Cloning and evaluation of puromycin/blasticidin lentiviral static barcode libraries

Lentiviral plasmids expressing puromycin or blasticidin resistance genes were cloned with Gibson assembly following manufacturer’s protocols. We introduced a 10N random barcode that could be efficiently assayed using scRNA-seq-based capture of 3’ polyadenylated transcripts. These libraries were cloned by introducing oligonucleotide libraries ordered from IDT using Gibson assembly. Top and bottom oligo libraries were ordered, cloned, and transformed before libraries were pooled, expanded through outgrowth for 8 hours and Midiprepped. Lentiviral plasmid libraries were sequence confirmed: plasmid libraries were amplified with two successive rounds of PCR as done to sequence genomic loci. PCR1 was performed using Phusion Green Hot Start II High-Fidelity PCR 2x Master Mix with the following cycle conditions: 98ºC for 2 minutes, [98ºC for 10 seconds, 64ºC for 20 seconds, 72ºC for 1 minute] where the number of cycles was empirically determined by qPR, followed by a 72ºC for 2 minutes. PCR2, gel extraction, library preparation, and sequencing were carried out as before.

**Puromycin 10N library top:**

CAGAAAGCCTGGCGCCTGACNNNNNNNNNNGTTTAAACCCGCTGATCAGCCTCGACTG

**Puromycin 10N library bottom:**

CAGTCGAGGCTGATCAGCGGGTTTAAACNNNNNNNNNNGTCAGGCGCCAGGCTTTCTG

**Blasticidin 10N library top:**

TTATGTGTGGGAGGGCTAACNNNNNNNNNNGTTTAAACCCGCTGATCAGCCTCGACTG

**Blasticidin 10N library bottom:**

CAGTCGAGGCTGATCAGCGGGTTTAAACNNNNNNNNNNGTTAGCCCTCCCACACATAA

**Forward primer for puromycin library amplification:**

TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG -[1N to 4N]-

CCTGGTGCATGACCAGAAAGC

**Forward primer for blasticidin library amplification:**

TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG -[1N to 4N]-

GTGAATTGCTGCCCTCTGGTTATGTG

**Reverse primer for puromycin and blasticidin library amplification:**

GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG -[1N to 4N]-

GCACAGTCGAGGCTGATCAG

To assess the balance of library elements, we used FastX toolkit (v0.0.14) to mask low quality bases (<q20) and collapse reads. We removed common flanking sequences and summed reads for unique barcode sequences. Lentivirus was produced from these plasmid libraries as described above (“Lentiviral production”).

Design and cloning of LTC\_v2 library

With three edit site sequences selected, we finalized our lineage tracing cassette design. These cassettes consist of the following sequence organization: 5’-UCOE- EF1a promoter- mCherry- T7 promoter- T3 promoter- PacI digestion site- 30nt sequencing barcodes- sequencing primer- MERFISH decodable intBC- Forward primer for sequencing assays- sequencing intBC- PacI digestion site- Edit Site 1 and flanking sequences- Edit Site 2 and flanking sequences- Edit Site 3 and flanking sequences- Reverse primer for sequencing assays- polyA. For lentiviral construct designs, we included 500nt of constant sequence downstream for optimizing imaging readout experiments and used a BGH polyA sequence.

We generated a lentiviral lineage tracing cassette plasmid library with matching 30nt sequencing barcodes and 183nt MERFISH barcodes (LTC\_v2). To design these orthogonal barcodes, we first randomly selected 9 of the 25nt orthogonal DNA among 240k published DNA barcodes (*151*), concatenated them in an randomly permuted order and selected the best permutation by checking if all 15mer are unique to this barcode. We repeated this process to select up to 6,000 candidate barcodes. We then predicted the self-binding partition function by “pfunc” from Nupack (*149*) and selected the candidate barcodes with highest free energy to minimize self binding. After filtering, we inserted the sequencing primer at position 30 of candidate barcodes and designated the first 30nt as sequencing barcodes, and then selected the next 183nt as MERFISH barcodes. We concatenated the 5’ sequence (T7, T3 and PacI site) and 56nt of 3’ sequence (Edit site 1) together as candidate intBCs for cross-hybridization testing. We first partitioned each 183nt MERFISH barcode into 6 non-overlapping 30nt segments, and used the reverse complement of these segments as candidate probes. To evaluate probe cross-hybridization, we enumerated the pairs of candidate intBCs, generated a Nupack Tube class with the candidate probes and these two candidate intBCs, and predicted the free energy and partition function for all possible bindings. After prediction, we summarized probe cross-hybridization for all candidate intBCs and retained intBCs with less than 1% of predicted average probe off-targeting probability. For barcodes that passed these requirements, we further filtered to have no 17nt matching to human, mouse, zebrafish and drosophila genomes.

As such, we ordered libraries of 2,171 intBCs from Twist Bioscience and amplified this pooled library using Phusion Green Hot Start II High-Fidelity PCR 2x Master Mix with the following cycle conditions: 98ºC for 2 minutes, [98ºC for 10 seconds, 65ºC for 20 seconds, 72ºC for 1 minute] where the number of cycles for amplification were determined by quantitative PCR, followed by 72ºC for 2 minutes. Given that libraries can be provided with variable quality, this qPCR determination was important for ensuring library quality. Following library amplification, samples were cleaned up with a QIAquick PCR purification column (Qiagen) according to manufacturer’s protocols. These amplified DNA fragments were assembled into PacI-digested and Quick CIP-treated (NEB) acceptor backbone via Gibson assembly using NEBuilder 2x HiFi DNA Assembly master mix as large scale 100 μL reactions described above. The entire 100 μL of Gibson assembly product was cleaned up using PCR cleanup columns (Qiagen) to remove salts prior to electroporation into MegaX cells as previously described. Colonies were scraped into a combined pool, grown out for 8 hours in 500 mL terrific broth (TB) containing 100 μg/mL carbenicillin, and midi-prepped using the ZymoPure II Plasmid Midiprep Kit (Zymo Research) following manufacturer’s procedures.

**LTCv2 Cloning Forward primer:** AATTAACCCTCACTAAAGGGATAATTTAATTAA

**LTCv2 Cloning Reverse primer:** GTCGTAATGACTAAGATGACTGCCATTAATTAA

Sequence validation and analysis of LTC\_v2 library

Lentiviral plasmid libraries were sequence confirmed as described for the 1,024 epegRNA libraries with modest differences. As before, plasmid libraries were amplified with two successive rounds of PCR as done to sequence genomic loci. PCR1 was performed using Phusion Green Hot Start II High-Fidelity PCR 2x Master Mix with the following cycle conditions: 98ºC for 2 minutes, [98ºC for 10 seconds, 65ºC for 20 seconds, 72ºC for 1 minute] where the cycle number was determined by qPCR, followed by 72ºC for 2 minutes, and a hold at 12ºC. PCR2, gel extraction, library preparation, and sequencing were carried out as before.

**intBC Sequencing Validation Forward primer:**

TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-[1N-4N]-GTAATTAACCCTCACTAAAGG

**intBC Sequencing Validation Reverse primer:**

TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-[1N-4N]-CATCGATACCTAATACGACTCACTATAGGGAGAG

To assess the balance of library elements, we used the FastX toolkit (v0.0.14) to mask low quality bases (<q20) and collapse reads. We removed common flanking sequences and summed reads matching the whitelist of intBC barcodes for sequences with Hamming distance <4 to account for sequencing errors and the length of intBC sequences.