Table of contents

1. Single-cell quantitative expression reporters (scQers)

To use multiplex reporter assays such as MPRA in biological contexts that extend beyond static cell lines towards multicellular systems requires combination with single-cell genomics modalities.

In order to quantitatively capture reporter expression with single-cell assays in the context of pooled screens, one needs to overcome the problem of determining which cells contains which reporters (much the same way one needs to identify which cell received which sgRNA in the context of CRISPR perturb-seq screens).

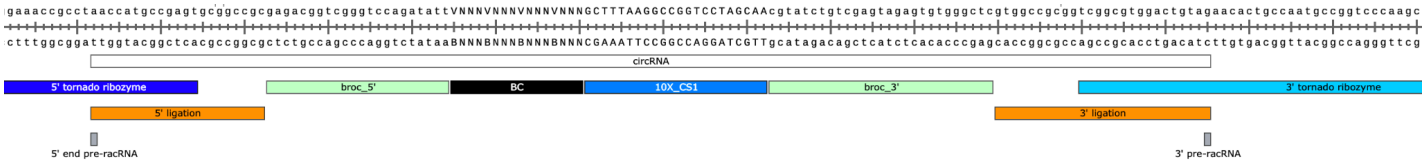
Relying on capture of the BC from the traditional Pol II mRNA (whose expression level encodes the activity of the CREs under investigation) is unsatisfactory for two reasons:
1) in some instances, the CRE will be completely inactive in certain cell types. How can one then detect an RNA that is not produced?
2) even in cases in which the reporter is expressed, conflating detection and quantification runs the risk of biasing the measurement.

To overcome this problem, we devised a two RNA reporter cassette, which harbors a constitutively expressed (U6) small barcoded RNA, used to detect presence/absence of reporters in cells. This is in addition to the usual MPRA reporter transcript. The short detection barcodes are circularized to improve stability and capture in single-cell (see note below on these ‘Tornado barcodes’), with substantial boost in performance compared to direct capture of other RNAs.

In these single-cell quantitative expression reporters (scQers, pronounced ‘skewers’ 🍢), triplet composed of the detection BC, the CREs to be tested, and the quantification BC of the Pol II mRNAs are pre-associated upstream during the cloning into a dictionary enabling signal deconvolution, and delivered to cells.

Interested researchers are encouraged to consult the detailed methods section (for information beyond cloning libraries and performing the pre-association) of the accompanying manuscript, and to reach out to us for questions.

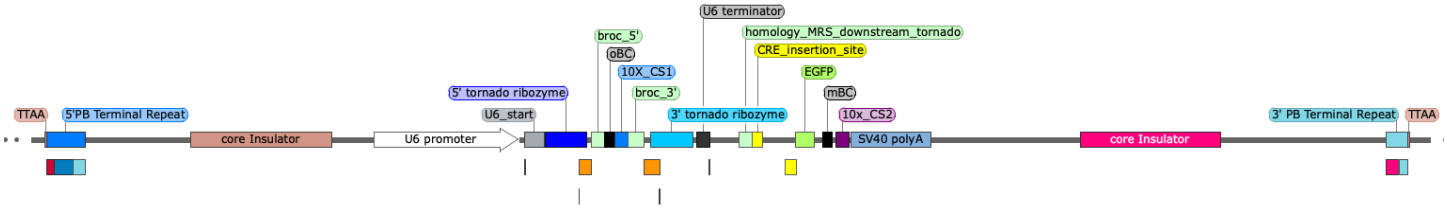
Side note on Tornado barcodes (oBC):
Tornado barcodes have 16 nt (5'VNNNVNNNVNNNVNNN, Vs prevent strings of 4 Ts that would lead to substantial termination[1] in about 4% of all Ns barcodes) barcodes and capture sequence 1 (CS1: 5' GCTTTAAGGCCGGTCCTAGCAA) inserted within the loop of the original Tornado-Broccoli cassette from Litke and Jaffrey[2].



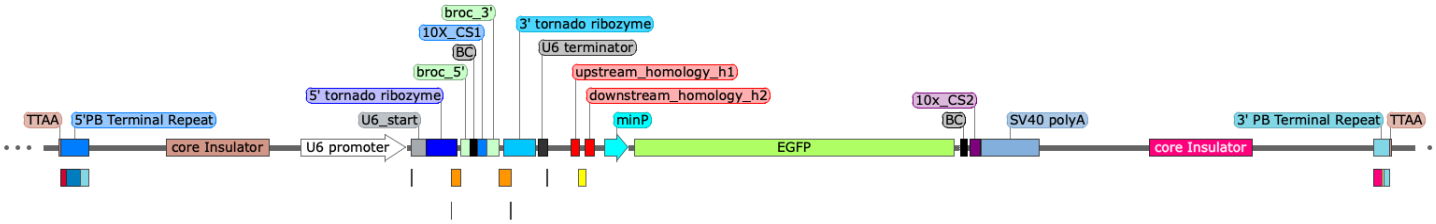
2. Available plasmids

Plasmid constituting cloning intermediates:

p025: paired barcode plasmid without reporter fragment integrated, with restriction sites to introduce reporter (BglII+EcoRI, see p043) and CRE libraries (MfeI+NheI, see below)



p043: paired barcode plasmid with minP-GFP reporter integrated.



p043 is the fastest starting point to generate your own library to measure cell-type specific activity of candidate regulatory elements (see schematic below). See section below for a description of the cloning approach.

We provide a complex library of p043 with pre-associated barcodes (complexity ~1M barcodes).

The list of oBC-mBC pairs can be found on GEO (here, accession: GSE217681, file name: GSE217681_p025_recloned_complex_mBC_oBC_subassembly.txt.gz).

Control plasmids (promoter series):
Directly usable as internal controls with libraries of interest and for test & optimization in new biotypes. These libraries are of relatively low complexity (~200 oBC-mBC triplet per library), but sufficient to perform proof-of-concept quality control runs.

p027: hU6-oBC_minP-puro-GFP-mBC
p028: hU6-oBC_EEF1A1p-puro-GFP-mBC
p029: hU6-oBC_noP-puro-GFP-mBC
p041: hU6-oBC_UBCp-puro-GFP-mBC
p042: hU6-oBC_Pgk1-puro-GFP-mBC

minP and noP serve as negative (basal expression) controls.
Pgk1 and UBC promoters are of intermediate (though still quite high) expression.
EEF1A1 promoter has consistently been the highest expression promoter in tested cells (hundreds of UMI per barcode per cell from scQers measurement with 10X genomics in both cell lines and stem-cell derived multicellular embryoids).

List of barcodes: full table of uniquely oBC-promoter-mBC triplet (a small proportion of BC pairs were not uniquely matched) can be found on GEO (here, accession: GSE217681, file name: GSE217681_final_subassembly_oBC_promoters_mBC_triplets_promoter_series.txt.gz).

All the above are inside a piggyBac transposon with flanking cHS4 insulator sequences.

For random transgenesis, we have found good performance with hyperactive piggyBac transposase (using the pCMV-hyPBase construct from here[3]).

3. Cloning approach

3.1. Plasmid amplification

Plasmid libraries p025 and p043 contain approximately 1M barcode pairs, pre-associated to facilitate library generation. It is crucial to maintain the library complexity until the last bottlenecking step as described below.

Any approach used to transform plasmids at high efficiency and high complexity should be applicable.
We used the following approach:

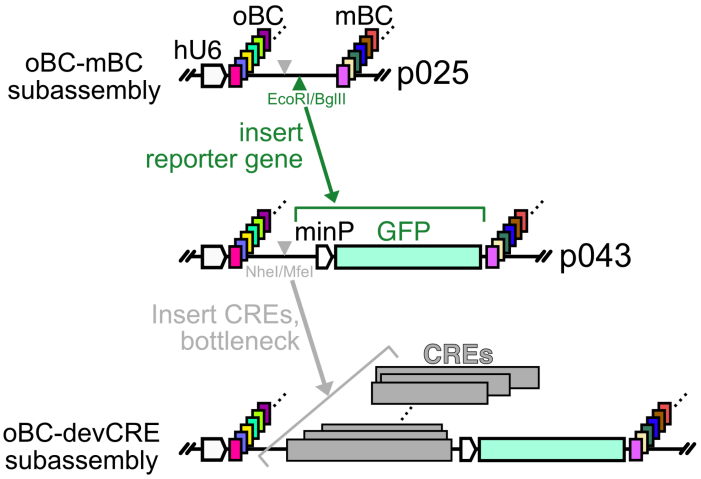
- Thaw 25 uL of NEB's 10-beta cells (C3020) on ice.
- Add 10 ng of p043 plasmid (diluted in distilled water) in <3 uL.
- Flick the tube to mix while avoiding bubbles.

- Transfer cells and plasmid mix to a pre-chilled (ice) 1 mm cuvette for electroporation (this for example), very carefully avoiding bubble (which can arc the electroporation, if necessary to avoid bubbles, pipette <25 uL). Keep the cuvette with cells on ice.
- Electroporate (e.g., Biorad gene pulser, 2 kV, 200 Ω, 25 uF). Tap the cuvette to eliminate bubbles and wipe ice/water from the cuvette with absorbent paper prior. Time constant should be in the range 4.0 to 4.3 ms. Short time constant with spark indicates problematic discharge. If this occurs, repeat, decreasing the amount of plasmid and pay attention to bubbles.
- Immediately after successful electroporation, add 475 uL recovery medium (e.g., SOC), transfer to a 1.5 mL tube, and outgrow at 37C with shaking.
- Serially dilute the electroporation, and plate 0.1% of the transformation on ampicillin plates to assess transformation efficiency.
- You can either keep the electroporated cells at 4C until high efficiency is confirmed, or outgrow overnight in LB with ampicillin (typically in 25-50 mL in 250 mL flask, 37C, 200 rpm on orbital shaker).
- Upon confirmation of high efficiency (you should see >1000 of colonies in the 0.1% plate, corresponding to >1M transformant), make a glycerol stock for future use, and purify plasmids by mini or midi prep as appropriate for downstream cloning (you will need a few micrograms

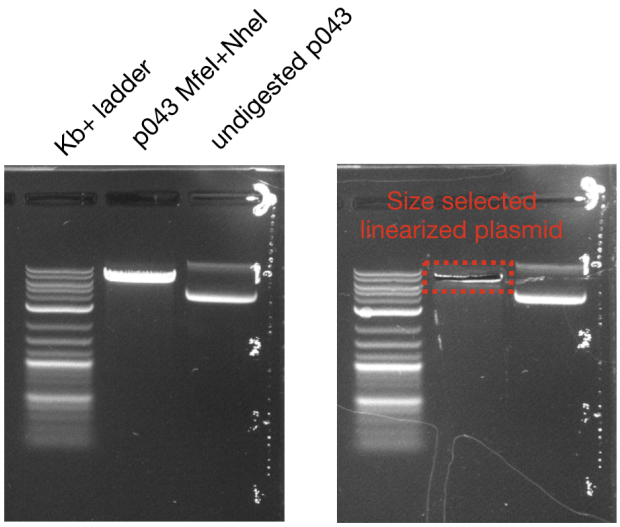
Note: other electrocompetent cells would presumably also work, but we caution that the flanking cHS4 sequences could cause recombination in less stable genetic bacterial background (e.g., 10-beta are *recA* null). We have not tested other strains.

3.2. Library generation from p043

We describe the approach starting from dually barcoded plasmid library p043, which already has minP-GFP integrated between the two barcodes:



Restriction sites for MfeI and NheI are present between the hU6-oBC and minP region of the plasmid, enabling introduction of a library of regulatory elements. Following double digestion of a few micrograms of p043, size select the linear product on agarose (band between 6 and 7 kb) and elute in low volume (e.g., with Zymoclean™ Gel DNA Recovery Kit).

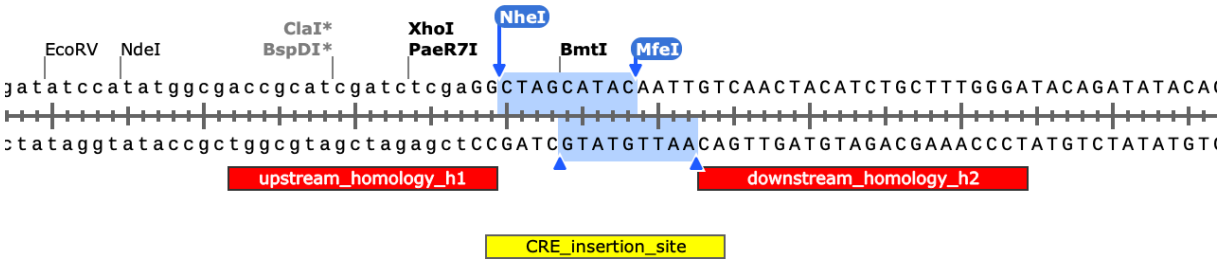


This p043_MfeI_NheI digested fragment is the backbone for a one-insert isothermal assembly to insert the library of enhancers/regulatory elements.

The library to be cloned needs to have the following handles (red) for isothermal assembly:

towards oBC ← 5’ accgcatcgatctcgagg(CRE SEQUENCE)gtcaactacatctgctttggga → towards minP

See below the portion of p043 with the insertion site:



For the isothermal assembly, we typically perform the following reaction:

4 uL total reaction volume with:

- 2 uL 2x NEBuilder HiFi DNA Assembly master mix
- 1 uL p043_MfeI_NheI size selected product (0.02 pmole, or about 100 ng).
- 1 uL CRE library (with handles as above) in 2-4 fold molar excess (mass depends on mean size)

You can perform a no insert negative control to confirm the quality of your digest and size selected products, and should see very little background (at least 100-fold fewer colonies compared to with insert reaction).

Following thorough mixing by pipetting, the assembly is incubated in a PCR tube at 50C for 60 min.
The reaction is then cleaned up with column purification (e.g., Zymo DNA Clean and Concentrator with 3:1 binding buffer), and eluted in 6 uL water (important: eluting in buffer can lead to arcing in electroporation).

High efficiency electroporation in *E. coli* proceeds as described above, with 2.5 uL of the cleaned up isothermal assembly as the material to electroporate. Successful isothermal assembly and electroporation should yield ~1M transformants.

An important point is that the final target number of transformants in your library needs to be tuned based on your application in a bottlenecking process. Without bottlenecking, multiple CREs will be associated with the same oBC-mBC pair because of the way the library is constructed.

In order to generate statistically unique oBC-CRE-mBC triplets, the final number of plasmids in your library should be much smaller than the total number of oBC-mBC pairs in the parental p043 library, which can be forced by limiting the final number of transformant in your pool. Based on simulations from the barcode representation, starting from p043, libraries comprising <80k final number of oBC-CRE-mBC (unique transformants) should harbor largely unique triplets (>90% unique).

The final complexity of your library should be determined by the number of CREs profiled. We recommend about 100 oBC-mBC pairs per CRE, so that the number of transformants desired can be estimated as (# CREs)×100. For example, with 500 CREs tested, you would want a target complexity of about 50k transformants.

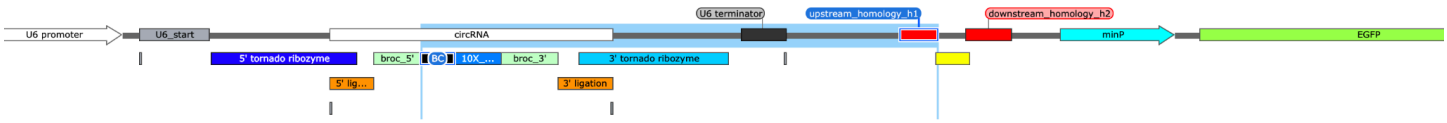
To titrate library complexity, serially dilute your electroporation (after the 60 min outgrowth at 37C), typically in steps of 5 or 10-folds, and plate a small proportion of a few of the dilutions to estimate complexity. You can keep the dilutions at 4C overnight, and determine which dilution to outgrow at scale to purify plasmids based on the desired complexity (see above note). The final purified plasmid library with the CREs inserted can then serve as template for the final step of dictionary generation.

4. CRE-barcode dictionary generation

4.1. Barcode association library generation

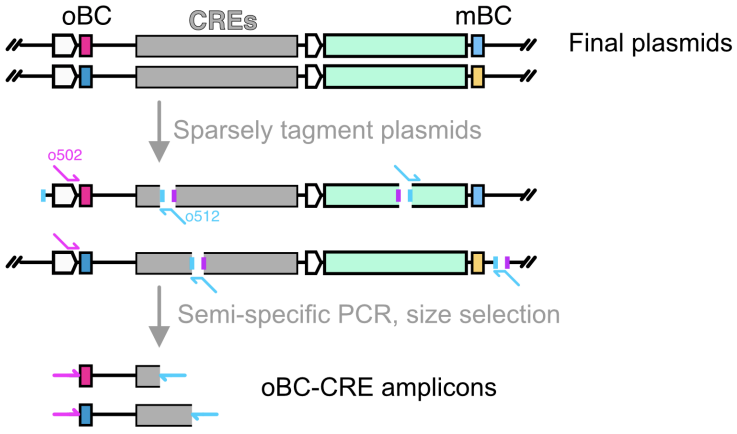
Library p043 already contains pre-associated oBC and mBCs (here, GSE217681_p025_recloned_complex_mBC_oBC_subassembly.txt.gz). To finalize the oBC-CRE-mBC triplet dictionary, one needs to associate oBC to CREs for the libraries generated in the step above. This can be performed with standard Illumina short read platforms with paired end sequencing (e.g., Nextseq).

We distinguish two regimes, depending on the typical size of the CREs in the library. Given that clustering on the flow cell becomes highly inefficient for amplicons of around 1kb, and that the constant region (without adapters) between the CRE and the oBC (see close up below of p043 prior to CRE integration) is about 250 bp, it becomes difficult to rely on PCR amplification to generate libraries sequenceable on the short read platforms.



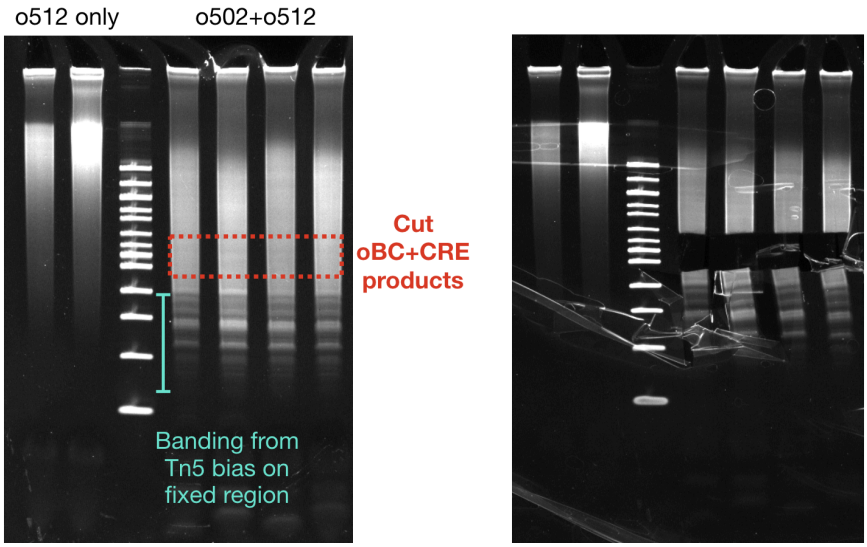
4.1.1. Case 1: long (>750 bp) regulatory elements (tagmentation + semi-specific PCR)

To associate long CREs to barcodes, we tagment the plasmid with Tn5, perform semi-specific PCR, and size select. The procedure is illustrated schematically (which does not show all the possible products resulting from tagmentation and PCR) below:

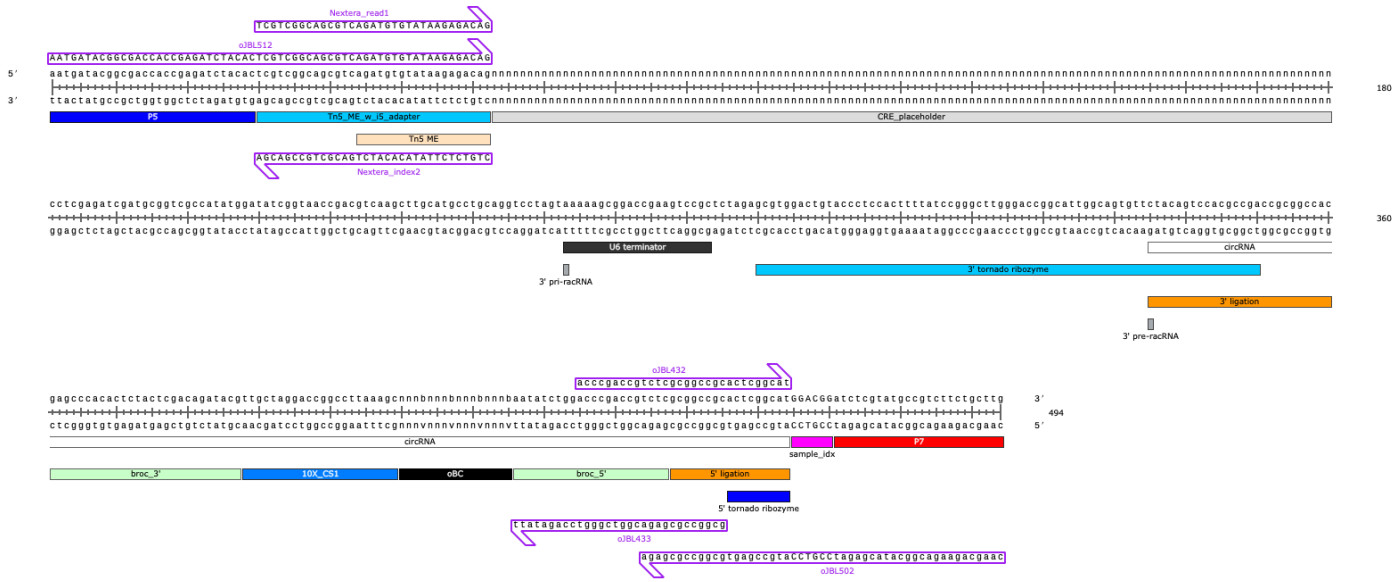


Briefly, the scQer plasmid library is tagmented with Tn5 (Illumina, Nextera Tagment DNA enzyme, cat. no. 15027916) at a concentration such that the expected fragment size would be larger than the oBC to minP distance, determined by a Tn5 titration experiment. The treatment described below was determined for CRE size of about 1 kb. Following tagmentation (5 uL 2x Tagmentation DNA buffer [Illumina, cat. no. 15027866], 0.4 uL Tn5 enzyme 1, 3.6 uL water, 1 uL 10 ng/uL plasmid library; 30 min at 37C), the tagmented plasmids were cleaned up (Zymo, clean and concentrator, 3:1 binding buffer), eluted in 10 uL Tris 8 10 mM. 1 ng (1 uL of the elution) was amplified via semi-specific PCR with a Nextera primer with a P5 handle (oJBL512, binding to all P5 tagmentation events) and a oBC-specific upstream primer (oJBL502, binding to specific portion of the plasmid) in 25 uL (8.9 uL water, 12.5 uL 2x NEBNext master mix, 1.25 uL 10 uM oJBL502, 1.25 uL oJBL512, 1 uL tagmented plasmids, 0.1 uL 200x SYBR green) with the following conditions (gap fill: 72C for 5 min, 98C for 30 sec, then 12 cycles of 98C for 10 sec, 65C for 30 sec, 72C for 1 min). As controls for the non-specific product size distribution, the tagmented plasmids were also amplified with oJBL512 exclusively. Following purification (Zymo, clean and concentrators), the amplified libraries were run on PAGE (6% TBE, 180V, 30 min). The amplicons with primers oJBL502+oJBL512 (semi-specific products) displayed reduced size distribution compared to oJBL512 alone amplified (non-specific) products, with most oJBL512 exclusive amplicons >1.2 kb. Semi-specific oJBL502+oJBL512 products between 450 bp and 800 bp were size selected on the PAGE gel, purified (minimum size size from CRE ≈ 75 bp), and sequenced (read 1: CRE sequence, Illumina Nextera primer [no custom], 34 cycles; index 1: P7-idx, primer oJBL432 15 cycles; read 2: oBC, primer oJBL433, 30 cycles).

Example of a PAGE size selection post semi-specific PCR:



Example sequencing amplicon:



Sequencing primers:

Read1: CRE (tagmentation position), Nextera read1 (Illumina primers, no custom needed). As many cycles as needed for unique mapping.

Index1: sample/library index with primer oJBL432. 6 cycles.

Read2: oBC with primer oJBL433. 16 cycles.

Primer sequences:

oJBL432 5’ acccgaccgtctcgcggccgcactcggcat

oJBL433 5’ gcggccgcgagacggtcgggtccagatatt

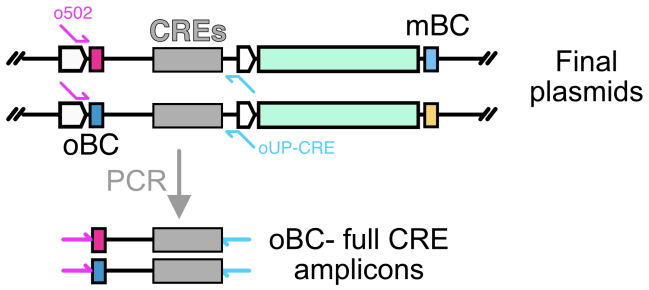
oJBL502 5’ caagcagaagacggcatacgagatCCGTCCatgccgagtgcggccgcgaga

[can have multiple such indexed primers for the different libraries as required]

oJBL512 5’AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTGTATAAGAGACAG

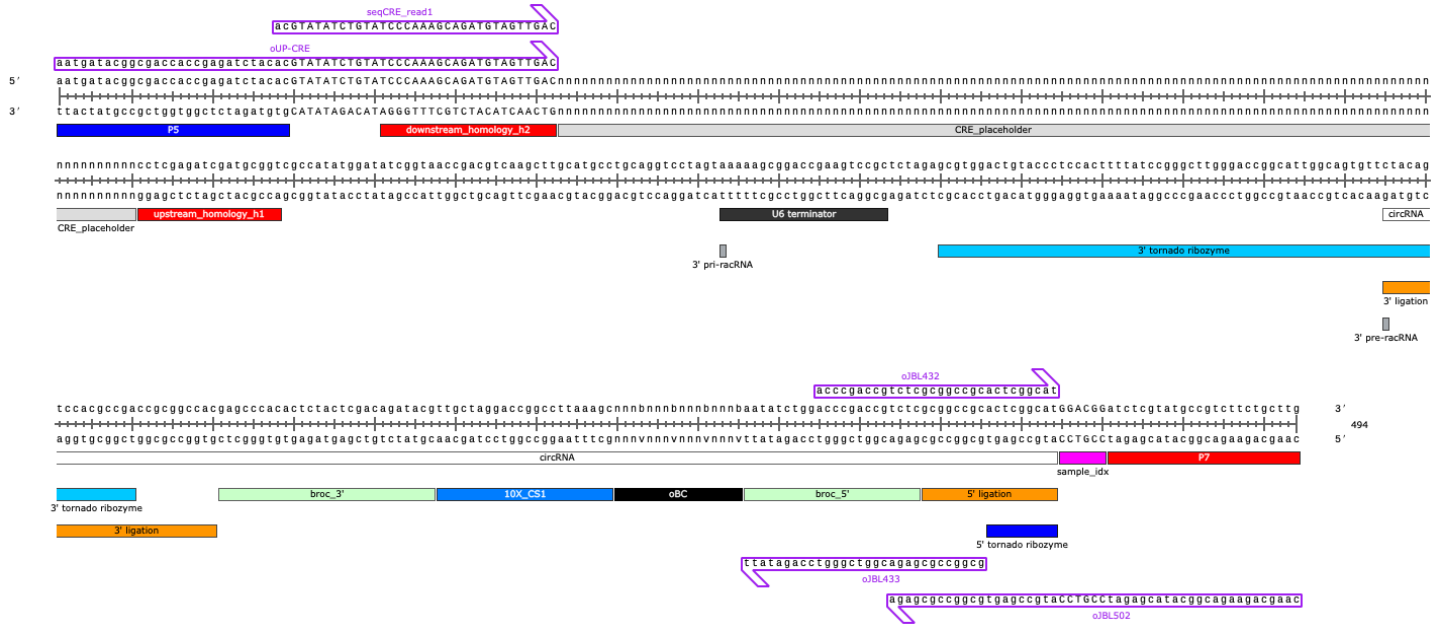
4.1.2. Case 2: short (<750 bp) regulatory elements (PCR)

For libraries of shorter CREs, a simple library generated by PCR from the two constant regions would suffice:



For example, with primers oJBL502 and oUP-CRE.
The ampicon could be sequenced similarly, but now with a read 1 primer coming from the constant region upstream of the CRE.

Resulting amplicon (flipped orientation compared to the schematic above):



Sequencing primers:
Read1: CRE with primer seqCRE_read1, as many as needed for unique mapping.
Index1: sample/library index with primer oJBL432, 6 cycles.
Read2: oBC with primer oJBL433, 16 cycles.
(note, another primer could also be used to sequence the other end of the CRE on the Index2 read).

Primers:

oJBL432	5’ acccgaccgtctcgcggccgcactcggcat
oJBL433	5’ gcggccgcgagacggtcgggtccagatatt
oJBL502	5’ caagcagaagacggcatacgagat**CCGTCC**atgccgagtgcggccgcgaga

[can have multiple such indexed primers, index above in **bold**, for the different libraries as required]

oUP-CRE 5’ aatgatacggcgaccaccgagatctacacGTATATCTGTATCCCAAAGCAGATGTAGTTGAC

seqCRE_read1 5’ GTATATCTGTATCCCAAAGCAGATGTAGTTGAC

We caution that we have not explicitly used this approach, since the tested CREs all had sizes prohibiting clustering of full sized products, and primers above might require optimization, but would likely work.

4.1.3. Verification of oBC and mBC plasmid composition/complexity

In addition to the subassembly procedure, the mBC and oBC complexity in the final library can be confirmed/assessed by PCR amplifying a library from the plasmid. Although rarely strictly necessary, we typically append a pseudo-UMI to the amplicon in a two PCR steps procedure, as detailed below for the two barcodes.

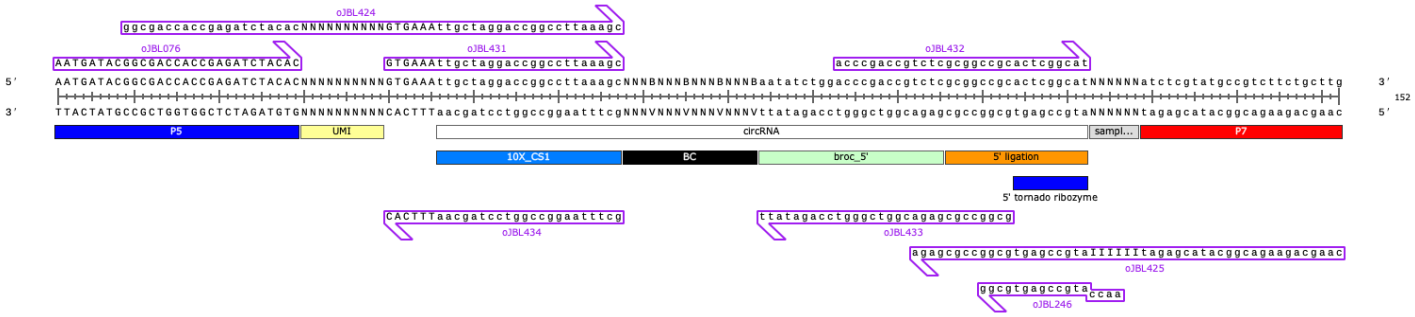
<u>oBC amplicon:</u>

Amplicon libraries from DNA were generated in two steps of PCR amplification with Kapa HiFi (Roche). 3 ng input plasmid was used.

PCR1: For low-cycle number PCR1, template DNA was mixed with 50 µL 2× Kapa HiFi master mix, 5 µL 10 µM oJBL246, 5 µL 10 µM oJBL424, and water to 100 µL. Cycling parameters: 1 min at 95C, and 4 cycles of: 20 s at 98C, 20 s at 60C, 30 s at 72C, followed by 4C hold. Primer oJBL424 contains 10 random Ns to serve as a pseudo-UMI (hereafter referred to as UMIs for brevity) to correct for PCR jackpotting. Reactions were cleaned up with Ampure XP beads (Beckman Coulter) at 1.75×, and eluted in 20 µL of 10 mM Tris 8.

PCR2: Illumina adapters and sequencing indices were appended through PCR2, with 4 µL of the eluate from PCR1 taken as input, and 25 µL 2× Kapa HiFi master mix, 0.25 µL 100× SYBr green, 2.5 µL 10 µM oJBL076, 2.5 µL 10 µM indexed primers (e.g., oJBL501), and water to 50 µL. Libraries were amplified with tracking by qPCR with: 1 min at 95C, and cycles up to the qPCR inflection point (typically 15-17 cycles) of: 20 s at 98C, 20 s at 60C, 30 s at 72C. Libraries were then cleaned up with Ampure XP beads at 1.75×.

We note that for the purpose of simple confirmation of library complexity, appending a pseudo-UMI is not strictly necessary, and a single step PCR library generation with primers oJBL425 (P7-sample idx) and something like CS1-P5 (aatgatacggcgaccaccgagatctacacGTGAAAttgctaggaccggccttaaagc) should work and can be sequenced with single-end read run.



Sequencing:
Read1 (barcode forward): 16 cycles with primer oJBL431
Index1 (index): 6 cycles with primer oJBL432
Read2 (barcode reverse): 16 cycles with primer oJBL433
Index2 (UMI): 10 cycles with primer oJBL434.

Primers:

oJBL246	5’ aaccatgccgagtgcgg
oJBL424	5’ ggcgaccaccgagatctacacNNNNNNNNNNGTGAAAttgctaggaccggccttaaagc
oJBL076	AATGATACGGCGACCACCGAGATCTACAC
oJBL501	caagcagaagacggcatacgagat**ATGTCA**atgccgagtgcggccgcgaga

[can have multiple such indexed primers, index above in **bold**, for the different libraries as required]

oJBL434	5’ gctttaaggccggtcctagcaaTTTCAC
oJBL433	5’ gcggccgcgagacggtcgggtccagatatt
oJBL432	5’ acccgaccgtctcgcggccgcactcggcat
oJBL431	5’ GTGAAAttgctaggaccggccttaaagc

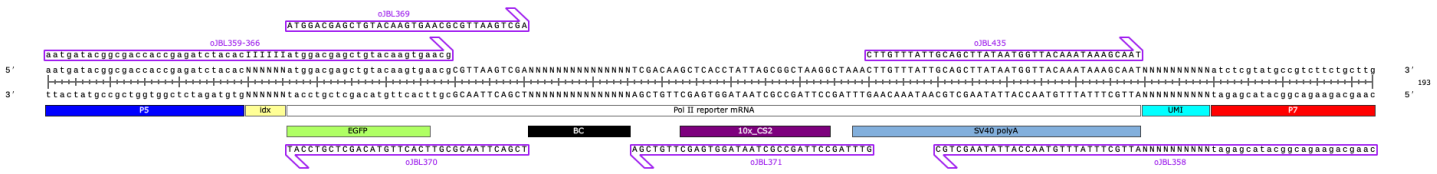
<u>mBC amplicon:</u>

Libraries from DNA were generated in two steps of PCR amplification with Kapa HiFi (Roche). 5 ng of plasmid DNA was used.

PCR1: For low-cycle number PCR1, DNA was mixed with 50 µL 2× Kapa HiFi master mix, 5 µL 10 µM oJBL039, 5 µL 10 µM oJBL358, and water to 100 µL. Cycling parameters: 1 min at 95C, and 4 cycles of: 20 s at 98C, 20 s at 60C, 30 s at 72C, followed by 4C hold. Primer oJBL358 contains 10 random Ns to serve as a pseudo-UMI (hereafter referred to as UMIs for brevity) to correct for PCR jackpotting. Reactions were cleaned up with Ampure XP beads (Beckman Coulter) at 1×, and eluted in 20 µL of 10 mM Tris 8.

PCR2: Illumina adapters and sequencing indices were appended through PCR2, with 4 µL of the eluate from PCR1 taken as input, and 25 µL 2× Kapa HiFi master mix, 0.25 µL 100× SYBr green, 2.5 µL 10 µM oJBL077, 2.5 µL 10 µM indexed primers (e.g., oJBL359), and water to 50 µL. Libraries were amplified with tracking by qPCR with: 1 min at 95C, and cycles up to the qPCR inflection point: 20 s at 98C, 20 s at 60C, 30 s at 72C. Libraries were then cleaned up with Ampure XP beads at 1×.

Similar as with the oBC, we note that for the purpose of simple confirmation of library complexity, appending a pseudo-UMI is not strictly necessary, and a single step PCR library generation with primers oJBL359 (P5-sample idx) and indexed primer on P7 side (similar to oJBL358, but with the Ns replaced by sample indices) should work and can be sequenced with single-end read run.



Note: primer oJBL039 is not visible in the final depiction of the amplicon above as it is upstream in GPF, and primers oJBL359 are nested internally.

Sequencing:
Read1 (barcode forward): 15 cycles with primer oJBL369
Index1 (UMI): 10 cycles with primer oJBL435
Read2 (barcode reverse): 15 cycles with primer oJBL371
Index2 (sample index): 6 cycles with primer oJBL370

Primers:

oJBL039	5' TGAGCAAAGACCCCAACGAG
oJBL358	5' caagcagaagacggcatacgagatNNNNNNNNNNNATTGCTTTATTTGTAACCATTATAAGCTGC
oJBL077	5' CAAGCAGAAGACGGCATACGAGAT
oJBL359	5' aatgatacggcgaccaccgagatctacacATCACGatggacgagctgtacaagtgaacg
oJBL435	5' CTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAAT
oJBL371	5' GTTTAGCCTTAGCCGCTAATAGGTGAGCTTGTCGA
oJBL370	5' TCGACTTAACGCGTTCACTTGTACAGCTCGTCCAT
oJBL369	5' ATGGACGAGCTGTACAAGTGAACGCGTTAAGTCGA

5. Sequencing data analysis of association library

Detailed github sections have been compiled for the different computational tasks, with example datasets to validate simple scripts on.

Links:
Single barcode libraries processing (e.g., assess mBC or oBC composition of plasmids).

Two barcodes association library processing (e.g., determining pairs of associated mBC-oBC in p25).

BC-CRE association library processing (e.g., determining which CREs are matched to which oBCs).

References

1. Gao, Z., Herrera-Carrillo, E. & Berkhout, B. Delineation of the Exact Transcription Termination Signal for Type 3 Polymerase III. *Mol. Ther. Nucleic Acids* **10**, 36–44 (2018).
2. Litke, J. L. & Jaffrey, S. R. Highly efficient expression of circular RNA aptamers in cells using autocatalytic transcripts. *Nat. Biotechnol.* **37**, 667–675 (2019).
3. Yusa, K., Zhou, L., Li, M. A., Bradley, A. & Craig, N. L. A hyperactive piggyBac transposase for mammalian applications. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 1531–1536 (2011).