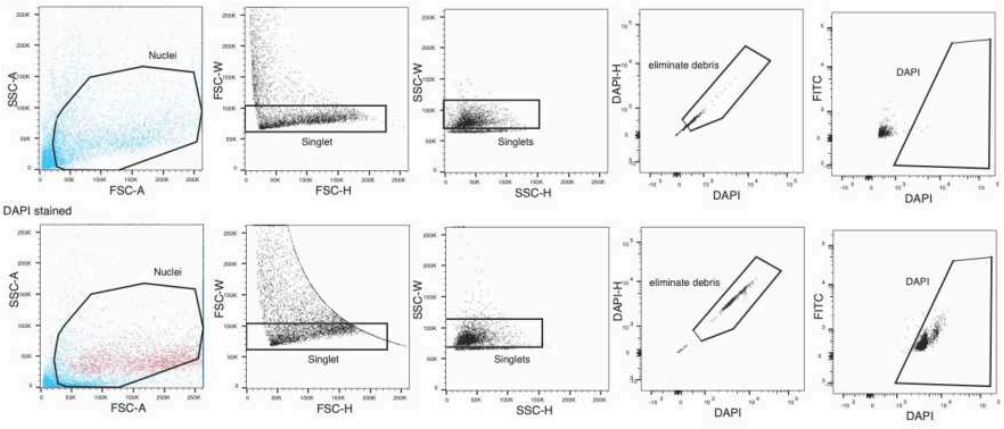
98 

Figure SP2.2. Nuclei FACS Exemplary gating strategy for single nuclei sorting. Set up the gates based on the negative control (top) and sort the DAPI+ population. DAPI+ single nuclei are back-gated in red; debris are in light blue.

Second strand synthesis

99 Prepare reaction master mix, distribute 1 µL to each well, pipette mix, cap the plate, and centrifuge
it for 10 seconds at 100 g

100

	A	B	C
		one well (ul)	one plate
	mRNA Second Strand Synthesis buffer (NEB)	0.6ul	67.5ul
	mRNA Second Strand Synthesis enzyme (NEB)	0.3ul	33.75ul
	mol bio ddH2O	0.1ul	11.25ul

Table SP2.3. iPS2-sci-seq second strand synthesis mix

101 Carry out second strand synthesis at 16 °C for 180 minutes, and terminate the reaction by incubating
at 75 °C for 20 minutes (lid kept at 80 °C for both steps)

102 Centrifuge the plate for 10 seconds at 100 g and either immediately proceed with the next section
or store at 4 °C overnight

Tagmentation

103 Prepare the tagmentation mix with the Nextera TD buffer (5.75 µL/well) and the TDE1 enzyme (0.01 µL/well), distribute 5.76 µL of mix to each well of the sorting plate, pipette mix, cap the plate, and centrifuge it for 10 seconds at 100 g

104 Incubate the plate at 55 °C for 5 min, immediately place it on ice, add 12 µL of Zymo DNA binding buffer to each well, and incubate at room temperature for 5 min.

105 Purify tagmented dsDNA using 1.5X volumes of SPRIselect beads (36 µL/well), following the manufacturer's instructions and eluting in 17 µL of elution buffer

106 Transfer 16 µL of eluted DNA to a new 96-well LoBind plate and either proceed with the next section or store at 4 °C overnight

Indexing PCR and NGS

107 Add 2 µL of 10 µM P5 primer and 2 µL of 10 µM multiplexed P7 primers (step 1e from Reagent setup) to each well, applying a unique combination of i5 and i7

108 Add 20 µL NEBNext High-Fidelity 2X PCR Master Mix to each well, pipette mix, cap the plate, and centrifuge it for 10 seconds at 100 g, and run the following PCR:

109

	cycles	temperature	timing	step
	1	75°C	3 min	
	1	98°C	30 sec	Initial Denaturation

	18-22 cycles	98°C	10 sec	Denaturation
	66°C	30 sec	annealing	
	72°C	1 min	extension	
	1	72°C	5 min	final extension

Table SP2.4. iPS2-sci-seq indexing PCR cycling conditions

110 Pool all samples of a 96-well plate in a 15 mL Falcon tube and concentrate on 4 columns of the Zymo DNA Clean & Concentrator kit, following the manufacturer's instructions except without performing any washes and eluting in a total volume 25 µL of elution buffer

111 Further purify using 0.85X volumes of SPRIselect beads, according to the manufacturer's instruc-
tions and eluting in 50 µL of elution buffer

112 Run 2 µL of a 1:10 dilution of the NGS library on a TapeStation High Sensitivity 5000 Screen Tape
assay, following the manual. Expect a main peak at ~350 bp (Figure SP2.3)

113 Perform NGS using NextSeq 1000/2000 and 50 cycles reagents: read 1 - 18 cycles; index 1 &
index 2 - 10 cycles each; read 2 - 52 cycles (>25,000 reads/nuclei)

114

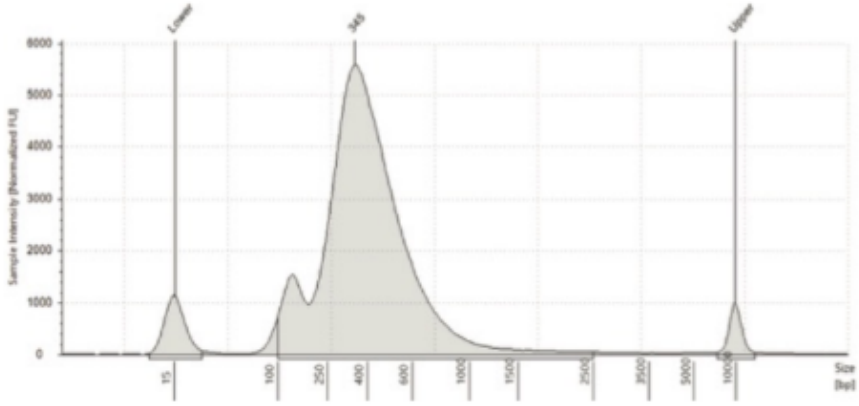


Figure SP2.3. iPS2-sci-seq NGS library QC
Exemplary TapeStation QC of iPS2-sci-seq NGS library (1:10 dilution).

Protocol 3: iPS2-10X-seq library preparation

115 iPS2-10X-seq relies on microfluidics partitioning-based scRNA-seq from 10X Genomics. Despite the higher costs, the ease and reproducibility of this commercial method have led to its widespread adoption. The procedure integrates the manufacturer's protocol to generate an additional NGS library containing UCI-BCs. This is sequenced in parallel to the standard gene expression library to enable *in silico* matching of UCI-BCs and single-cell transcriptomes through their shared unique cell barcodes. Of note, iPS2-10X-seq provides built in support for the cost-saving strategy of reasonably overloading the microfluidics chip, since cell doublets are readily identified by the co-expression of two different UCI-BCs and filtered out as part of the standard *catcheR_10Xcatch* pipeline. Thus, it is possible to load the manufacturer-recommended cell number for barcoded sample pools even without additional multiplexing barcodes.
Other cost-saving strategies include using cell multiplexing oligos (CMOs) or barcoded antibodies against housekeeping cell surface antigens to multiplex tet-untreated and tet-treated cells in a single microfluidics channel, enabling robust control over knockdown experimental design without performing two separate reactions. This Supplemental Protocol describes the following procedures:

1. Prepare the gene expression library
2. Prepare the shRNA barcode library
3. Library pooling and NGS

Prepare the gene expression library

116 Obtain a single-cell suspension using a dissociation protocol that preserves >95% cell viability. Avoid the use of DNase, and remove any EDTA with at least three washes.

116.1 **OPTIONAL:** Label multiple samples using TotalSeq-B hashtags and/or other antibodies with ADTs, following the manufacturer's instructions (CG000149 Rev D).

116.2 **OPTIONAL:** Label multiple samples using cell multiplexing oligos (CMOs) in the 3' CellPlex Kit from 10X Genomics, following the manufacturer's instructions (CG000391 Rev B).

116.3 **OPTIONAL:** Sort live cells after staining with Fixable Viability Dye eFluor 780.
Note: *this step increases the fraction of cells that deliver high-quality transcriptomes, and is particularly recommended for organoids that are difficult to dissociate. Avoid the use of other live/dead staining that can intercalate with DNA.*

117 Centrifuge cells/cell pools at 150 g for 5 min at 4°C, and resuspend in ice cold DPBS 1% BSA aiming for 1 × 10^6^ cells/mL.

118 Manually count cells, and calculate the cell volume needed to load the target cell number per channel of Chromium Next GEM Chip G.
Note: since iPS2-seq cells express shRNA barcodes, chips can be reasonably over-loaded even without any CMO or hashtags (~49,500 cells/reaction at a ~24% doublet rate).

119 Follow the 10X Genomics protocols to obtain and quality control the gene expression library (GEX) and, when appropriate, the CMO library and/or the ADT library.

120 Save a portion of the pre-amplified full-length cDNA for the next procedure.

Prepare the shRNA barcode library

121 Perform a two-step PCR to enrich the UCI-BC region in the 3' UTR of the OPTtetR cDNA.

122 Perform the first PCR according to the calculations and cycling conditions below:

122.1

	A	B
	Reagent	Volume (ul)
	NEBNex High-Fidelity 2X PCR Master Mix	10 ul
	iPS2-10X-seq_truseq	1 ul
	iPS2-10X-seq_inner	1 ul
	cDNA 30-60ng/uL	variable
	H2O (bring to volume)	to 10ul
	Total volume	20 ul

Table SP3.1. iPS2-10X-seq first PCR mix

122.2

	A	B	C	D
	Steps	Temperature	Time	Cycles
	Initial denaturation	98°C	30 sec	1
	Denaturation	98°C	10 sec	15
	Annealing	60°C	20 sec	
	Extension	72°C	20 sec	
	Final extension	72°C	30 sec	1

Table SP3.2. Cycling conditions for iPS2-10X-seq first PCR

122.3 **Note:** When performing iPS2-multi-seq, adjust this step by increasing the input material to 200 ng and reducing the cycle amplification to 11 cycles. Adaptations of the current protocol were essential to obtain clean UCI-BC library profiles.

123 Purify the PCR with 1.8X volumes of SPRIselect beads, following the manufacturer's instructions and eluting in 20 µL of elution buffer.

124 Run and quantify 2 µL of undiluted PCR on a TapeStation High Sensitivity 5000 Screen Tape assay, following the manual. Expect a main peak at ~285 bp (Figure SP3.1A).

125 Dilute the first PCR to 0.1 ng/µL and use it as a template for the second PCR according to the calculations and cycling conditions below:

125.1

	A	B
	Reagent	Volume (µL)
	NEBNext High-Fidelity 2X PCR Master Mix	10
	Dual Index Kit TT Set A primers	4
	First PCR (0.1 ng/uL)	variable
	Nuclease-free water	to 20ul
	Total volume	20

Table SP3.3. iPS2-10X-seq second PCR mix

125.2

	A	B	C	D
	Step	Temperature	Timing	Cycles
	Initial denaturation	98 °C	30 sec	1
	Denaturation	98 °C	10 sec	10
	annealing	58 °C	20 sec	
	extension	72 °C	20 sec	
	final extension	72 °C	30 sec	1

Table SP2.7. Cycling conditions for iPS2-10X-seq second PCR

126 **Note:** use a different index from the one utilized to prepare the GEX, CMO, and/or library, to allow subsequent demultiplexing of the UCI-BC library.

127 Purify and perform a double-sided size selection using 0.7X 26 0.9X volumes of SPRIselect beads, following the manufacturer's instructions and eluting in 20 µL of elution buffer.

128 Run 2 µL of undiluted PCR on a TapeStation High Sensitivity 5000 Screen Tape assay, following the manual. Expect a main peak at ~385 bp (Figure SP3.1B).

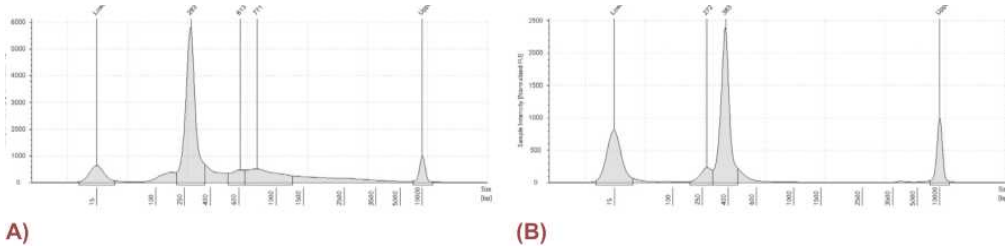
129 

Figure SP3.1. Exemplary TapeStation QC of iPS2-10X-seq first (A) and second (B) PCR.

Library pooling and NGS

130 Quantify the GEX and, when appropriate, CMO and/or ADT libraries (step 3 of Prepare the gene expression library), and the UCI-BC library (step 11 of Prepare the shRNA barcode library) with a Qubit dsDNA High Sensitivity kit.

131 Pool GEX, CMO, ADT and UCI-BC libraries at a relative molar ratio of 60:10:10:3.

132 Perform NGS using NextSeq 1000/2000 and 100 cycles reagents, with the following settings: read 1 - 28 cycles; index 1 - 10 cycles; index 2 - 10 cycles each; read 2 - 90 cycles (3e30,000 reads/target cells).

Protocol 4: Data processing and analysis with catcheR

133 **iPS2-seq design and analysis with catcheR**
This Supplemental Protocol describes the following procedures:

1. Overview
2. Installation
3. Oligonucleotides design
4. Pooled cloning step 1 QC
5. Pooled cloning step 2 and hiPSC genome editing QC
6. iPS2-10X-seq perturbation deconvolution
7. iPS2-sci-seq perturbation deconvolution
8. Barcode reassignment
9. Perturbation effect analysis

Overview

134 *catcheR* is a comprehensive bioinformatic package for designing and analyzing iPS2-seq experiments. For the complete documentation, see **https://marialuisaratto.github.io/catcheRdocs/**

It comprises the following functions:

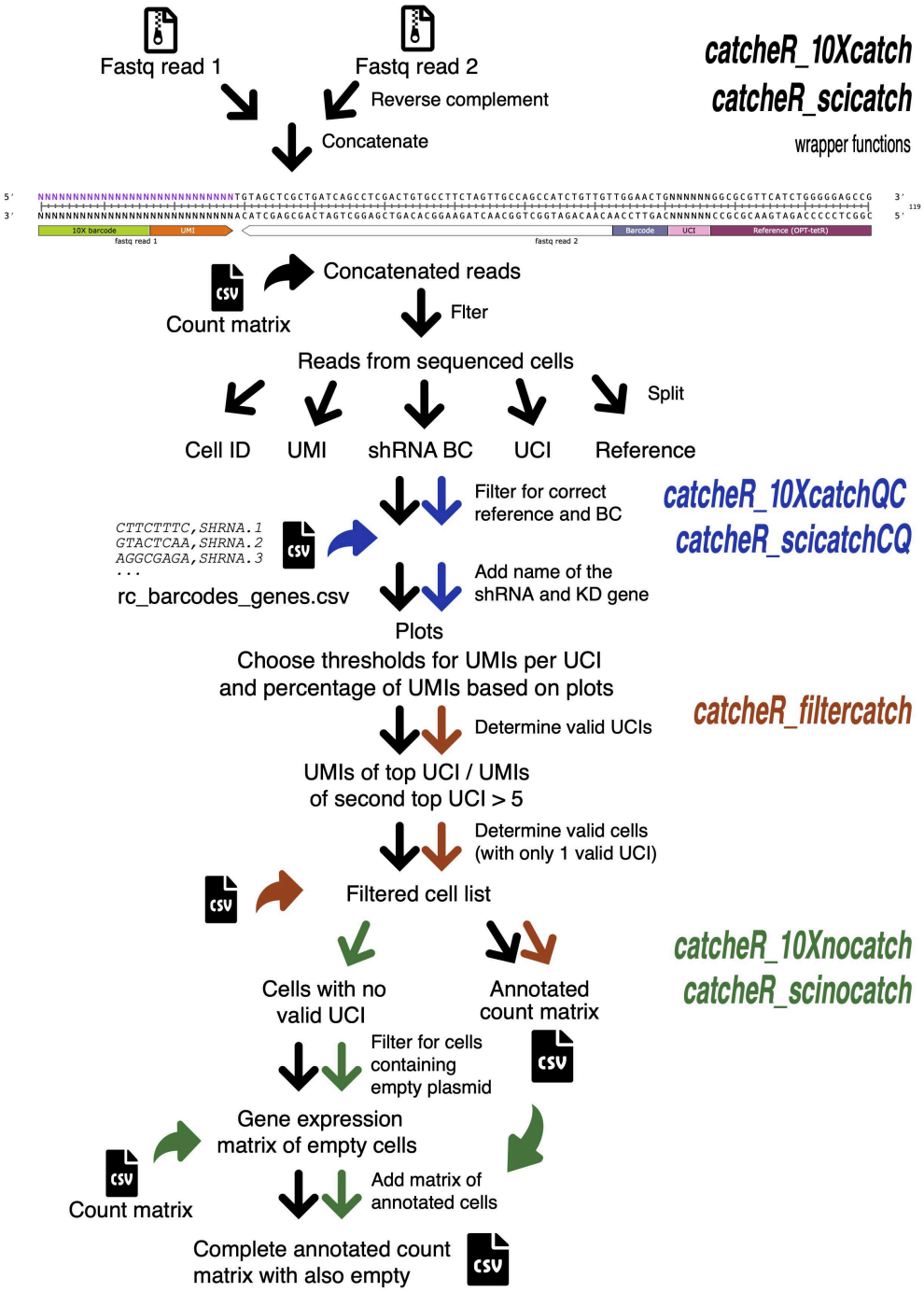


Figure SP4.1. Overview of catcheR
Logical relationship and main inputs/outputs for the catcheR analytical pipeline to assign perturbation in iPS2-seq experiments. Functions used for oligo design, plasmid/hiPSC QC, and perturbation effect analyses are not illustrated.

134.1 *catcheR_design*, which designs oligonucleotides for Supplemental Protocol 1 - Design shRNA oligonucleotides, facilitating shRNA library cloning.

134.2 *catcheR_step1QC*, which analyzes the results of Supplemental Protocol 1 - Intermediate plasmid pool QC, assessing pooled cloning step 1 plasmids for barcode swaps.

134.3 *catcheR_step2QC*, which analyzes the results of Supplemental Protocol 1 - Final plasmid pool QC or hiPSC pool QC, assessing pooled cloning step 2 or genome-edited hiPSC pools for shRNA representation.

134.4 *catcheR_scicount*, which analyzes 2-level indexing sci-RNA-seq data, facilitating the generation of gene expression matrix for iPS2-sci-seq experiments.

134.5 *catcheR_scicatch*, which assigns shRNA perturbations to single nuclei transcriptomes obtained by Supplemental Protocol 2, enabling the primary analysis of iPS2-sci-seq.

134.6 *catcheR_10Xcatch*, which assigns shRNA perturbations to single cell transcriptomes obtained by Supplemental Protocol 3, enabling the primary analysis of iPS2-10X-seq.

134.7 *catcheR_scicatchQC* and *catcheR_10XcatchQC*, which use the outputs of *catcheR_scicatch and catcheR_10Xcatch*, respectively, to fine-tune shRNA assignment thresholds.

134.8 *catcheR_filtercatch*, which leverages on the output of *catcheR_scicatchQC* and *catcheR_10XcatchQC* to filter single nuclei/cell transcriptomes expressing a single shRNA.

134.9 *catcheR_sortcatch,* which quality controls the cell-by-gene matrix based on the results of *catcheR_step1QC*, reassigning hPSC clones with barcode swaps to the correct shRNA.

134.10 *catcheR_scinoatch* and *catcheR_10Xnoatch*, which identify cells expressing no shRNA in iPS2-sci-seq and iPS2-10X-seq experiments, respectively, adding them to the cell-by-gene matrix to be used as additional controls.

134.11 *catcheR_load*, which loads gene expression matrices annotated with shRNA perturbations into a Monocle object, preparing the dataset for downstream analysis.

134.12 *catcheR_pseudotime*, which analyzes the effects of shRNA perturbations on pseudotime dynamics, highlighting shifts along differentiation trajectories (e.g., Figures 2J, 2K).

134.13 *catcheR_modules*, which assesses perturbation-induced changes in gene module expression, such as coordinated activation or repression of functional programs (e.g., Figures 5H–5J).

134.14 *catcheR_enrichment*, which quantifies differences in perturbation representation across experimental samples (e.g., Figures 5F, 5G) or cell clusters (e.g., Figure 5E).

Installation

135 *catcheR* is available at https://github.com/alessandro-bertero/catcheR. The GitHub repository folder "scripts" contains all the bash and R scripts that can be run independently. However, in order to ensure reproducible analyses, we strongly recommend to install *catcheR* package from GitHub, since its functions run all the analysis inside Docker containers.

135.1 Install Docker engine, following the instructions at https://docs.docker.com/engine/install/

135.2 Install *catcheR*
In R (≥ 3.0.2) install *devtools*, if not already present

```
install.packages("devtools")
Install catcheR from GitHub
install_github("alessandro-bertero/catcheR")
Install rundocker from GitHub
install_github("Reproducible-Bioinformatics/rundocker")
```

135.3 Load *catcheR* and *rundocker* in your R environment

```
library(catcheR)
library(rrundocker)
```

Oligonucleotides design

136 In a new working folder, prepare the following files:

- A comma-separated values (CSV) file with three columns listing: (1) the forward oligos from the TRC shRNA library; (2) the corresponding barcodes (BC); (3) the shRNA names. Below is an example for the shRNA described in Figure SP1.2:

```
CCGGCAAACTCCTCTGCGATTCTCGGAACATCCAGACAGAGACTGTTTTTG,CAGTTCCA,SMAD2
.1
```

- (Optional) - A txt file with a newline-separated list of 5′-3′ restriction sites, or other sequences, to be avoided in the shRNAs. By default these are SalI, SwaI and AscI:

```
GTCGAC
ATTTAAAT
GGCGCGCC
```

137 Run catcheR_design:

```
catcheR_design(
group=c("docker","sudo"),
folder,
sequences,
gibson.five = "AGTTCCTCTACTAGTGATAGAGATCCC",
gibson.three = "GTAGCTCGGTGATCAGC",
fixed = "GTCGACATTTAAATGGCGCGCC",
restriction.sites = NULL)
```

138 *catcheR_design* arguments:

1. group: string with two options: sudo or docker, depending on the user group. For a detailed explanation of Docker user groups, see this page
2. folder: string with the working folder path
3. sequences: string with the CSV file name from step 1a
4. gibson.five: (optional) string with the 5′ Gibson homology
5. gibson.three: (optional) string with the 3′ Gibson homology
6. fixed: (optional) character string with the multicloning site
7. restriction.sites: (optional) string with the txt file name from Step 25

138.1 Example usage:

```
catcheR_design(
group = "docker",
folder = "path/to/folder",
sequences = "filename.csv",
restriction.sites = "filename.txt")
```

Pooled cloning step 1 QC

139 This step complements Supplemental Protocol 1 - Intermediate plasmid pool QC.
In a new working folder, prepare the following files:

- (a) Fastq/fq or fastq.gz files with demultiplexed read 1 from the NGS run
- (b) A CSV file with the shRNA names and their full sequences

```
SMAD2.1,GCAGATCCTCTCTGCTGGATTCTCGGAACATCCAGACAGAGACTG
```

- (c) (Optional) - A txt file with a newline-separated list of clones of interest as "BC_UCI"

```
CAAGACCC_CATCGT
```

140 Run catcheR_step1QC:

```
catcheR_step1QC(
group=c("docker","sudo"),
folder,
fastq.read1,
DIs = 100,
ratio = 10,
plot.threshold = 2000,
clones = NULL)
```

140.1 *catcheR_step1QC* arguments:

- (a) group: string with two options: sudo or docker, depending on the user group (info)
- (b) folder: string with the working folder path
- (c) fastq.read1: string with the read 1 filename from step 1a

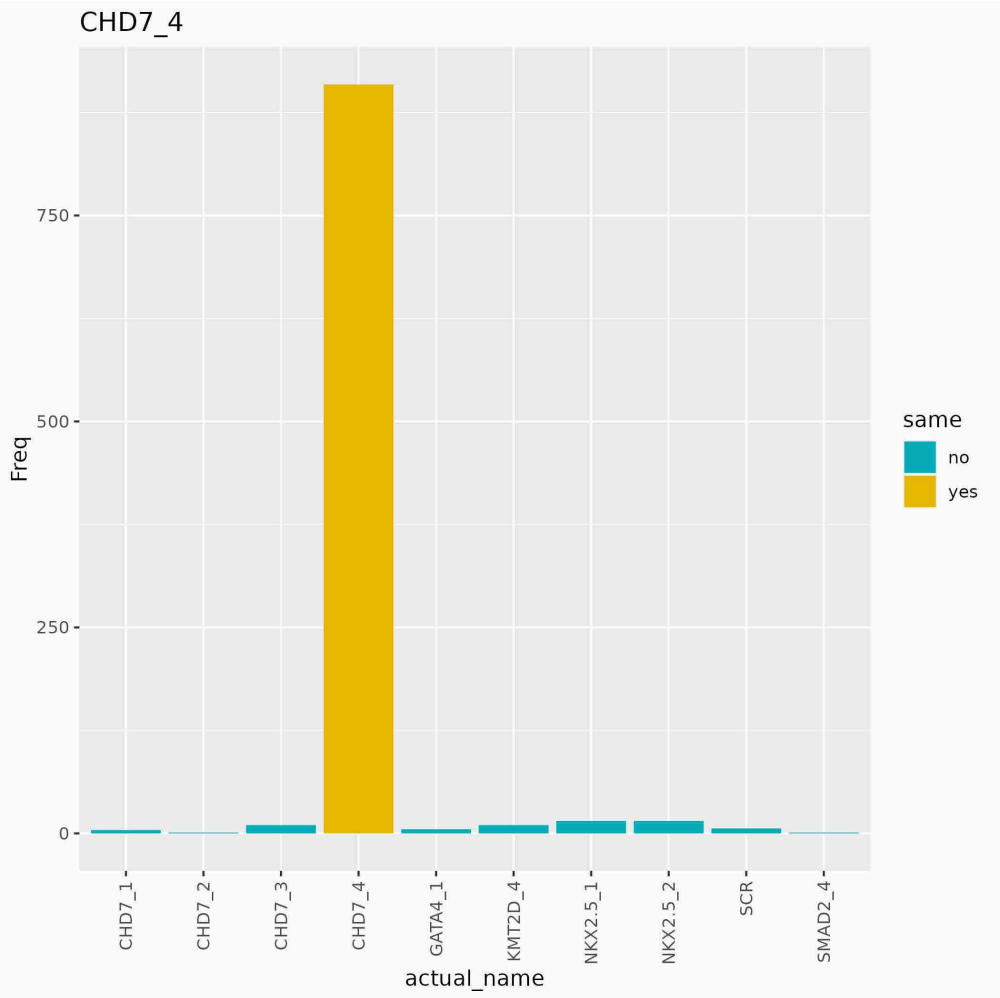
- (d) DIs: integer of the minimum number of diversity indexes (DIs, pseudo-unique reads) of the most represented shRNA matching to a given UCI-BC; in combination with ”ratio”, it selects UCI-BCs for which it is possible to reliably assign an shRNA
- (e) ratio: integer of the minimum ratio between the number of DIs of the most represented and second most represented shRNAs matching to a given UCI-BC
- (f) plot.threshold: integer of the minimum number of DIs per UCI-BC for output 2b
- (g) clones: (optional) a string with the txt file from step 1c

140.2 Example usage:

```
catcheR_step1QC (
group = "docker",
folder = "path/to/folder",
fastq.read1 = " filename .fq",
clones = " filename .txt ")
```

140.3 *catcheR_step1QC* key outputs:

- (a) ”reliable_clones_swaps.csv”, which lists UCI-BCs with reliable evidence of shRNA-barcode swap, and is used as input for Barcode reassignment
- (b) Bar chart of the number of DIs for each clone above ”plot.threshold”
- (c) Bar chart of the number of DIs associated to each BCs, shRNAs, and reliable swaps
- (d) (If ”clones” argument is provided) - CSV files and bar charts with the number of DIs for each shRNA matching to each clone of interest (Figure SP4.2)


CHD7_4
Freq
750
500
250
0
CHD7_1
CHD7_2
CHD7_3
CHD7_4
GATA4_1
KMT2D_4
NKX2.5_1
NKX2.5_2
SCR
SMAD2_4
actual_name
same
no
yes

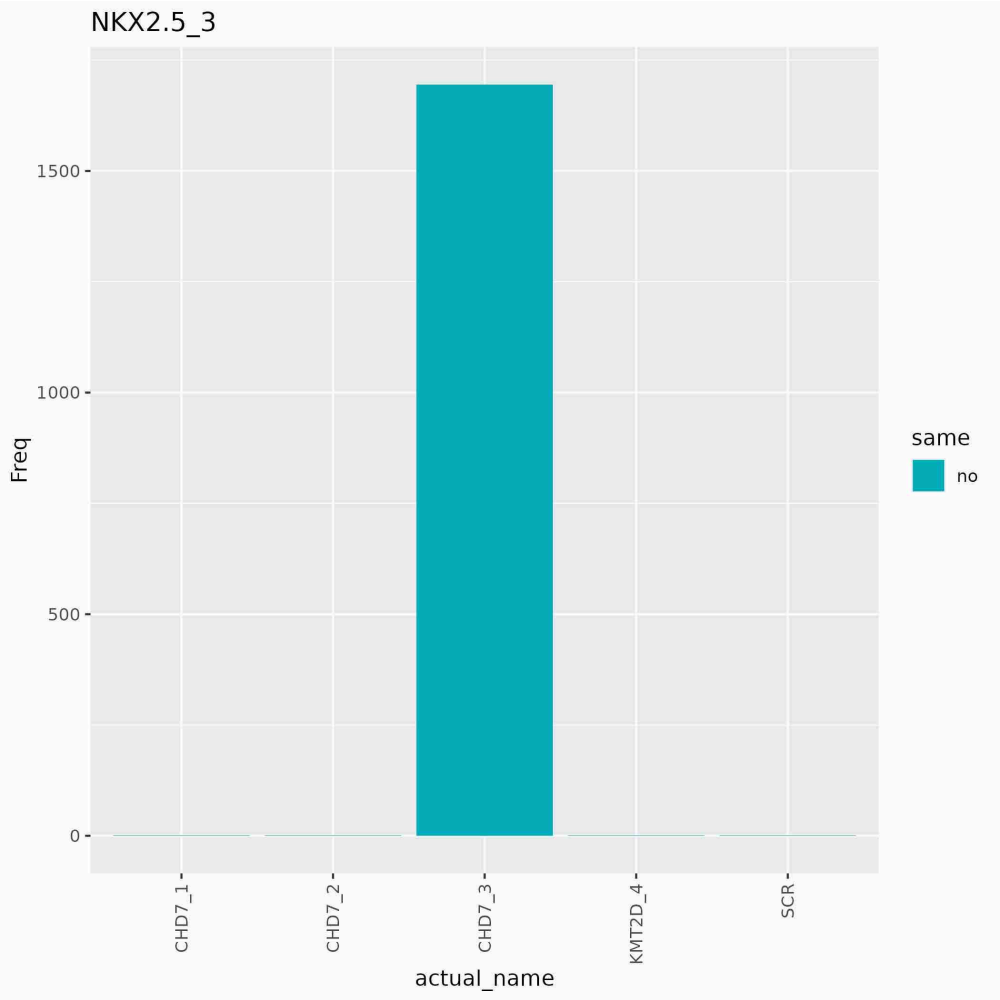


Figure SP4.2. Example of quantification of DIs for bacterial clones carrying the expected shRNA (A) or with robust evidence of a barcode swap that can be reassigned (B; Figure 1B). These graphs allow visual evaluation of appropriate thresholds for "DI" and "ratio".

Pooled cloning step 2 and hiPSC genome editing QC

141 In a new working folder, prepare the following files:

- (a) Fastq/fq or fastq.gz files with demultiplexed read 1 from the NGS run
- (b) "rc_barcodes_genes.csv", a CSV file with two columns: (1) the shRNA BCs; (2) the matching shRNA names in the format "GENE.shRNAID"
- (c) (Optional) - A txt file with clones of interest (step 1c of Pooled cloning step 1 QC)

142 Run catcheR_step2QC:

```
catcheR_step2QC(
group=c("docker","sudo"),
folder,
fastq.read1,
DIs = 1000,
clones = NULL)
```

142.1 catcheR_step2QC arguments:

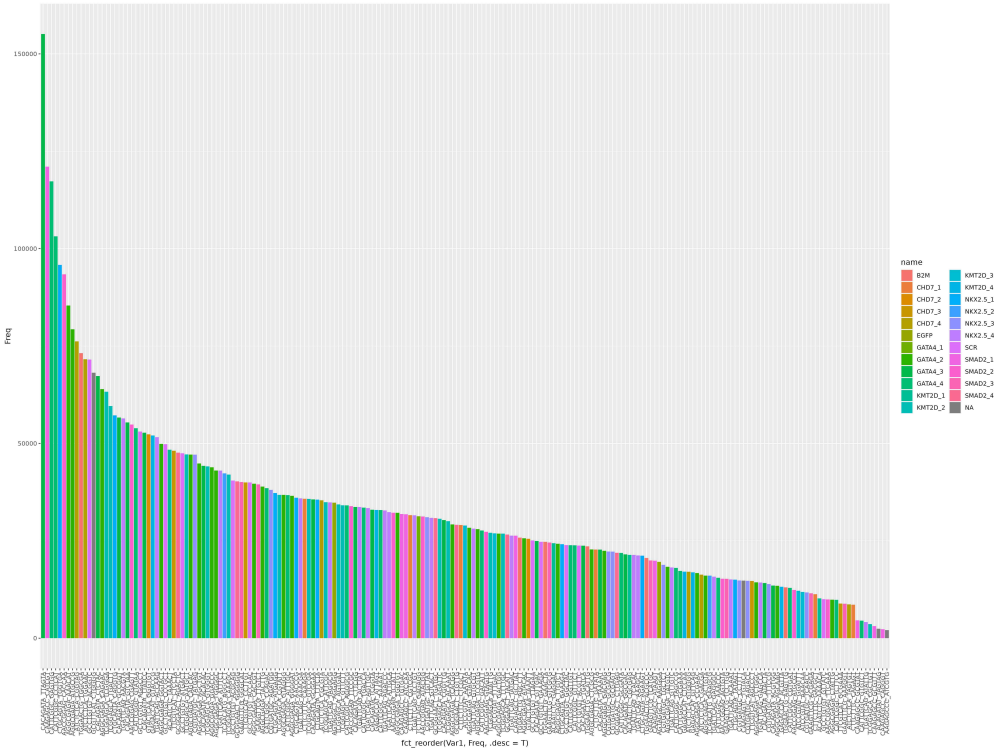
- (a) group: string with two options: sudo or docker, depending on the user group
- (b) folder: string with the working folder path
- (c) fastq.read1: string with the read 1 file from step 1c
- (d) DIs: integer of the minimum number of DIs for a given UCI-BC; it selects reliably-measured UCI-BCs
- (e) clones: (optional) a string with the txt file from step 1c

142.2 Example usage:

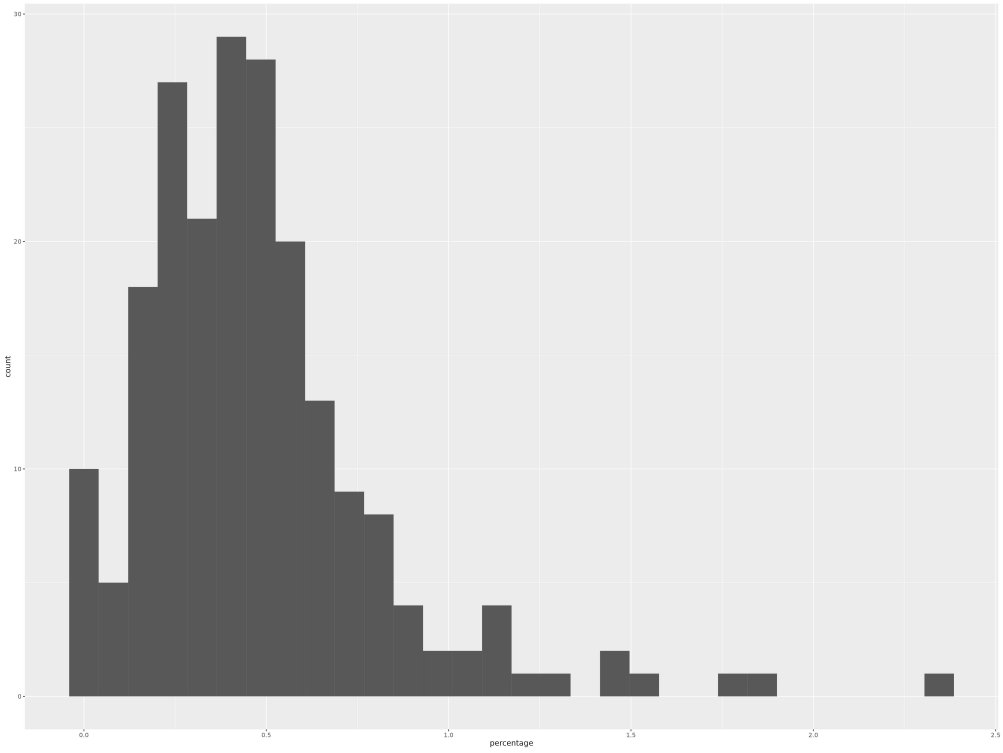
```
catcheR_step2QC(
group = "docker",
folder = "path/to/working/folder",
fastq.read1 = "filename.fq",
clones = "filename.txt")
```

142.3 catcheR_step2QC key outputs:

- (a) Pie charts of DIs per shRNA and per gene targets associated to each clone
- (b) Text file listing all clones above the DI threshold
- (c) Bar chart of the number of DIs for each clone above the DI threshold (Figure SP4.3A)
- (d) Frequency histogram of the percentage of clones above the DI threshold for each shRNA (Figure SP4.3B)


Freq
150000
100000
50000
0
fct_reorder(Var1, Freq, .desc = T)
name
B2M
CHD7_1
CHD7_2
CHD7_3
CHD7_4
EGFP
GATA4_1
GATA4_2
GATA4_3
GATA4_4
KMT2D_1
KMT2D_2
KMT2D_3
KMT2D_4
NKX2.5_1
NKX2.5_2
NKX2.5_3
NKX2.5_4
SCR
SMAD2_1
SMAD2_2
SMAD2_3
SMAD2_4
NA

A



B

Figure SP4.3. Example of quantification of DIs for bacterial clones represented by more than 1,000 DIs (A), and of the frequency of bacterial clones for each shRNA. These graphs confirm whether shRNAs are normally distributed in the final plasmid pool.

Obtain iPS2-10X-seq, iPS2-CITE-seq, or iPS2-multi-seq count matrice

143 Download and install **cellranger** (for iPS2-10X-seq and iPS2-CITE-seq) or **cellranger-arc** (for iPS2-multi-seq). Alternatively, Docker containers for **cellranger v7**, **cellranger v9** (recommended for iPS2-CITE-seq), or **cellranger-arc** can be pulled from our Docker repository.

143.1 For iPS2-10X-seq and iPS2-CITE-seq, demultiplex Illumina base call (BCL) files using cellranger mkfastq, following the standard 10X Genomics user guide. In the sample sheet CSV, include the index sequences used in *Supplemental Protocol 3* for the GEX and UCI-BC libraries, and, when applicable, CMO and/or ADT libraries.

143.2 For iPS2-multi-seq, use cellranger-arc mkfastq to demultiplex GEX + ATAC dual-index libraries, following the Cell Ranger ARC user guide. Ensure the sample sheet format is compatible with dual-modality sequencing runs, and include the index sequences used for both gene expression and chromatin accessibility libraries.

144 Run FastQC to assess the quality of the FASTQ files for each library type.

145 Generate cell-by-gene count matrices using cellranger count (for single-sample experiments) or cellranger multi (for multiplexed experiments, including iPS2-CITE-seq). For iPS2-multi-seq, use cellranger-arc count to obtain both gene expression and chromatin accessibility matrices.

145.1 Note: In multiplexed experiments (e.g., using CMO or ADT barcodes in iPS2-CITE-seq), count matrices from individual samples can be aggregated using cellranger aggr to produce a unified dataset. This enables joint analysis in a single run of catcheR_10Xcatch by specifying the number of samples via the samples argument.

146 Use cellranger mat2csv to convert the sparse matrix output into a dense CSV format for compatibility with downstream tools. For iPS2-multi-seq, run cellranger-arc mat2csv separately for GEX and ATAC outputs, if needed.
After obtaining the count matrix, proceed to *iPS2-10X-seq perturbation deconvolution*.

Obtain iPS2-sci-seq count matrices

147 *catcheR_scicount* is a wrapper of the "bbi-sci" pipeline developed by the Brotman Baty Institute for Precision Medicine, which was implemented and dockerized in catcheR to be used on any operating system.

Demultiplex Illumina base calls to fastq files

- (a) Create a "SampleSheet.csv" file with as many sample rows as the PCR wells from Indexing PCR and NGS of Supplemental Protocol 2, where "Sample_ID" is the well identified with format [A-H]01-[12], and index and index2 are the i7 and i5 indexes used in the corresponding row and column (refer to Table SP2.1)
- (b) Run Illumina bcl2fastq following Illumina's manual
- (c) Run fastQC to confirm the quality of the fastq files

148 In a new working folder:

- (a) Create the subfolder "fastq", and copy all the demultiplexed fastq.gz files. Ensure all file names begin with the well coordinate (e.g. A01) followed by an underscore
- (b) Create a tab separated txt file called "sci-RNA-seq-8.RT.oligos" with the association between RT wells and RT barcode sequences (refer to *Table SP2.1*)
- (c) Create the subfolder "GENOMES", and copy the annotated genome (i.e., **GRCh38**)

149 Run catcheR_scicount:

```
catcheR_scicount(
group=c("docker","sudo"),
folder,
sample.name,
UMI.cutoff)
```

149.1 catcheR_scicount arguments:

- (a) group: string with two options: sudo or docker, depending on the user group (***info***)
- (b) folder: string with the working folder path
- (c) sample.name: string with the name of the experiment
- (d) UMI.cutoff: integer of the minimum number of UMI per nucleus needed to consider the single cell transcriptome valid

149.2 Example usage:

```
catcheR_scicount(
group = "docker",
folder = "path/to/file",
sample.name = "experiment",
UMI.cutoff = 500)
```

149.3 catcheR_scicount outputs, found in the final-output folder: knee-plot of UMI per cell (Figure S1H), statistics files, sparse cell-by-gene count matrix, dense gene matrix "exp_mat.csv" and "exp_mat_no0.csv" (filtered to exclude genes with 0 counts in all cells), also in Rdata format.
After obtaining the count matrix, proceed to *iPS2-sci-seq perturbation deconvolution*.

iPS2-10X-seq perturbation deconvolution

150 shRNA perturbations can be assigned to single cells using *catcheR_10Xcatch*, which identifies NGS reads containing UCI-BCs, matches them to the corresponding transcriptome based on their shared cellular barcodes, applies filters for background noise, and selects cells with robust evidence of a single integration. UCI-BC filtering involves: (1) filtering out UCI-BCs supported by less than a certain number of UMIs in a given transcriptome, to account for PCR and sequencing artifacts; (2) filtering out UCI-BCs that represent less than a certain fraction of all UMIs for UCI-BCs in a given transcriptome, to reduce noise arising from free mRNA; and (3) filtering out UCI-BCs whose UMI fraction in a given transcriptome is not several fold greater than the second most common UCI-BC, to eliminate cells that contain multiple UCI-BCs when the second most common one falls just below one of the first two thresholds. The resulting list of *bona fide* shRNA integrations per cell is then used to filter those expressing a single shRNA. This section outlines a typical analysis workflow for iPS2-10X-seq, iPS2-CITE-seq or iPS2-multi-seq experiments, which follow analogous preprocessing and count matrix structures based on 10X Genomics technology. The pipeline begins with *catcheR_10Xcatch*, which performs a full analysis using automatic thresholding. If needed, these thresholds can be subsequently refined

with *catcheR_10XcatchQC* and applied for re-filtering using *catcheR_filtercatch*. Finally, *catcheR_10Xnocatch* allows incorporation of cells that do not express any shRNA as additional negative controls in the final annotated cell-by-gene matrix.

In a new working folder:

150.1 Copy the fastq/fq or fastq.gz files with demultiplexed read 1 and read 2 of the UCI-BC library (step 3 of Obtain iPS2-10X-seq, iPS2-CITE-seq, or iPS2-multi-seq count matrices)

150.2 Copy the cell-by-gene count matrix in CSV format (step 6 of Obtain iPS2-10X-seq, iPS2-CITE-seq, or iPS2-multi-seq count matrices)

150.3 Create a CSV file named “rc_barcodes_genes.csv” with two columns: (1) the shRNA BCs; (2) the matching shRNA names in the format “GENE.shRNAID”

151 Run catcheR_10Xcatch to execute the complete analysis with automatic thresholding:

```
catcheR_10Xcatch(
group=c("docker","sudo"),
folder,
fastq.read1, fastq.read2,
expression.matrix,
reference = "GGGCCGCTCATCTGGGGAGCCG",
UCI.length = 6,
threads = 2,
percentage = 15,
mode = "bimodal",
ratio = 5,
samples = 1,
x = 100, y = 400)
```

151.1 Example usage:

```
catcheR_10Xcatch(group = "docker",
folder = "path/to/folder",
fastq.read1 = "R1.fastq.gz", fastq.read2 = "R2.fastq.gz",
expression.matrix = "filename.csv",
threads = 12)
```

151.2 catcheR_10Xcatch arguments:

- (a) group: string with two options: sudo or docker, depending on the user group (**info**)
- (b) folder: string with the working folder path
- (c) fastq.read1: string with the filename of read 1 fastq/fq or fastq.gz (step 1a)
- (d) fastq.read2: string with the filename of read 2 fastq/fq or fastq.gz (step 1a)
- (e) expression.matrix: string with the filename of the count matrix file CSV (step 1b)
- (f) reference: string with the reverse complement of the sequence before the shRNA-BC at the start of read 2 (default is the 3′ end of the OPTtetR cDNA, optional argument; Figure S2F)
- (g) UCI.length: integer of the UCI length (default is 6, optional argument)
- (h) threads: integer of the threads to be used for parallelization (default is 2, optional argument)
- (i) percentage: integer of the percentage of UMIs supporting a given UCI over the total UMIs supporting all UCIs in a given cell; it is used as threshold to consider the UCI valid (Figure SP4.4D); the recommended default is 15 (optional argument)
- (j) mode: string with two options: bimodal or noise, defining the mode for automatic thresholding the minimum number of UMIs per UCI to consider the UCI valid (optional argument)

1. ″bimodal″ (default) sets the threshold at the valley of the bimodal UMIxUCI distribution (Figure SP4.4C)
2. ″noise″ sets the threshold at 1.35 * the number of UCIs supported by a single UMI; used if the UMIxUCI distribution is not bimodal

- (k) ratio: the ratio between the number of UMIs supporting the most represented UCI and the number of UMIs supporting the second most represented UCI. The parameter is needed to identify the cell as a single integration cell. The default is 5 (optional argument)
- (l) samples: the number of samples present in the same experiment (cells from different experiments will have the corresponding number after the cell ID in the gene expression matrix, so that multiple reactions of the same experiment can be unified). The default is 1 (optional argument)
- (m) x: an integer indicating the upper limit on the x-axis of the cropped version of the plot ″UMIxUCI″. Default is 100 (optional argument)
- (n) y: an integer indicating the upper limit on the y axis of the cropped version of the plot ″UMIxUCI″. Default is 400 (optional argument)

151.3 catcheR_10Xcatch key outputs (will be stored inside the "Result" folder of each sample):

- (a) ″log.txt″ with the starting number of reads
- (b) ″log2.txt″ with the number of cells, UMIs, UCIs, and the calculated thresholds of UMI per UCI for both ″bimodal″ and ″noise″ (only one is chosen based on the ″mode″ argument; the other one is provided for a potential reanalysis)
- (c) Bar charts of UMI counts per shRNA and target gene (Figure SP4.4A and Figure SP4.4B)

- (d) Frequency histograms of UCIs supported by a certain number of UMIs (UMIxUCI), to interpret the signal/noise of the experiment and possibly set a custom UMIxUCI threshold for subsequent reanalysis (Figure SP4.4C) **Note**: here and below, copies of the same UCI expressed by different cells are plotted and analyzed separately
- (e) Frequency histogram of UCIs supported by a certain fraction of UMIs over the total number of UMIs supporting all UCIs in a given cell (UMIpercentagexUCI), to further assess signal/noise and possibly adjust the default threshold (Figure SP4.4D
- (f) Dot plots that combine the UMIpercentagexUCI and UMIxUCI data on the x and y axes, respectively, with each dot representing one or more UCI (quantified by the dot size). Dot colors indicate either the number of valid UCIs (i.e., shRNAs) in the cell containing a given UCI (Figure SP4.4E), or whether said UCI is the only valid one in the relevant cell and is thus assigned to such cell (Figure SP4.4F)
- (g) "log_part3.txt" with how many cells were assigned to a single shRNA or were filtered out due to zero or multiple shRNAs
- (h) "silencing_matrix.csv" is the key output used for the secondary analyses: the cell-by-gene count matrix provided as input, filtered and annotated with the shRNA encoded in each cell. It is also provided in RDS format for easy loading into R. Cell names are modified as follows:

```
cellID_UMIxUCI_BC_GENE_UCI
```

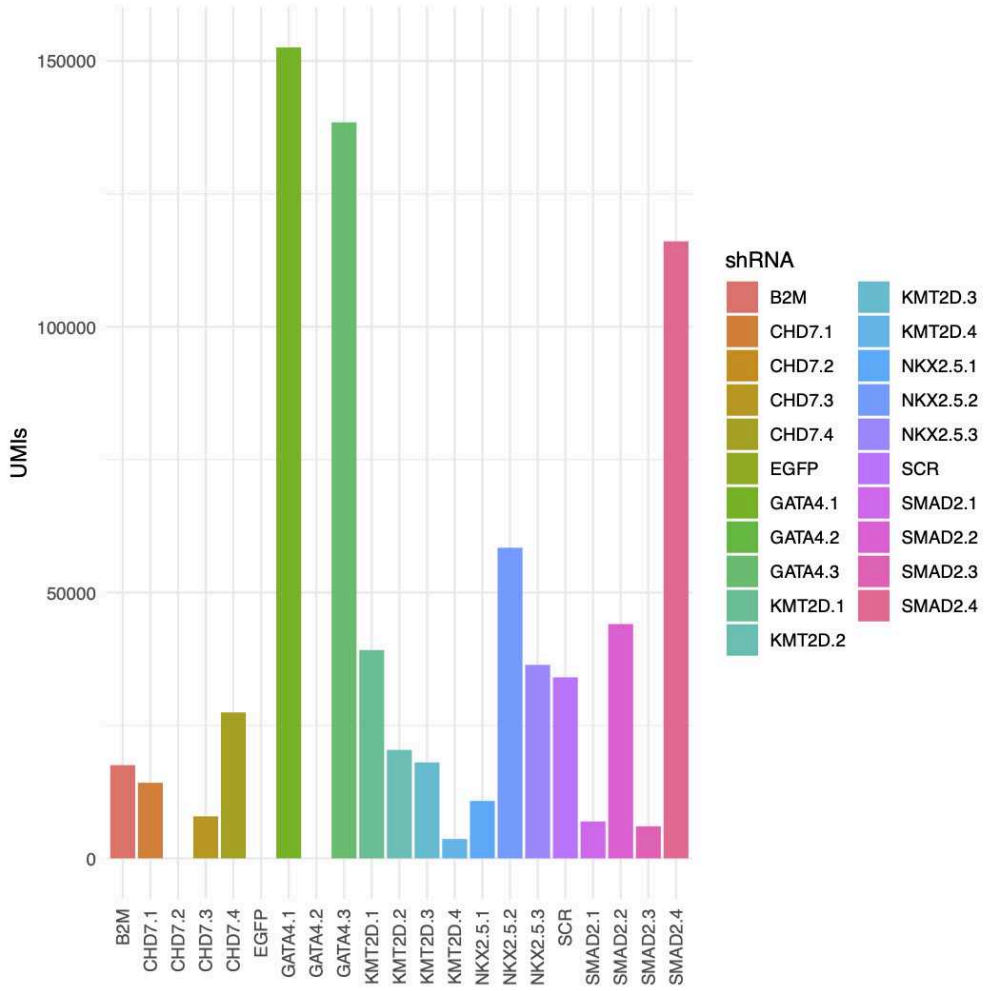
Where:

1. cellID is the original cell name (i.e., the 16 bp of the 10X RT barcode)
2. UMIsxUCI is the number of UMI associated with that perturbation
3. BC is the shRNA barcode
4. GENE is the shRNA target
5. UCI is the Unique Clonal Identifier (shared by all cells originating from the same hPSC clone)

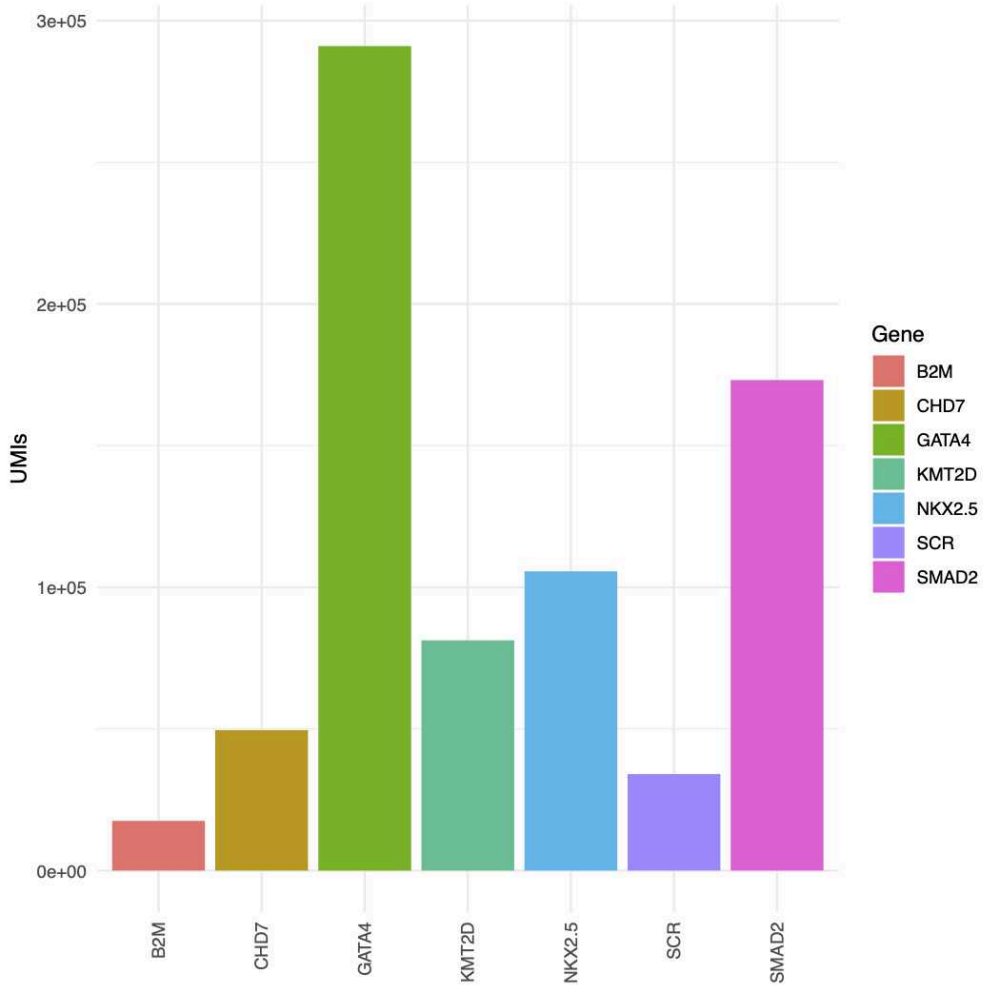
Example of a cell name after the analysis:

```
TTCTAACCACAGTCGC_180_CGTGATGC_NKX2 .5 _ACAGTG
```

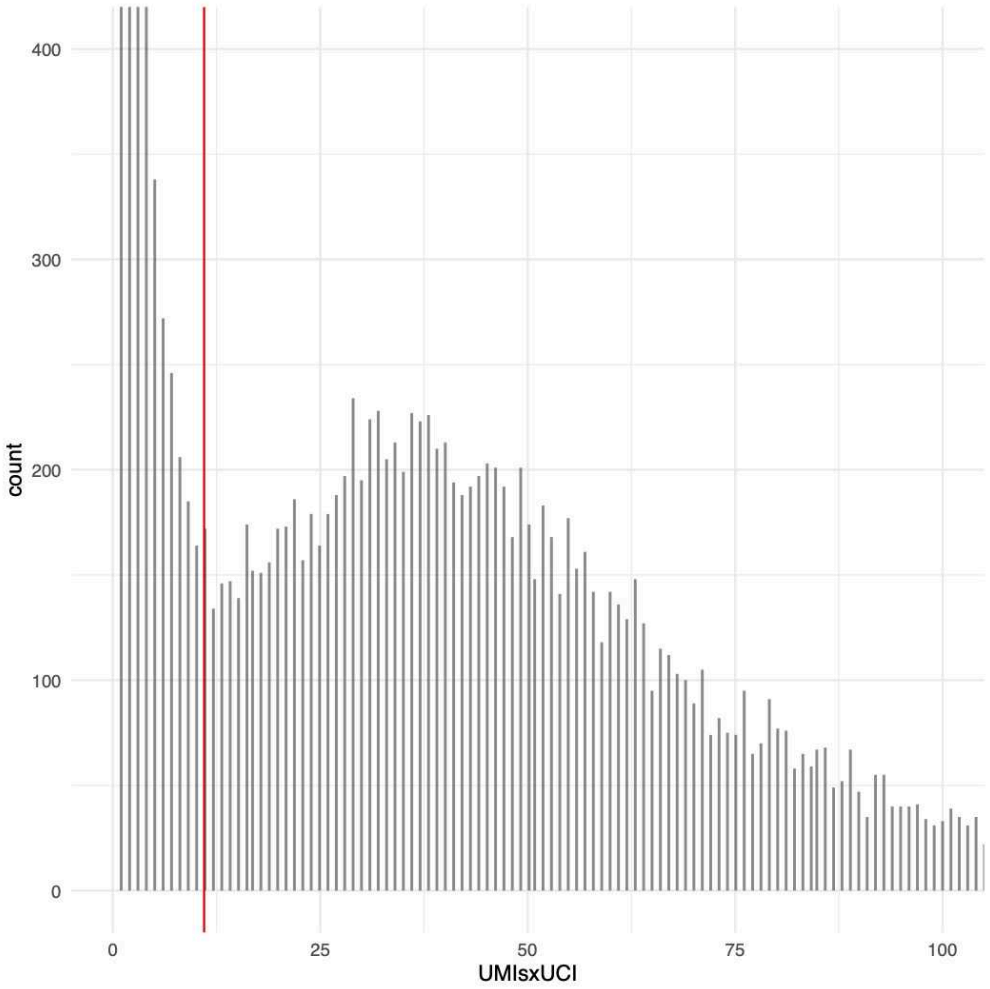
The annotated matrix can be directly used with the functions implemented in catcheR (see Perturbation effect analysis) to perform a range of secondary analyses described in the manuscript. Additional scripts are available at our GitHub repository for customized or extended analyses, most of which are now incorporated into catcheR.


150000
100000
50000
0
UMIs
B2M
CHD7.1
CHD7.2
CHD7.3
CHD7.4
EGFP
GATA4.1
GATA4.2
GATA4.3
KMT2D.1
KMT2D.2
KMT2D.3
KMT2D.4
NKX2.5.1
NKX2.5.2
NKX2.5.3
SCR
SMAD2.1
SMAD2.2
SMAD2.3
SMAD2.4
shRNA
B2M
CHD7.1
CHD7.2
CHD7.3
CHD7.4
EGFP
GATA4.1
GATA4.2
GATA4.3
KMT2D.1
KMT2D.2
KMT2D.3
KMT2D.4
NKX2.5.1
NKX2.5.2
NKX2.5.3
SCR
SMAD2.1
SMAD2.2
SMAD2.3
SMAD2.4

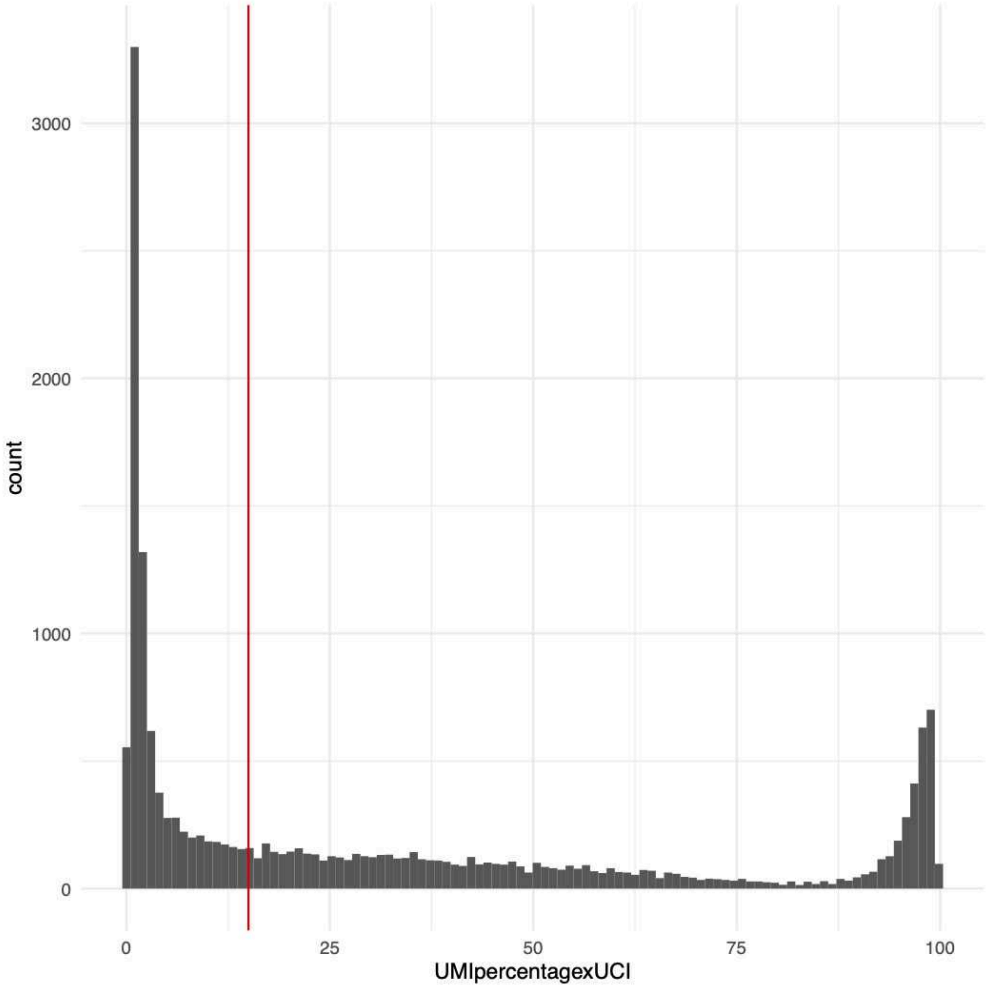
A


3e+05
2e+05
1e+05
0e+00
UMIs
B2M
CHD7
GATA4
KMT2D
NKX2.5
SCR
SMAD2
Gene
B2M
CHD7
GATA4
KMT2D
NKX2.5
SCR
SMAD2

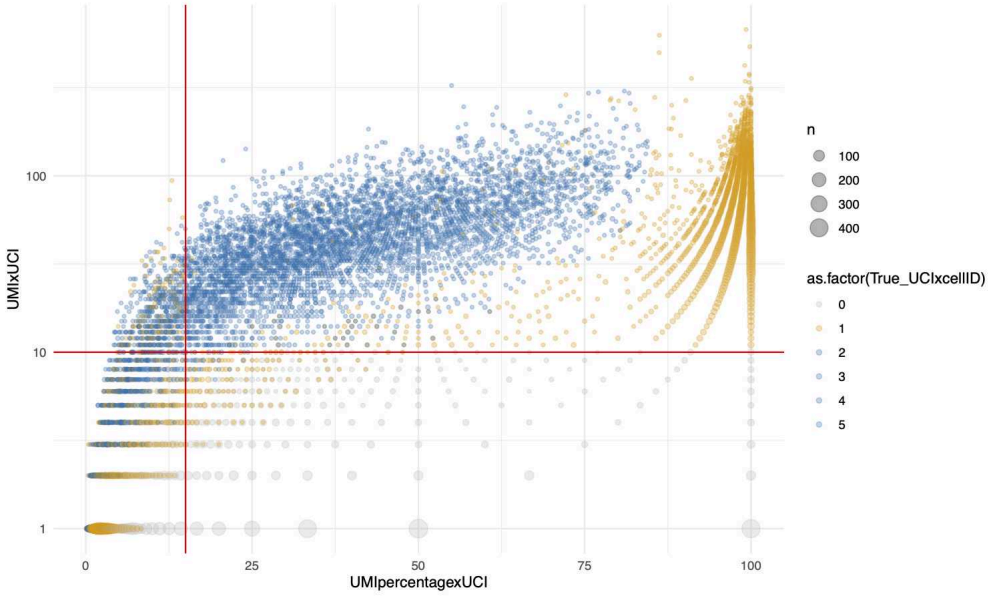
B



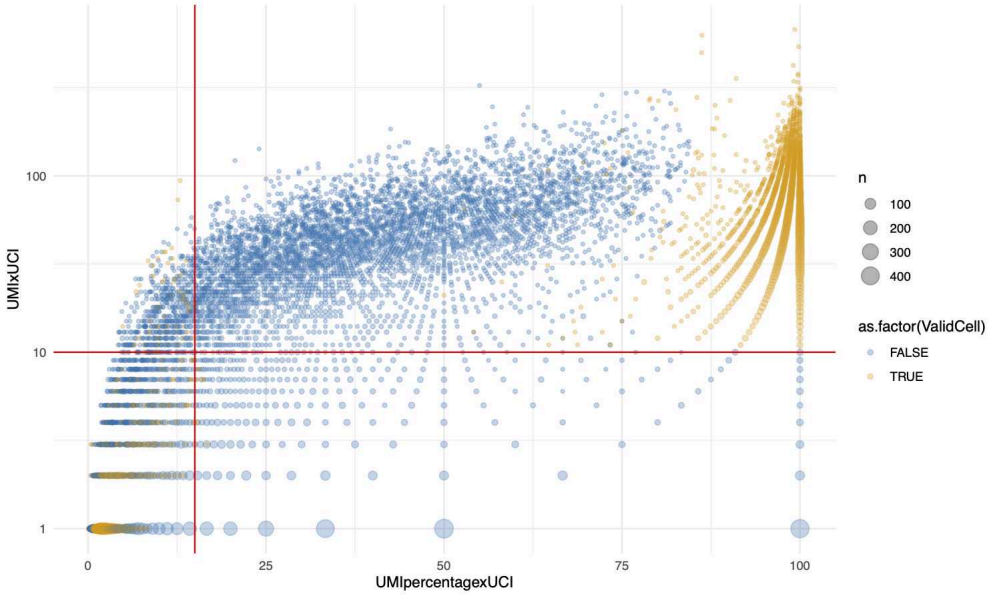
C



D



E



F

Figure SP4.4. Exemplary key outputs of a catcheR_10Xcatch analysis: "barcode_distribution.pdf" (A); "gene_distribution.pdf" (B); "UMIxUCI_400_100.pdf" with the automatic UMIxUCI threshold based on "bimodal" (C); "percentage_of_UMIxUCI_dist.pdf" with the default 15% "percentage" threshold (D); "2D_percentage_of_UMIxUCI_UMI_count_trueorfalse.pdf" (E); and "2D_percentage_of_UMIxUCI_UMI_ValidCells.pdf" (F).

OPTIONAL: fine tune iPS2-10X-seq perturbation assignment

152 In the same working folder used to run *catcheR_10Xcatch,* run *catcheR_10XcatchQC*, which uses the txt file outputs to regenerate the quality control plots described in **Figure SP4.4** and suggest new thresholds to enable subsequent filtering.

```
catcheR_10XcatchQC(
    group=c("docker","sudo"),
    folder,
    reference = "GCGCGCTTCATCTCGGGGAGCCG",
    mode = "bimodal",
    sample = 1
    x = 100, y = 400)
```

152.1 Example usage:

```
catcheR_10XcatchQC(
    group = "docker",
    folder = "path/to/working/folder",
    mode = "noise")
```

The function is to be run separately for each sample (specified by argument sample with default 1). *catcheR*_10XcatchQC *arguments and outputs are the same as* catcheR*_10Xcatch*.

153 In the same working folder used to run *catcheR_10Xcatch*, run *catcheR_filtercatch*, which uses the output of *catcheR_10Xcatch* to filter cells based on custom thresholds.

```
catcheR_filtercatch(
    group=c("docker","sudo"),
    folder,
    expression.matrix,
    UMI.count,
    percentage = 15,
    ratio = 5,
    sample = 1)
```

153.1 Example usage:

```
catcheR_filtercatch(group = "docker",
    folder = "path/to/working/folder",
    expression.matrix = "filename.csv",
    UMI.count = 5, sample = 1)
```

The function is to be run separately for each sample (specified by argument sample with default 1). *catcheR_filtercatch* argument percentage *i*s the minimum percentage of UMI for a given UCI over the total UMIs of UCIs in that cell, to consider the UCI valid. UMI.count *i*s an integer of the custom UMI threshold. ratio is the minimum ratio between UMIs of top UCI and UMIs of 2nd top UCI. The other arguments and the outputs are the same as *catcheR_10Xcatch*.

OPTIONAL: identify cells expressing no shRNA in iPS2-10X-seq

154 In the same working folder used to run catcheR_*1*0Xcatch, run catcheR_10Xnocatch:

```
catcheR_10Xnocatch(
  group=c("docker","sudo"),
  folder,
  expression.matrix,
  threshold,
  sample = 1,
  reference = "TAGCCGTTCACTCGGGGACGCG")
```

154.1 Example usage:

```
catcheR_10Xnocatch(group = "docker",
  folder = "path/to/folder",
  expression.matrix = "filename.csv",
  threshold = 5)
```

154.2 catcheR_10Xnocatch argument threshold is an integer of the minimal UMI count for the "empty" reference (Figure S2F) needed to confidently identify it as having integrated said plasmid. All other arguments are the same as catcheR_10Xcatch.

catcheR_10Xnocatch output is an updated "silencing_matrix.csv", also available in RDS format, in which names of empty cells are modified as "cellID__?_empty_NA_empty".

OPTIONAL: merge multiple samples

155 In case fine-tuning is applied after catcheR_10Xcatch and multiple samples are present, run catcheR_merge to aggregate the results in a single matrix. This step is done automatically when running catcheR_10Xcatch.

```
catcheR_merge(
  group=c("docker","sudo"),
  folder,
  samples = 2,
  empty = T)
```

155.1 The argument "empty" determines whether the cells identified by catcheR_nocatch should be added to the matrix.

155.2 Example usage:

```
catcheR_merge(group = "docker",
  folder = "/20tb/ratto/catcheR/test_CM5/",
  samples = 4)
```

iPS2-sci-seq perturbation deconvolution

156 This section describes a typical analysis pipeline similar to the one described in the previous section, but leveraging on the catcheR*scicatch, catcheR*scicatchQC, catcheR*filtercatch, and catcheR*scinoactch functions.
In a new working folder:

Create a subfolder called "fastq", and copy all the demultiplexed fastq.gz files (step 2 of Obtain iPS2-sci-seq count matrices). Ensure that all file names begin with the well coordinate (e.g. A01)

157 Copy the cell-by-gene expression matrix CSV file obtained with catcheR_scicount function (step 4 of Obtain iPS2-sci-seq count matrices)

157.1 Create a CSV file called "rc*barcodes*genes.csv" with two columns: (1) the shRNA BCs; (2) the matching shRNA names in the format "GENE.shRNAID"

157.2 Copy the txt file "sci-RNA-seq-8.RT.oligos" used by catcheR_scicount (step 3b of Obtain iPS2-sci-seq count matrices)

158 Run *catcheR_scicatch* to execute the complete analysis with automatic thresholding; the arguments are the same as for *catcheR_10Xcatch* (step 2 of iPS2-10X-seq perturbation deconvolution), except that no filenames are provided since fastq files must be in the "fastq" subfolder in the working directory:

```
catcheR_scicatch(
  group=c("docker","sudo"),
  folder,
  expression.matrix,
  reference = "GGCGGCTTCATCTGGGGGACGCG",
  UCI.length = 6,
  threads = 2,
  percentage = 15,
  ratio = 5,
  mode = "bimodal",
  x = 100, y = 400)
```

158.1 Example usage:

```
catcheR_scicatch(group = "docker",
 folder = "path/to/working/folder",
 expression.matrix = "filename.csv",
 threads = 12)
```

158.2 catcheR_scicatch key outputs are the same as for catcheR_10Xcatch (step 2 of iPS2 10X-seq perturbation deconvolution and Figure SP4.4), except that in "silencing_matrix.csv" "cell ID" indicates the PCR well and RT barcode ID (which can be leveraged to identify nuclei from different samples pooled on the same RT plate) Example of cell name after the analysis:
P24__RT_27_7_GCCTGTGT_SCR_ACGGTG

158.3 In addition, catcheR_scicatch provides two quality control outputs to evaluate experimental biases during sci-RNA-seq library preparation: demux and RT detail about how many cells were identified from each row and column of the PCR and RT plates, respectively

OPTIONAL: fine tune iPS2-sci-seq perturbation assignment

159 Run *catcheR_scicatch* to execute the complete analysis with automatic thresholding; the arguments are the same as for catcheR_10Xcatch (step 32.2), except that no filenames are provided since fastq files must be in the "fastq" subfolder in the working directory:

```
catcheR_scicatch(
  group=c("docker","sudo"),
  folder,
  expression.matrix,
  reference = "GGCGCTTCATCTGGGGGACGCG",
  UCI.length = 6,
  threads = 2,
  percentage = 15,
  ratio = 5,
  mode = "bimodal",
  x = 100, y = 400)
```

159.1 Example usage:

```
catcheR_scicatch(group = "docker",
  folder = "path/to/working/folder",
  expression.matrix = "filename.csv",
  threads = 12)
```

159.2 catcheR_scicatch key outputs are the same as for catcheR_10Xcatch (step 2 of iPS2-10X-seq perturbation deconvolution and Figure SP4.4), except that in "silencing_matrix.csv", "cell ID" indicates the PCR well and RT barcode ID (which can be leveraged to identify nuclei from different samples pooled on the same RT plate).
Example of cell name after the analysis:

```
P24__RT_27_7_GCCTGTGT_SCR_ACCGTC
```

In addition, catcheR_scicatch provides two quality control outputs to evaluate experimental biases during sci-RNA-seq library preparation: demux and RT detail about how many cells were identified from each row and column of the PCR and RT plates, respectively.

160 Run catcheR_scicatchQC following the steps described for catcheR_10XcatchQC in *OPTIONAL: fine tune iPS2-10X-seq perturbation assignment*: the functions share the same arguments and outputs, but catcheR_scicatchQC used the output of catcheR_scicatch.

160.1 Example usage:

```
catcheR_scicatchQC(
  group = "docker",
  folder = "path/to/working/folder",
  mode = "noise")
```

161 Run catcheR_scinocatch following the steps described for catcheR_10Xnocatch in OPTIONAL: identify cells expressing no shRNA in iPS2-10X-seq: the functions share the same arguments and outputs, but catcheR_scinocatch uses the output of catcheR_scicatch.

161.1 Example usage:

```
catcheR_scinocatch(
  group = "docker",
  folder = "path/to/working/folder",
  expression.matrix = "filename.csv",
  threshold = 5,
  reference = "TACGCGTTCATCTGGGGGAG")
```

Barcode reassignment

162 catcheR_sortcatch is an optional function that corrects the annotated cell-by-gene count matrix obtained with catcheR_10Xcatch or catcheR_scicatch reassigning the perturbation of any cell belonging to a HPSC clone with reliable evidence of a shRNA-barcode swap, based on the results of catcheR_step1QC.
In a new working folder:

- (a) Copy the annotated count matrix (i.e., ”silencing_matrix.csv”)
- (b) Copy the CSV file with the list of UCI-BCs with reliable evidence of a shRNA-barcode swap (i.e., "reliable_clones_swaps.csv”; output 2a of Pooled cloning step 1 QC)

163 Run catcheR_sortcatch:

```
catcheR_sortcatch(
  group=c("sudo","docker"),
  folder,
  expression.matrix,
  swaps)
```

163.1 catcheR_sortcatch arguments:

- (a) group: string with two options: sudo or docker, depending on the user group
- (b) folder: string with the working folder path
- (c) expression.matrix: string with the filename of the annotated count matrix CSV
- (d) swaps: a character string with the filename of the txt file (step 1b)

163.2 catcheR_sortcatch output is an updated annotated gene expression matrix CSV file called "silencing_matrix_updated.csv", in which "BC" and "GENE" have been modified to reflect the actual shRNA encoded in each cell.

Perturbation effect analysis

164 The second part of *catcheR* provides an exploratory analysis with visual and statistical summaries to highlight perturbation effects.

Annotation
Before proceeding, the gene expression matrix should be annotated with gene symbols using the *scannobyGtf* function from the R package rCASC. As part of quality control, we recommend evaluating the fraction of ribosomal and mitochondrial reads — e.g., using the *mitoRiboUmi* function from the same package — and considering the exclusion of cells with abnormally high proportions, which may indicate poor quality or stress.
Note: After this step, row names of the matrix (the genes) will have the following format:

```
GeneSymbol:EnsemblID
```

Example:

```
ENSG00000000003:TSPAN6
```

Data loading

165 In a new working folder:

- Copy the count matrix annotated with gene names during the previous step (filtered_*annotated*_silencing_*matrix*_complete_*all*_samples.csv).
- Copy the file rc_*barcodes*_genes.csv described in step 1a of Pooled cloning step 2 and hiPSC genome editing QC.
- Create a new-line separated file listing the control genes (e.g. SCR, *B2M*).
- Create a new-line separated file listing the control samples, if any (e.g. 1,3). These are the sample names also used by *aggr* (see CSV file used as input for *aggr*).
- Create a plain text file listing the sample replicate labels, one per line (e.g., batch1, batch1, batch2, batch2). The order of entries must match the order of samples in the input matrix exactly. This file is required for downstream batch-aware analyses.
- If the dataset includes samples from different experiments or processing batches, batch correction is recommended.
- Create a CSV file listing each sample along with its corresponding annotation name, which will be used for display instead of the sample number (this file is mandatory). Example of CSV file to download on GitHub.
- Create a newline-separated file listing the genes of interest whose expression you wish to visualize on the UMAP (optional).

166 Run catcheR_load:

```
catcheR_load(
group="docker",
folder,
expression.matrix,
control_genes,
control_samples = NULL,
replicates = NULL,
sample_names,
resolution = 8e-4,
genes = NULL)
```

166.1 Example usage:

```
catcheR_load(
group="docker",
folder="/path/to/working/folder/",
expression.matrix =
"annotated_silencing_matrix_complete_all_samples.csv",
control_genes = "controls.txt",
control_samples = "noTET.txt",
replicates = "replicates.txt",
sample_names = "samples.csv",
resolution = 8e-4,
genes = "genelist.txt")
```

166.2 The argument _resolution_ refers to the resolution parameter used by Monocle's _cluster_cells_ function for clustering.

166.3 catcheR_load outputs:

- 1. expression_data.csv and cell_metadata.csv, which can be used to create a Monocle Cell Data Set (CDS) and are also included in the ready-to-load R object starting_cds.Rdata
- 2. UMAP.pdf plots the dimensionality reduction and UMAP*gene*expression.pdf shows the gene expression on the UMAP of the genes provided by the argument "genes"
- 3. UMAP_clustering.pdf shows the clustering obtained on the UMAP with the provided resolution
- 4. processed_cds.RData is the Cell Data Set after normalization, dimensionality reduction, clustering and calculation of trajectories

166.4 At the end of this step, the data are stored in a format compatible with the monocle3 package. However, you can switch to other single-cell analysis frameworks such as Seurat or Scanpy at any point in the workflow. *Note*: For standard iPS2-seq perturbation analysis, always continue using the CDS object generated by *catcheR*.

```
library(SeuratWrappers)
library(Seurat)
seurat = as.Seurat(cds, assay = NULL)
scanpy_sce = as.SingleCellExperiment(seurat)
```

166.5 The follow up analysis are: catcheR_pseudotime, catcheR_modules and catcheR_enrichment.

- With the *Pseudotime* function, *catcheR* calculates the cumulative frequency of cells sharing the same gene, shRNA, or clone, compared to each control, along the pseudotime trajectory. *catcheR* can compare the cumulative frequency curves by computing directional Kolmogorov-Smirnov statistics between the target and control
- With the *Genes modules* function, *catcheR* identifies gene modules within the dataset and computes the cumulative frequency of cells with the same gene, shRNA, or clone —relative to each control—based on the expression of each gene module. Also in this case, cumulative frequency curves are compared using a directional Kolmogorov-Smirnov test
- With the *Enrichment / depletion analysis* function, catcheR calculates the enrichment of cells with the same gene, shRNA, or clone within clusters, in comparison to each control. catcheR can compare each target and control quantity and obtain statistics using the Fisher exact test

Pseudotime

167 Since Monocle pseudotime calculation requires the user to select the starting point interactively, the first step of the pseudotime analysis needs to be done separately within monocle3.
After loading the processed_cds.RData in R, calculate the pseudotime trajectory and plot it with the R script below:

```
library(monocle3)
cds = order_cells(cds)
pt = as.data.frame(pseudotime(cds))
names(pt) = c("pseudotime")
write.csv(pt, paste0(dir, "/pseudotime.csv"))

plot_cells(cds, color_cells_by = "pseudotime",
label_cellgroups=FALSE, label_leaves=FALSE,
label_branch_points=FALSE, graph_label_size=1.5)
```

168 Run catcheR_pseudotime:

```
catcheR_pseudotime( group=c("docker","sudo"), folder, cds,
pseudotime, all = FALSE)
```

168.1 Example usage:

```
catcheR_pseudotime(
group="docker",
folder="/path/to/working/folder/",
cds = "processed_cds.RData",
pseudotime = "pseudotime.csv")
```

168.2 Argument "all" is a logical operator indicating whether to perform the Kolmogorov-Smirnov test against the controls together (all = F) or against each control separately (all = T). The default argument is false

168.3 Below is a list of the catcheR_pseudotime outputs:

- (a) The "cumulative*frequency*pseudotime" plots show the number of cells at each given point of the pseudotime for different groups of cells at different comparison levels - gene, shRNA, and clone
- (b) The "ks_statistics" CSV files that include the Kolmogorov-Smirnov test results comparing cumulative the frequency based on the pseudotime between knockdown and different controls
- (c) The Volcano plots of the results of the Kolmogorov-Smirnov test show the significance and fold change between the pseudotime cumulative curves. Different controls and levels are used
- (d) The "correlated_pseudotime_gene_exp.pdf" showing the expression on the UMAP of the genes most correlated with the pseudotime

Gene Modules

169 This function uses Monocle to find gene modules which expression varies in different perturbation groups, i.e., perturbed gene or shRNA or clones.

Run catcheR_modules:

```
catcheR_modules(group=c("docker", "sudo"), folder, cds,
resolution=1e-2)
```

169.1 Example usage:

```
catcheR_modules(group="docker",
folder="/30tb/3tb/data/ratto/testing/", cds =
"processed_cds.RData")
```

169.2 The resolution parameter refers to the value used in Monocle's find_*gene*_modules function, which influences the number of gene modules identified and the number of genes included in each module (higher resolution typically results in more, smaller modules)

169.3 Below are listed the outputs of catcheR_modules:

- (a) gene_modules.csv, listing the genes present in each module
- (b) Heatmap plots display the Z-scores of each module, either all modules or the top 10 most variable ones, across perturbation groups
- (c) The "modules*cells" folder contains tables with aggregated module expression values for each cell. These can be used as input for catcheR*pseudotime in place of the pseudotime CSV file, allowing the analysis of cumulative frequency based on module expression

169.4 Run catcheR_pseudotime on the CSV files stored in the "modules_cells" folder to use the data generated by catcheR_modules as described in the previous example in the Pseudotime section

Enrichment / depletion

170 This function evaluates whether perturbation groups are enriched or depleted in cells (number) or in specific cell subpopulations (clusters).

Run catcheR_enrichment:

```
catcheR_enrichment(
group=c("docker","sudo"),
folder,
file,
meta,
timepoint = "PSC",
control_gene = "SCR",
min_cells_cluster = 70,
min_cells_shRNA = 40)
```

170.1 Example usage:

```
catcheR_enrichment(
group = "docker",
folder = "/3tb/data/ratto/aggr/test/",
file = "processedcds.RData",
meta = "cell_metadata.csv",
timepoint = "PSC",
control_gene = "SCR")
```

170.2 The required input is the cds file generated from catcheR_load. Timepoint refers to the baseline time point used as a control for statistical analysis. It is required when experiments include multiple time points and aim to assess enrichment or depletion across these time points. control_gene specifies the control gene used as a reference for statistical comparisons.

171 Below are the outputs of catcheR_enrichment:

- (a) Plots of cells in each perturbation group
- (b) Volcano plot showing enrichment or depletion of cell numbers in perturbation groups compared to the control, based on fold-change ($\log_2$ ratio) and statistical significance. A corresponding bar plot displays the $\log_2$ fold-change for each perturbation
- (c) Barplots showing the distribution of cells from different perturbation groups in the clusters
- (d) Volcano plot showing the results of Fisher's exact test, comparing the distribution of cells from perturbation groups across Monocle-derived clusters. The plot displays $-\log_{10}$ adjusted p-values *versus* $\log_2$ fold-changes relative to the control group
- (e) Table with Fisher's statistics