CROPseq-multi: a versatile solution for multiplexed perturbation and decoding in pooled CRISPR screens

Russell T. Walton[1,2,*], Yue Qin[1,3], Paul C. Blainey[1,2,4,*]

1. Broad Institute of MIT and Harvard, Cambridge, MA, USA
2. Department of Biological Engineering, MIT, Cambridge, MA, USA
3. Eric and Wendy Schmidt Center, Broad Institute of MIT and Harvard, Cambridge, MA, USA.
4. Koch Institute for Integrative Cancer Research, MIT, Cambridge, MA, USA

* Correspondence: rwalton@broadinstitute.org and pblainey@broadinstitute.org

Abstract

Forward genetic screens seek to dissect complex biological systems by systematically perturbing genetic elements and observing the resulting phenotypes. While standard screening methodologies introduce individual perturbations, multiplexing perturbations improves the performance of single-target screens and enables combinatorial screens for the study of genetic interactions. Current tools for multiplexing perturbations are incompatible with pooled screening methodologies that require mRNA-embedded barcodes, including some microscopy and single cell sequencing approaches. Here, we report the development of **CROPseq-multi**, a CROPseq[1]-inspired lentiviral system to multiplex *Streptococcus pyogenes* (Sp) Cas9-based perturbations with mRNA-embedded barcodes. CROPseq-multi has equivalent per-guide activity to CROPseq and low lentiviral recombination frequencies. CROPseq-multi is compatible with enrichment screening methodologies and optical pooled screens, and is extensible to screens with single-cell sequencing readouts. For optical pooled screens, an optimized and multiplexed *in situ* detection protocol improves barcode detection efficiency 10-fold, enables detection of recombination events, and increases decoding efficiency 3-fold relative to CROPseq. CROPseq-multi is a widely applicable multiplexing solution for diverse SpCas9-based genetic screening approaches.

Introduction

Most pooled screens achieve perturbation of genetic components through lentiviral delivery of one or more Cas9 single guide RNAs (sgRNAs). Lentiviral delivery systems are near-ubiquitous across *in vitro* pooled screens conducted today due to their ability to efficiently deliver perturbations to a wide range of cell types and retain a record of the perturbation identity in the form of the lentiviral cargo integrated in the host-cell genome. Most commonly, a single perturbation is delivered to each cell, however multiplexing perturbations offers several advantages. Despite improvements in algorithms for guide design[2–7], the selection of active and specific guides remains challenging and pairing guides can increase on-target performance. The use of multiple guides targeting the same genetic element (“single-target screens”) has been shown to improve on-target activity in CRISPR knockout (CRISPR-KO), activation (CRISPRa), and interference (CRISPRi) screens[8–11]. Multiplexing perturbations also enables targeting multiple distinct genetic elements (“combinatorial screens”) to identify genetic interactions – phenotypes that arise from an interaction of genetic components.

Multiplexing perturbations in screens typically requires multi-perturbation lentiviral vector designs (“multiplexing vectors”). In principle, single-perturbation vectors can enable random sampling of perturbation combinations in pooled screens, via either high multiplicity of infection (MOI) (with limited

control of multiplicity) or serial transduction and selection[12,13]. However, random sampling of perturbation combinations requires profiling an immense combinatorial space, which scales as $\binom{m}{n}$, for *m* targets and combinations of *n* perturbations per cell. For example, pairwise interactions of only 200 gene targets results in 19,900 unique combinations, roughly equivalent in scale to single-target genome-wide screens in common practice today. This limitation precludes the use of single-perturbation vectors for multiplexed single-target screens and poses a major limitation to their utility for combinatorial screens.

Instead, multiplexing vectors are delivered with a single selection step at a low MOI. Multiplexing vectors uniquely enable multiplexing in single-target screens and offer an alternative to random sampling in combinatorial screens. In addition to enabling an exhaustive search of the combinatorial space, multiplexing vectors allow the perturbation of a biologically informed subset of target combinations[14–18]. In such screens, the size of multiplexed vector libraries scales linearly with the number of selected target combinations.

Existing lentiviral perturbation systems offer performance and functionality tradeoffs with respect to multiplexing capacity, lentiviral recombination frequencies, and mRNA-barcoding compatibility (**Supplementary Table 1**). Derived from standard single-perturbation designs (**Supplementary Figure 1A**), vectors encoding serial pol III (typically U6) promoters and SpCas9 sgRNAs ("guides") are a common approach to multiplex perturbations[19,9,20] (**Supplementary Figure 1B**). A major limitation of these multiplexing systems is the roughly 400 base pair (bp) distance separating sgRNA spacer sequences, leading to lentiviral recombination and unintended sgRNA combinations in about 30% of cells[19,9,20] (**Table 1**). These recombination events are an inherent property of lentiviral systems as lentivirus are pseudodiploid and, in the process of infection, reverse transcription during minus-strand synthesis exhibits frequent template switching in a homology and distance-dependent manner[21–23] (**Supplementary Figure 1C, D**). If perturbation identities are assigned based on the observation of only one sgRNA, these recombination events will be misassigned and contribute to experimental noise[21]. If the screening methodology captures both sgRNA identities, recombination events can be detected and filtered out. However, high recombination rates remain a major burden when perturbation coverage is challenging or costly to maintain, such as in single-cell RNA-sequencing (scRNA-seq)-based screens[20].

Other designs seek to minimize recombination by reducing the distance separating sgRNAs. The Big Papi vector[24] and similar designs[25] employ antiparallel orthogonal pol III promoters and sgRNAs, reducing the distance separating spacers to under 200 bp and decreasing recombination to about 9%[26]. Cas12a systems capitalize on the native crRNA array processing ability of Cas12a enzymes[27,28], enabling separation of spacers by only 20 bp and likely reducing recombination to negligible levels[10,29,25,8]. However, in contrast to the widely used SpCas9, Cas12a enzymes are limited in guide design by relatively restrictive protospacer adjacent motifs, are less effective in CRISPR-KO screens unless compensated by multiple Cas12a guides per gene[10,25,29], lag in development with applications including CRISPR-KO, CRISPRa, CRISPRi, base editing, and prime editing[30–32], and are limited in the availability of existing cell lines and animal models. Regardless, all current multiplexing solutions lack compatibility with mRNA-barcoding, which is required for some screening modalities, including popular single-cell RNA-sequencing and *in situ* detection workflows[1,12,19,33].

LentiGuide-Barcode (LentiGuideBC)[12] vectors and similar designs (e.g. Perturb-seq[33], MOSAIC-seq[34], CRISP-seq[35]) typically sacrifice multiplexing capability in pooled screens for the ability to express a linked barcode in mRNA (**Supplementary Figure 1E**). In principle, these vector designs are not fundamentally incompatible with multiplexing; however, constructing libraries with at least three distal programmed sequence elements (i.e. two sgRNAs and a barcode) is challenging. To date, these designs have required multi-step arrayed cloning[19], making the approach impractical for most pooled screens. In these designs, the linked barcode is separated from the sgRNA spacer by at least 1,700 bp comprising the pol II promoter

and selection gene, resulting in lentiviral recombination near the theoretical maximum of 50%[21]. As only about half of cells are correctly genotyped, this recombination contributes to a major loss of statistical power[21]. Efforts to decrease this distance by moving the U6-sgRNA downstream of the pol II promoter and resistance gene have resulted in poor guide activity[21], potentially due to transcriptional interference[1,21]. Our group described co-packaging integration-deficient templates to mitigate recombination, albeit at the cost of a 100-fold reduction in lentiviral titer[22].

The CROPseq approach offers a solution to mRNA-barcoding that is not impacted by lentiviral recombination[1] (**Supplementary Figure 2A, B**). By embedding the U6 promoter and sgRNA within the lentiviral 3' long terminal repeat (LTR), the CROPseq design leverages the high-fidelity intramolecular duplication of the 3' LTR to the 5' end of the lentiviral genome during genome integration (**Supplementary Figure 2B**). This duplication results in two copies of the sgRNA, with the 5' copy expressing functional sgRNAs without transcriptional interference and the 3' copy transcribed as mRNA in the 3' untranslated region (UTR) of the pol II-transcribed selection gene, compatible with mRNA-detection approaches. With these considerations in mind, we sought to engineer a lentiviral system to enable multiplexed SpCas9-based perturbations with mRNA-barcoding.

Results

Design of a CROPseq-inspired multiplexing vector

We reasoned that the lentiviral 3' LTR could be compatible with a minimal multiplexing system. While CROPseq vectors enable faithful duplication of the 3' LTR via intramolecular recombination during plus-strand synthesis, adjacent guides and/or barcode elements are still vulnerable to intermolecular recombination during minus-strand synthesis (**Supplementary Figure 2C**). To minimize the 3' LTR insertion and sgRNA separation, we opted to use endogenous transfer RNA (tRNA) processing as our multiplexing solution, as others have implemented[36–42] (**Figure 1A**). Encoded in about 72 bp, tRNA molecules recruit endogenous RNAses P and Z to cleave the tRNA at the 5' and 3' ends, such that, when positioned between sgRNAs within a single transcript, the RNA is processed into separate, functional sgRNAs. While only a single tRNA is required for multiplexing (i.e. U6-sgRNA-tRNA-sgRNA), we additionally preceded the first sgRNA with a tRNA (i.e. U6-tRNA-sgRNA-tRNA-sgRNA) as tRNAs encode their own promoter elements and, in some contexts, may improve expression alone or in conjunction with a U6 promoter[36,41,42]. Additionally, each tRNA eliminates the requirement for a 5' guanine base on the following guide that is otherwise required in U6 transcription systems and is often encoded as a mismatched 20th or 21st base of the spacer. This design increased the size of the 3' LTR insertion from 352 bp in CROPseq to 643 bp in CROPseq-multi (**Supplementary Figure 3A**). tRNA-encoding sequences can be processed by the endogenous RNases out of mRNA[42]. Processing of tRNAs out of the lentiviral genomic RNA and the selection gene transcript would interfere with lentiviral production, selection, and detection of barcoded mRNA. To enable selective processing of tRNAs from pol III (but not pol II) transcription products, we reversed the orientation of the elements within the 3' LTR, such that the U6 promoter, tRNAs, and sgRNAs are encoded on the lentiviral minus strand. We termed our CROPseq-inspired multiplexing solution **CROPseq-multi** (**Figure 1A**).

In addition to our guide multiplexing changes, we added 12 bp barcodes internal to the sgRNAs (iBARs) as freely specified additional readout elements linked to the sgRNA[43] (**Figure 1A, B, Supplementary Figure 3A**). Linked barcodes are advantageous to minimize the sequence length needed to uniquely identify library members, to represent combinations of guides that may not be individually unique, and to encode additional information such as clonal identity. As linked barcodes are prone to distance-dependent recombination, the iBAR system is attractive for the placement of barcodes only 19 bp from the spacer, within the synthetic

loop that joins the crRNA and tracrRNA into a sgRNA (**Figure 1B, Supplementary Figure 3A**). As the iBARs are transcribed both antisense as mRNA and within the sgRNA scaffold, their detection should be compatible with both mRNA-based[12,19,33,35] and direct-capture[9] (U6 product) protocols.

We first addressed the compatibility of CROPseq-multi with lentiviral production and transduction. It has been suggested that the lentiviral 3' LTR may be incompatible with the larger insertions needed to encode multiple perturbations[21,23]. Insertions of up to 1,200 bp within the 3' LTR have been evaluated, albeit with reduced viral titers[44]. While our design was compatible with lentiviral delivery, we observed a roughly 10-fold reduction in functional titer of CROPseq-multi compared to CROPseq (**Figure 1C**). Reasoning that increased size of the 3' LTR insertion might explain the low titer, we compared viral titers of CROPseq-multi

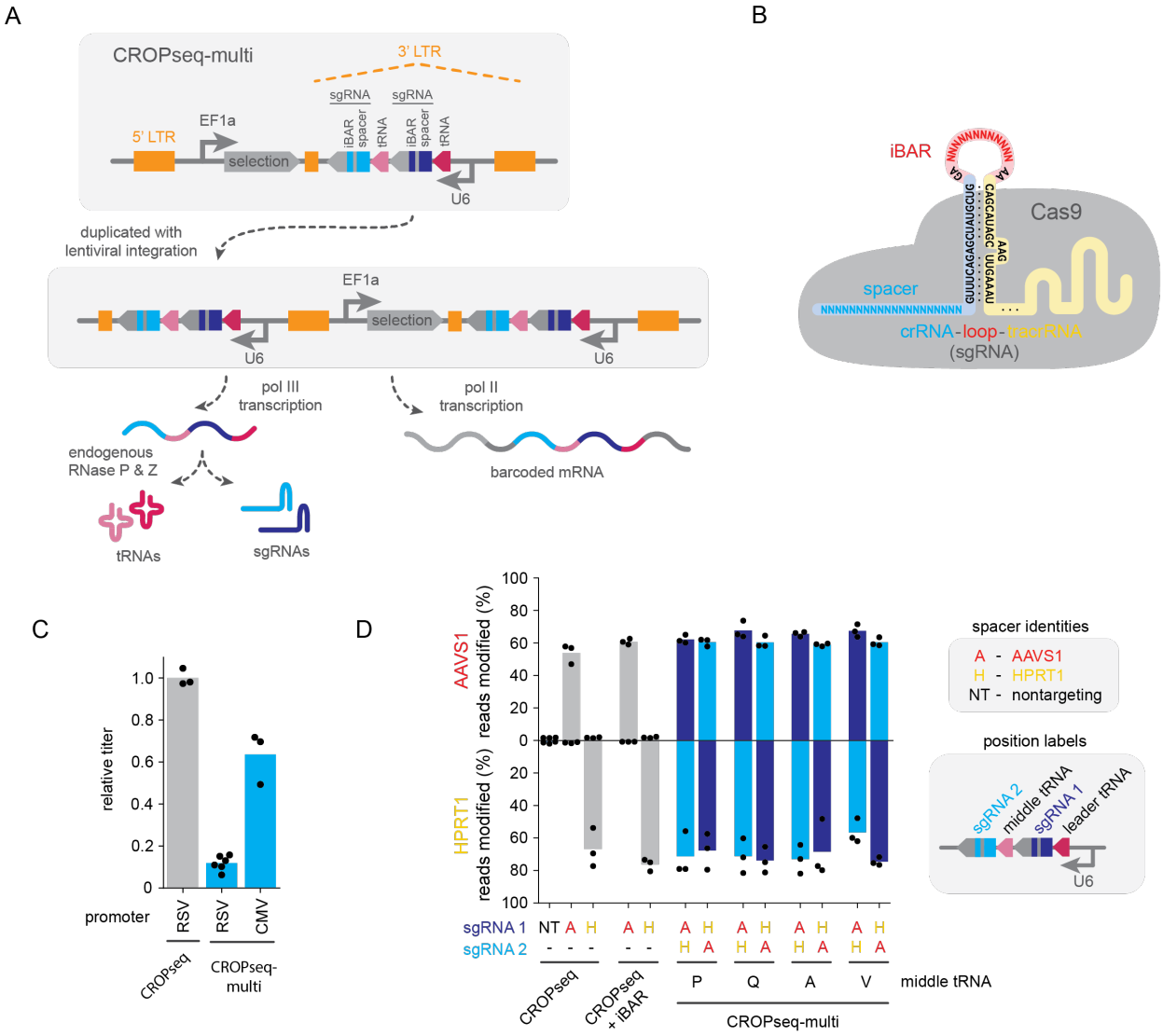


Figure 1. CROPseq-multi: a CROPseq-inspired multiplexing system.

(A) CROPseq-multi encodes two sgRNAs with internal barcodes (iBARs), multiplexed using tRNAs, within the lentiviral 3' LTR. The 3' LTR is duplicated during lentiviral integration, producing a second copy of the sgRNAs. (B) iBARs place linked barcodes within the loop joining the crRNA and tracrRNA into the sgRNA. (C) Lentiviral titers of CROPseq-multi relative to CROPseq. (D) Genome editing activities of CROPseq and CROPseq-multi vectors in SpCas9-expressing A549 lung adenocarcinoma cells, quantified by next-generation sequencing. Mean and n=3 biological replicates shown. LTR, long terminal repeats; RSV, Rous sarcoma virus; CMV, human cytomegalovirus.

with one guide (445 bp insertion) or two guides (643 bp insertion) against CROPseq (352 bp insertion). Surprisingly, we observed the same 10-fold reduction in titer with the single-guide CROPseq-multi, suggesting that titer did not decrease linearly with increasing 3' LTR insertion size and that the sequence elements (i.e. tRNA(s)) and/or orientation may contribute (**Supplementary Figure 3B**). We hypothesized that an alternate lentiviral promoter could improve titer if transcriptional interference with the U6 promoter and tRNAs was hindering lentiviral genome production. While most CROPseq vectors employ a Rous sarcoma virus (RSV) promoter to drive lentiviral genome expression during viral production, the human cytomegalovirus (CMV) promoter can improve titer in some lentiviral vectors[45]. Swapping the RSV promoter with a CMV promoter rescued the titer of the full two-guide CROPseq-multi system to 64% that of CROPseq (**Figure 1C**). We proceeded to use the CMV promoter as the default construction for CROPseq-multi.

Editing performance of CROPseq-multi

To assay the performance of CROPseq-multi for genetic perturbation, we transduced SpCas9-expressing A549 lung adenocarcinoma cells with CROPseq and CROPseq-multi vectors encoding guides targeting the AAVS1 and HPRT1 loci and evaluated genome editing efficiency by next-generation sequencing (NGS) (**Figure 1D**). Ideally, multiplexing systems possess equivalent per-guide activity to single-plex systems and display minimal positional bias (i.e. equal guide activity in all positions). We first validated that 12 bp iBARs were not detrimental to activity in the CROPseq vector architecture (**Figure 1D**). Next we tested both orientations of spacers and four guide-intervening or "middle" tRNAs in CROPseq-multi. None of the tested CROPseq-multi constructs had significantly different activity for either target compared to CROPseq or one-another (T-test with Bonferroni correction) (**Figure 1D**). We selected designs employing human proline, glutamine, and alanine tRNAs ($tRNA_P$, $tRNA_Q$, and $tRNA_A$, respectively; see **Supplementary Table 2**) for further development because they displayed the most consistent activity of the four tRNAs tested.

Lentiviral recombination with CROPseq-multi

To evaluate recombination rates, we performed arrayed and pooled lentiviral production with pairs of CROPseq-multi vectors and measured the identities of barcode elements after transduction and integration in genomic DNA. Arrayed lentiviral preparation guarantees that co-packaged genomes are identical, meaning that template switching does not disrupt perturbation pairing, and serves as a control for other sources of recombination, such as polymerase chain reaction (PCR) amplification. In pooled lentiviral preparation, as is performed in pooled screens, lentiviral genomes are essentially randomly paired within virions. In a pooled preparation of two vectors, 50% of virions will harbor different perturbation pairs that can result in observable intermolecular recombination events (**Figure 2A**). In a high-complexity library, nearly all virions are expected to package different perturbation pairs, so the recombination rate should approach 2x the observed rate in a two-vector assay. In all conditions, recombination was measured by next-generation sequencing of genomic integrations.

Intermolecular recombination events that decoupled spacers from their iBARs were detected but at rates below 1% that are acceptable for many applications (**Figure 2B**). For two CROPseq-multi vectors with a common middle tRNA, intermolecular recombination events resulting in incorrect pairings of sgRNAs ("pair-swaps") were observed at about 5.9% (**Figure 2B**). As the middle tRNAs we validated are divergent in sequence (**Supplementary Figure 3C**), we reasoned that if co-packaged CROPseq-multi constructs encoded orthogonal middle tRNAs, the homologous barcode-separating distance would be reduced to only 75 bp (**Supplementary Figure 3A**). In a pooled screening setting, this could be implemented by designing libraries with multiple orthogonal middle tRNAs to reduce the probability of co-packaging constructs with a common middle tRNA. Correspondingly, the use of orthogonal tRNAs reduced observed pair-swap recombination events to 3.2% (**Figure 2B**). We converted observed rates to expected underlying rates and

compared with previously reported recombination frequencies from other dual barcode systems (either multiplexed guides or guides with linked barcodes) (**Figure 2C, Supplementary Table 1**). Modeling barcode swapping frequency as a function of the length of homologous intervening sequence, we found that lentiviral recombination frequency could be modeled by fixed per-bp probability of template switching of $5.3x10^{-4}$ per bp, within the range of previously reported measurements of $4.9\text{-}13.5x10^{-4}$ per bp[46] (**Figure 2C**).

With both arrayed and pooled lentiviral preparations, we observed sgRNA deletion events in about 2% of integrations that appeared to be driven by sgRNA scaffold homology (**Figure 2B**), despite the use of two

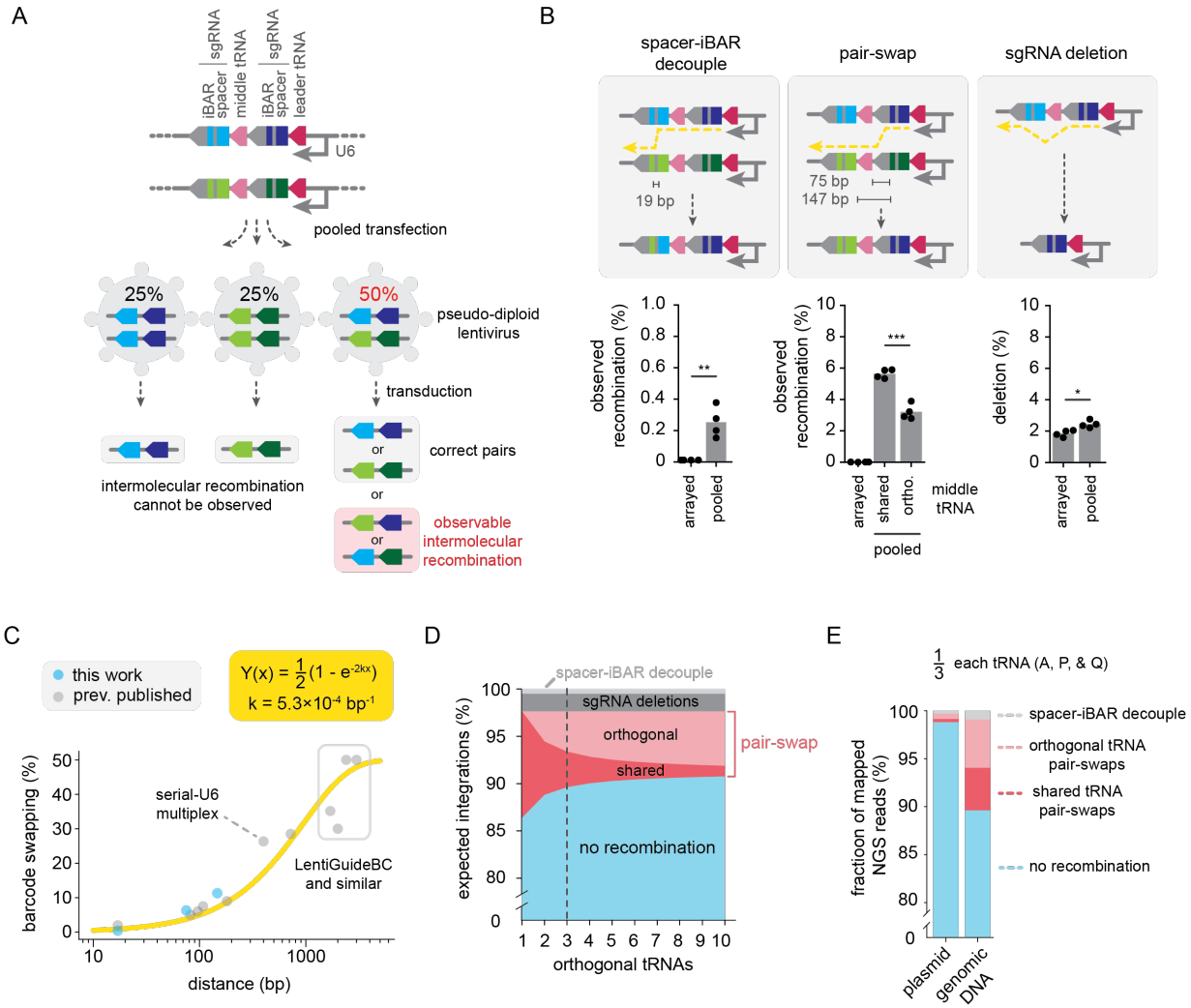


Figure 2. Lentiviral recombination with CROPseq-multi.

(A) Assaying lentiviral recombination with pooled lentiviral production. (B) Observed lentiviral recombination of CROPseq-multi barcode elements. (T-test: *p<0.05, **p<0.01, ***p<0.001) Illustrations depict one example of each type of recombination. (C) Lentiviral barcode swapping rates as a function of the distance separating barcode elements (e.g. spacer-spacer, spacer-linked barcode, etc.) in published barcoding systems and CROPseq-multi (see **Supplementary Table 1**). (D) Expected recombination frequencies in libraries with orthogonal tRNAs. Expected recombination rates for a three-tRNA system indicated with a dotted line. (E) Recombination in a CROPseq-multi library with equal proportions of three orthogonal tRNAs in plasmid DNA (prior to lentiviral recombination) and genomic DNA (after lentiviral recombination).

orthogonal scaffold sequences (**Supplementary Figure 3D**). These events exclusively result in the deletion of the second spacer. The resulting guide scaffolds are chimeras of the two orthogonal scaffolds and may vary in functionality. Examination of recombination outcomes revealed that the modest but significant increase in deletion frequencies in the pooled condition was driven by spacer homology – an artifact the use of identical spacers in opposite positions in two of the three constructs used in our assay (i.e. $sgRNA_{AAVS1}$-tRNA-$sgRNA_{HPRT1}$ and $sgRNA_{HPRT1}$-tRNA-$sgRNA_{AAVS1}$) – and should not be a factor in more common high-complexity libraries. Considering the observed deletion rates and the expected underlying spacer-iBAR decoupling and pair-swap rates based on the observed rates, we estimated the fraction of CROPseq-multi integration events without recombination for complex libraries to be 86% with the use of a single middle tRNA or 90% when implemented with three orthogonal middle tRNAs (**Figure 2D**).

Recombination and uniformity of pooled CROPseq-multi libraries

We next sought to validate the construction and performance of CROPseq-multi libraries in a pooled setting. We evaluated an assembly protocol (**Supplementary Figure 4A**) for sequence fidelity, recombination, and uniformity for a 1,080-member library encoding three different middle tRNAs. Mapping rates of individual barcode elements (spacer 1, iBAR 1, middle tRNA, spacer 2, and iBAR 2) ranged from 94-98% and all elements together mapped jointly to the library as designed in 84% of reads, with NGS error rates likely contributing to a substantial fraction of unmapped reads (**Supplementary Figure 4B**). We initially found that libraries were prone to PCR-mediated recombination during oligo pool amplification. Optimization of the amplification conditions revealed that PCR-mediated recombination was almost entirely determined by oligo pool template concentrations and the use of low template concentrations (12-24 pg/µL) could effectively eliminate recombination (**Supplementary Figure 4C**). With optimized amplification conditions, we observed less than 1% recombination in our plasmid libraries (**Figure 2E, Supplementary Figure 4C, D**). This oligo pool amplification strategy did not compromise library uniformity when scaled to the appropriate number of amplification reactions (see **Methods**) and, when paired with an optimized assembly procedure, resulted in 90:10 ratios (the ratio of the 90th and 10th percentiles of abundance) below 2 and Gini coefficients below 0.14 (**Supplementary Figure 4E, F**).

We additionally optimized an alternate oligo pool amplification strategy to repurpose the second iBAR sequence as a clonal barcode (**Supplementary Figure 4G**). This approach requires no changes to oligo pool design, only alternate amplification primers, to generate libraries with high complexity clonal barcodes (**Supplementary Figure 4G, H**).

Following pooled lentiviral preparation and transduction, we assayed recombination in genomic DNA and observed 90.5% of genomic integrations without recombination (**Figure 2E**). This analysis excludes deletion events that were not captured by a modified NGS library preparation procedure but otherwise recapitulates the expected contributions to recombination from spacer-iBAR decoupling and pair-swap events based on our arrayed measurements (**Figure 2D, E**).

Performance of CROPseq-multi in a pooled CRISPRi essentiality screen

To evaluate performance in a pooled screen, we used a CROPseq-multi CRISPRi library consisting of 1,080 constructs to target genes across a range of essentiality scores in DepMap[47,48] as well as non-targeting controls. We designed three constructs for every pair of gene-targeting guides: one “dual-targeting” construct encoding two gene-targeting guides (1-2) and two “single-targeting” constructs, each with one of the two targeting guides replaced by a non-targeting (NT) control (both 1-NT and NT-2) (**Figure 3A**). Each set of three constructs shares a common middle tRNA, which was assigned at random from the three we validated such that the three tRNAs are approximately equally represented in the library.

We transduced retinal pigment epithelial (RPE1) cells expressing dCas9 with *ZIM3*-derived KRAB domain[20] with the CROPseq-multi library and assessed depletion at 7 and 14 days (**Supplementary Figure 5A**). For essential genes, defined here by RPE1-specific DepMap gene effect scores below -0.5, dual-targeting constructs had an average log2 fold change (LFC) 0.45 and 1.2 greater in magnitude compared to single-targeting constructs at days 7 and 14, respectively (**Figure 3A, Supplementary Figure 5B, C**). For classification of essentiality, dual-targeting constructs outperformed single-targeting constructs, assessed by area under receiver operator characteristic curve (AUROC) scores (**Figure 3B, C**). At day 7, dual-targeting constructs had AUROC scores of 0.88 compared to 0.76-0.79 for single-targeting constructs (**Figure 3B**). At day 14, dual-targeting construct AUROC scores improved to 0.94 compared with 0.85-0.88 for single-targeting constructs (**Figure 3C**). Single-targeting constructs with second-position guides outperformed single-targeting constructs with first-position guides by AUROC scores of 0.88 versus 0.85 at day 7 (**Figure 3B**) and 0.79 versus 0.76 at day 14 (**Figure 3C**). This trend was consistent across middle tRNA identities (**Supplementary Figure 5D**). The lack of matched controls for direct comparison of middle tRNA performance (i.e. the same guide pairs with each middle tRNA) and position bias (i.e. the same guide pair in both orientations) is a limitation of this analysis. Future screens should address position bias and tRNA performance more exhaustively.

We also compared the performance of individual dual-targeting constructs to composite metrics (mean and minimum) of the single-targeting constructs encoding the same guides (**Figure 3A**). At day 7, the mean and minimum of single-targeting pairs improved the AUROC scores to 0.84, below the 0.89 AUROC score for dual-targeting constructs (**Figure 3B**). At day 14, composite metrics of single-targeting pairs again improved to AUROC scores of 0.93, compared to the dual-targeting construct performance of 0.94 (**Figure**

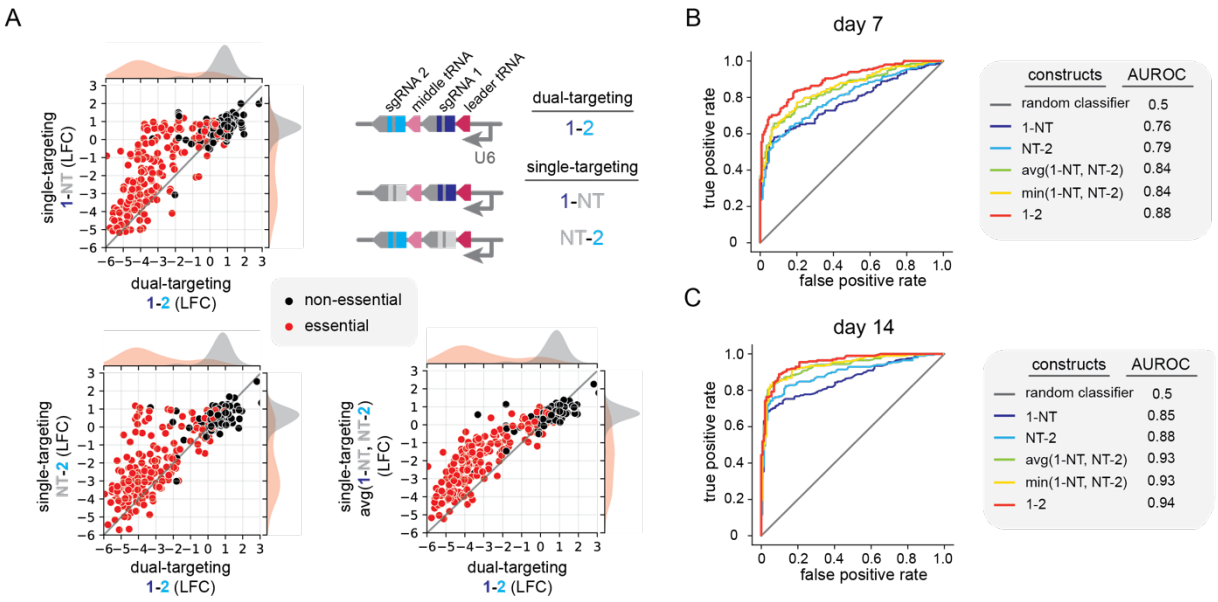


Figure 3. Performance of CROPseq-multi in a pooled CRISPRi essentiality screen. (A) Log2 fold changes of dual-targeting (1-2) CROPseq-multi constructs versus the corresponding single-targeting (1-NT and NT-2) constructs and construct averages. Log2 fold changes are the mean of two biological replicates. (B,C) Receiver operator characteristic curves for classification of gene essentiality at days 7 (B) and 14 (C). LFC, log2 fold change; AUROC, area under receiver operator characteristic curve.

3C). Consistent with previous studies[8,9,11], these results suggest that multiplexed perturbation systems outperform single-perturbation systems for single-target screens and can do so with smaller library sizes.

Multiplexed detection of CROPseq-multi with *in situ* sequencing for optical pooled screens

Optical pooled screens enable high-content, single-cell-resolved phenotypic measurements, spanning spatial scales and including temporal dynamics and diverse molecular measurements, at the scale of tens of millions of cells[12,49,50]. However, the unique constraints of *in situ* barcode detection have made it challenging to multiplex perturbations with optical pooled screens. Briefly, *in situ* detection of barcodes for optical pooled screens involves fixation and permeabilization of cells, reverse transcription of barcoded mRNA to cDNA, fixation of the cDNA, copying the barcode sequence into a padlock probe ("gapfill"), ligation of the padlock into a circular ssDNA template, rolling circle amplification, and sequencing by synthesis to decode perturbations[12,51] (**Figure 4A**). Robust mRNA expression is required for efficient detection and multiple perturbations must either be encoded by a single barcode or detected separately, as multiple barcodes. As sequencing reagent costs and imaging time impact screening throughput, minimizing the requisite number of sequencing cycles (i.e. barcode bases) to uniquely identify perturbations is desirable.

We designed padlock probes flanking each spacer and iBAR such that both are captured within the gapfill (**Figure 4A, Supplementary Figure 6A**). In contrast to the CROPseq vector, the orientation of *in situ* sequencing for CROPseq-multi is opposite the orientation of the sgRNA (**Supplementary Figure 6B**). We first optimized the detection efficiency of the first spacer and iBAR pair in A549 cells transduced at an MOI of 0.1 with a library of three CROPseq-multi vectors encoding orthogonal middle tRNAs. With our standard protocol, detection efficiencies (reads per cell) were low for the CROPseq-multi iBAR, averaging 1.3 reads per cell, compared to an average of 2.9 reads per cell for the CROPseq spacer (**Figure 4B, C**). We implemented two protocol changes to improve detection efficiencies for CROPseq-multi (**Figure 4A**). First, we aimed to improve mRNA retention by altering the primary fixation, adding 0.007% glutaraldehyde to the standard 4% paraformaldehyde (PFA) fixative. Second, we sought to optimize cDNA retention by using a biotinylated reverse transcription primer and adding a streptavidin incubation between the reverse transcription and cDNA fixation steps to improve cDNA anchorage within the fixed cells (**Figure 4A**). The optimized cDNA retention alone improved detection efficiency to an average 7.5 reads per cell for CROPseq-multi (**Figure 4B, C**). The optimized primary fixation alone did not substantially improve detection, but in combination with the optimized cDNA retention, detection efficiency improved further to an average 18.8 reads per cell (**Figure 4B, C**). These modifications had relatively modest effects on detection of the CROPseq vector (**Figure 4B**). In RPE1 cells, these protocol changes similarly improved detection efficiencies (**Supplementary Figure 6C, D**) and fine tuning of the primary fixation conditions suggested an optimal concentration of glutaraldehyde in the range of 0.007-0.003% in 4% PFA (**Supplementary Figure 6E**).

An additional feature of the CROPseq-multi design is the use of two barcodes to facilitate the detection of recombination events and to improve perturbation decoding efficiency. First, detection of both barcodes would enable identification and filtering of lentiviral recombination events. Second, with simultaneous detection of both barcodes as separate reads ("multiplexed detection"), a total of two nucleotides of a barcode pair can be decoded per sequencing cycle – one nucleotide from each barcode per cycle. Of note, this strategy is dependent on the ability to reliably detect both barcodes in each cell, which should be facilitated by the optimized detection protocol.

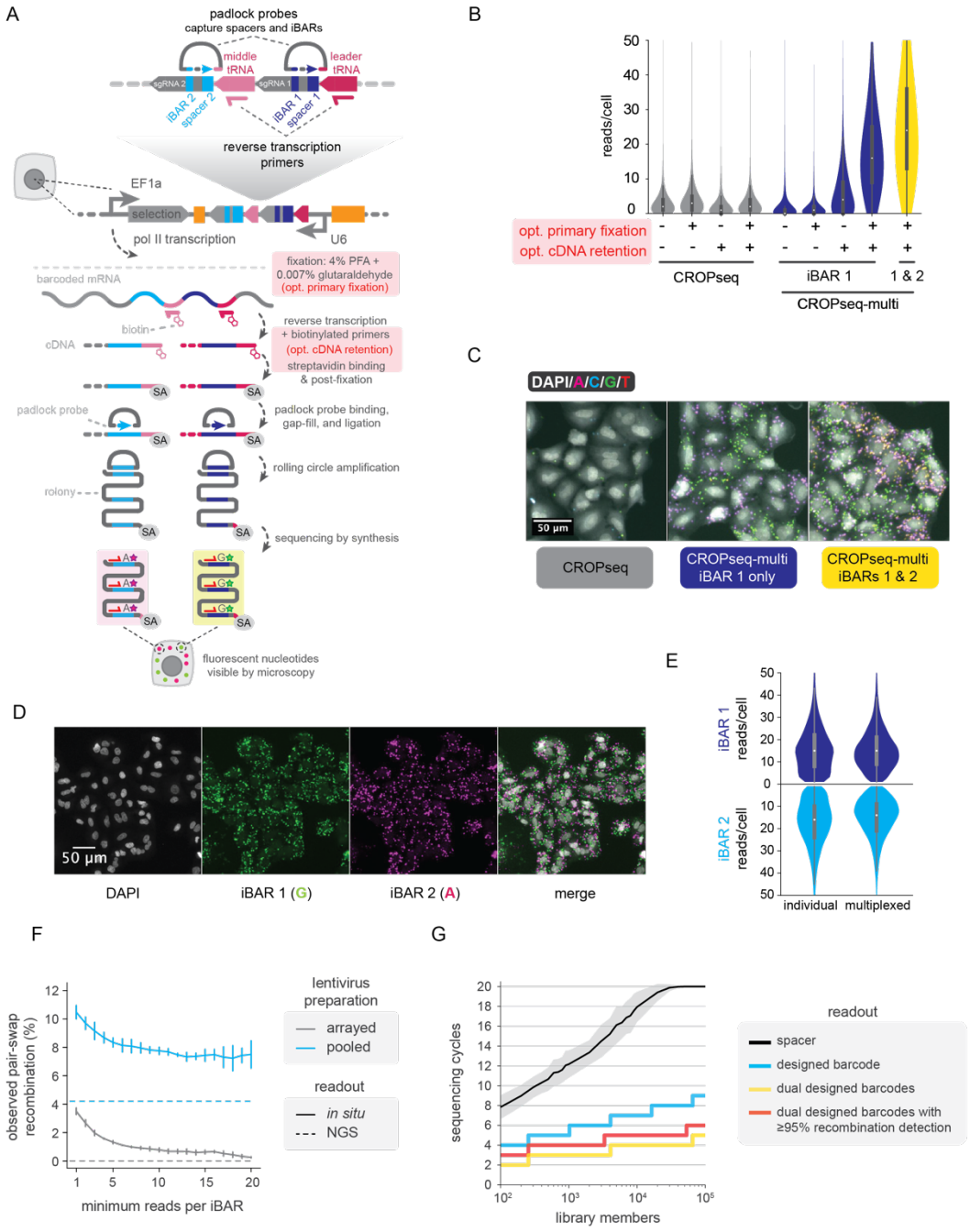


Figure 4. Multiplexed *in situ* detection of CROPseq-multi for optical pooled screens. (A) Overview of the *in situ* sequencing workflow. (B) Optimizing *in situ* detection of CROPseq-multi. (C) Representative *in situ* sequencing images in A549 cells of CROPseq detection with the standard detection protocol and CROPseq-multi with an optimized detection protocol. (D) Representative images of *in situ* sequencing reads with multiplexed detection in A549 cells. (E) Detection efficiency of CROPseq-multi iBARs individually and with multiplexed detection. (F) Quantifying recombination in a three-vector pool with *in situ* sequencing and NGS. (G) Modeling the sequencing cycles necessary to uniquely identify library members with different decoding methods. For decoding via the spacer, libraries were simulated by randomly sampling guides from a genome-wide CRISPRi library (see **Methods**); mean and standard deviation are shown. PFA, paraformaldehyde; SA, streptavidin; NGS, next-generation sequencing.

We evaluated multiplexed detection of iBARs 1 and 2 with our optimized protocol, using additional reverse transcription primers and padlock probes to detect the second spacer and iBAR pair. As these oligonucleotide reagents hybridize to sgRNA-adjacent tRNA, each orthogonal tRNA requires a different set of *in situ* detection oligos (**Supplementary Figure 6A**). For multiplexed detection with three orthogonal tRNAs, a total of four sets of reverse transcription primers and padlock probes are used, one set for the first spacer and iBAR and three sets (one for each middle tRNA) for the second spacer and iBAR. With multiplexed detection, we observed a mean of 27.7 total reads per cell in A549 cells (**Figure 4B**). Detection efficiencies were similar across constructs encoding different middle tRNAs (**Supplementary Figure 6F**). Multiplexed detection could impact per-barcode detection efficiencies due to optical crowding at high read densities. We evaluated detection of iBAR 1 and iBAR 2 both individually and multiplexed and observed only modestly lower per-barcode detection efficiencies when multiplexed (**Figure 4C, D**). The absence of substantial spatial overlap of *in situ* sequencing reads may suggest that detection events for each spacer/iBAR largely originate from different mRNA molecules. We found that reads could be assigned to iBAR 1 or iBAR 2 based on either unique sequence mapping or the use of a fluorophore-conjugated oligo to label one of the iBARs (**Supplementary Figure 6G, H**).

We used multiplexed *in situ* detection to decode perturbations and quantify recombination in cells transduced with three CROPseq-multi constructs at an MOI of 0.1, prepared in either arrayed or pooled lentiviral settings. Assaying recombination with three unique vectors, the observed recombination rate is expected to reflect ⅔ of the underlying recombination rate, analogous to a 2-vector recombination assay (**Figure 2A**). NGS of samples transduced with arrayed and pooled lentiviral preparations revealed observed pair-swap rates of <0.01% and 4.2%, respectively (**Figure 4E**). Unlike NGS measurements, quantification of recombination *in situ* is sensitive to cell segmentation accuracy, technical artifacts such as transcript diffusion, and lentiviral MOI. We varied the stringency of read assignment to cells by varying the required minimum read counts per iBAR between 1 and 30, reasoning that cells with few reads for either iBAR might be the result of deletion recombination events, silencing of the lentiviral transgene, or incomplete selection, and, together with imperfect cell segmentation, could appear as false-positive pair-swap events (**Figure 4F**). Relative to NGS, we observed modestly higher pair swap frequencies for the same samples *in situ*, ranging from 0.2-3.5% and 7.2-10.5% with arrayed and pooled lentiviral preparation, respectively, with the highest read count stringencies resulting in the lowest observed recombination frequencies (**Figure 4F**). In a screening context, filtering out all such low-confidence assignments and incorrect pairings will be the primary objective so the distinction between pair-swap events and these potential modes of false-positives will be less important.

In optical pooled screens, throughput is largely dictated by imaging time, which scales linearly with sequencing cycles during sample genotyping. Decoding via spacer sequences typically requires sequencing all 20 bases for genome-scale libraries (**Figure 4G**). Imposing orthogonality requirements at the guide-design stage is one compromise that we have employed for CRISPR-KO screens, however this necessitates the selection of guides with suboptimal predicted on- and off-target activities and is infeasible for applications with relatively strict guide design constraints, including CRISPRa and CRISPRi. The use of even a single linked barcode can reduce the required cycle number for a given library size to roughly half that of decoding with the spacer sequence of a typical sgRNA library (**Figure 4G**). With multiplexed detection of two linked barcodes, the required number of cycles is again halved relative to a single linked barcode, as two bases are decoded per sequencing cycle (**Figure 4G**). For most libraries, a single additional sequencing cycle with multiplexed decoding is sufficient to detect ≥95% of recombination events, corresponding to a roughly 3-fold decrease in cycle number relative to decoding with spacer sequences (**Figure 4G**). For example, fully decoding a genome-wide CROPseq library with 4 guides per gene (about 80,000 vectors) would require sequencing all 20 cycles of the spacer. With CROPseq-multi, an equivalent number of vectors, encoding twice as many guides per gene, could be decoded with only 6 cycles of

sequencing while detecting recombination events. Correspondingly, multiplexed decoding reduces the sequencing reagent costs by the same factor.

Discussion

CROPseq-multi is a generalized multiplexing solution for single-target and combinatorial Cas9-based CRISPR screens that addresses numerous technical challenges, including robust guide activity, low lentiviral recombination, and compatibility with mRNA-barcoding. For optical pooled screens, CROPseq-multi enables superior detection and improved decoding efficiency and should readily enable combinatorial screens.

In single-target CRISPR-KO, CRISPRa, and CRISPRi screens, multiplexing systems typically achieve superior on-target performance with smaller library sizes compared to single-perturbation systems[8,9,11]. Our CRISPRi enrichment screen targeting essential genes with CROPseq-multi supports this claim (**Figure 3 B, C**). Beyond screening modalities commonly employed today, we expect CROPseq-multi to be readily extensible to prime editing approaches[42,52], which rely on the delivery of two guide components, a prime editing gRNA (pegRNA) and nicking sgRNA, for efficient editing.

We evaluated tRNA performance and position bias for a small set of tRNAs, however further characterization may yield additional design considerations. The promoter activity of tRNAs may vary across biological models and it is feasible that sequence features such as those relating to RNA secondary structure may favor some tRNA-sgRNA pairings over others. In Cas9-expressing A549 cells, CROPseq-multi constructs with either orientation of two guides and any of the three tRNAs we used throughout this study had indistinguishable genome editing efficiencies (**Figure 1D**). In our pooled CRISPRi screen, second-position guides averaged better performance than first-position guides by AUROC scores of 0.88 to 0.85 at day 14, though this evaluation lacked a comparison of identical guides in both positions (**Figure 3B, C**). One potential explanation for this putative difference is the activity of the middle tRNA as an additional promoter for the second-position guide. In this scenario, reducing position bias may imply reducing activity of the second-position guide. For single-target screens, where the primary objective is maximal on-target activity, we don't expect reducing second-position activity to be advantageous. For screens where position bias is a major concern, such as combinatorial screens, designing libraries with guide pairs represented in both orders is one option to measure and correct potential biases.

It may be possible to extend CROPseq-multi to three or more sgRNA perturbations, though challenges exist relating to molecular cloning and lentiviral performance. Commercially available oligo library synthesis options are currently limited to 300 nucleotides in length, sufficient to encode at most two spacers separated by an sgRNA scaffold and tRNA. Pending improvements in oligo synthesis, serial assembly steps may enable pooled cloning of higher order perturbation combinations. Lentiviral recombination is also expected to pose a greater challenge with extended arrays of repetitive elements; the incorporation of additional orthogonal sgRNA scaffolds and tRNAs to combat this effect will require careful testing and validation. Lentiviral titer may also be negatively impacted by encoding additional perturbations within the 3' LTR, although we did not observe a difference in lentiviral titer between CROPseq-multi constructs containing one or two sgRNAs (**Supplementary Figure 3B**). For these reasons, we expect Cas12a systems will remain attractive solutions for multiplexing 3 or more perturbations. We expect CROPseq-multi-inspired designs (i.e. 3' LTR-embedded antisense crRNA arrays) will enable multiplexed Cas12a screens with mRNA-barcoding compatibility.

Further improvements in decoding efficiency may be possible through a multiplexed detection approach; multiplexed readout of three or more barcodes would further reduce the requisite number of sequencing

cycles to identify perturbations by decoding more bits of information per cycle. However, further increasing barcode readout multiplexity may be less impactful as additional reductions in sequencing reagent cost will be comparatively small and because phenotypic measurements at relatively high magnification may become the primary throughput bottleneck in the context of increasingly efficient library decoding.

Recently developed approaches for *in situ* genotyping with *in vitro* transcription provide an alternative strategy to mRNA-detection for optical pooled screens[53,54]. A similar strategy may be compatible with a T7-promoter-modified CROPseq-multi approach for compatibility with both mRNA-based and *in vitro* transcription-based detection. The generation of a single barcode focus per lentiviral integration, a characteristic feature of *in vitro* transcription approaches, may be incompatible with multiplexed decoding, which relies on spatial demultiplexing of multiple barcode foci that originate from the same lentiviral integration. While this may preclude decoding efficiency improvements with multiplexed detection, alternate barcoding strategies or serial decoding may still enable the identification of recombination events with these *in situ* detection approaches.

Beyond decoding efficiency and detection of recombination events, multiple barcodes can enable additional functionalities such as barcoding of sub- or clonal populations. Barcoding of subpopulations in pooled CRISPR screens has been leveraged to improve enrichment screen performance[43,55,56] and may be extensible to profiling screening techniques such as optical pooled screens and scRNA-seq-based screens. Additionally, clonal barcoding in screens may enable characterization of clonal heterogeneity and subclonal dynamics.

In conclusion, CROPseq-multi is a versatile multiplexing platform for diverse CRISPR screening methodologies. We addressed challenges in multiplexing perturbations while maintaining design compatibility across enrichment, single-cell sequencing, and optical pooled screens. This versatility provides the opportunity for direct comparison and integration of different screening modalities, for example scRNA-seq and imaging-based approaches. CROPseq-multi will enable single-target screens with smaller libraries and improved performance and we expect the compatibility of CROPseq-multi with high-content screening techniques to enable new directions for combinatorial screens. In particular, we anticipate combinatorial optical pooled screens will be a powerful approach to interrogate genetic interactions at high throughput and with rich, single-cell-resolved phenotypic measurements.

Methods

Arrayed molecular cloning

The CROPseq-multi-Puro entry vector was derived from CROPseq-puro-v2 (Plasmid #127458) by modifying the lentiviral promoter and 3' LTR with isothermal assembly. Alternate selection markers were subcloned from CROPseq-multi-Puro by isothermal assembly.

CROPseq-multi entry vectors with Puromycin (Addgene ID 216217), Zeocin (Addgene ID 216218), and EBFP2-NLS (Addgene ID 216219) selection markers are available through Addgene. To facilitate the generation of custom derivatives (e.g. alternative pol II promoters, selection genes, etc.), CROPseq-multi vectors will be supported in the Broad Institute Genetic Perturbation Platform's modular vector assembly system, Fragmid[45] (https://portals.broadinstitute.org/gppx/fragmid/public).

For arrayed cloning of CROPseq-multi dual-guide vectors, dual guide inserts were purchased as gene fragments from Twist Biosciences and cloned into the entry vector using golden gate assembly. Briefly, 10 nanograms (ng) of entry vector and 2 ng of gene fragment were assembled in a 2.5 µL reaction of 1X T4 DNA ligase buffer (New England Biolabs B0202S) and 1X BsmBI-v2 Golden Gate Enzyme Mix (New England Biolabs E1602S) with the thermal cycling protocol: 30 cycles of (42ºC for 1 min, 16ºC for 1 min), then 60ºC for 5 min. All arrayed cloning was performed in NEB Stable chemically competent cells (New England Biolabs C3040H) according to the manufacturer's recommendations and all cultures were grown at 30 ºC. Liquid cultures were grown shaking at 225 RPM and 30 ºC in 2xYT media with 100 µg/mL of carbenicillin.

Pooled molecular cloning of CROPseq-multi dual-guide libraries

For pooled cloning of CROPseq-multi dual-guide vectors, 300 nucleotide oligo libraries were purchased from Twist Biosciences and cloned into the entry vector by restriction digestion and ligation. Code for the design of CROPseq-multi oligo libraries is available on GitHub. Oligo pool amplification PCRs were assembled with 1X KAPA HiFi HotStart Ready Mix (Roche KK2601), 1X EvaGreen qPCR dye (Biotium 31000), 1M Betaine (MilliporeSigma B0300), 300 nM forward and reverse primers (Integrated DNA Technologies), and 12-24 pg/µL of template in 25 µL reactions. If multiple sublibaries with distinct amplification primer pairs are encoded in one oligo pool, the template concentration for the sublibrary is given by:

[sublibrary template] = [oligo pool] * (sublibrary percentage of oligo pool)

PCRs were conducted with the following thermal cycling protocol: 95 ºC for 3 min, 14-16 cycles of (98 ºC for 20 s, 62 ºC for 15 s, 72 ºC for 15 s), then 72 ºC for 1 min. We strongly caution against the use of the manufacturer's recommended amplification conditions, 10 ng of oligo pool template per 25 µL PCR (400 pg/µL), as of this publication date. Across three oligo pool orders, we observed that the frequency of recombination within plasmid libraries was determined by the oligo pool template concentration during amplification and template concentrations greater than 200 pg/µL resulted in double-digit recombination rates in plasmid libraries (**Supplementary Figure 4C**). For oligo pools that encode multiple sublibraries, the sublibrary concentration, not the total oligo pool concentration, determines recombination frequencies, so long as the amplification primers are specific to the sublibrary. Rather than high oligo pool template concentrations, we found scaling the number of amplification reactions was effective to maintain uniformity without compromising recombination.

While we found the manufacturer's total ssDNA quantification is accurate, the amplifiable proportion of ssDNA (presumably the fraction of full-length products) varies as much as 20-fold between oligo pool orders when quantified by qPCR (**Supplementary Figure 4D**). Correspondingly, the ng amount of oligo pool needed to maintain library uniformity varies by the same factor. Pending manufacturer improvements in oligo pool synthesis, quantification, and/or quality control, we can recommend two approaches for scaling PCR reactions to maintain uniformity:

1. *Conservative estimate*: Perform 33 PCR reactions of 25 µL with 12-24 pg/µL of oligo pool template (10-20 ng total template) for roughly 1,000-member libraries. For fewer PCR reactions, higher oligo template concentrations can be used but will result in predictably higher recombination rates (**Supplementary Figure 4C**). For smaller libraries, this may be the most straightforward approach.

2. *qPCR quantification*: For larger libraries, it may be desirable to determine a minimum number of reactions necessary to maintain uniformity. Perform qPCR quantification of oligo pools alongside a standard curve to determine the concentration of full length oligos. For our libraries, a per-oligo representation of 10^5 in oligo pool amplification has been sufficient to maintain 90:10 ratios (the ratio in abundance of the 90th percentile construct to the 10th percentile construct) of about 2. Perform the requisite number of reactions to amplify 10^5 full-length oligos per library member, while maintaining the effective template concentration to 12-24 pg/µL. Given the variability we have observed in oligo pool quantification, the minimum number of reactions for roughly 1,000-member libraries has varied between 1.5 and 30 reactions.

Following amplification, reactions were pooled and purified with the QIAquick PCR & Gel Cleanup Kit (Qiagen 28506) and quantified with the Qubit High Sensitivity dsDNA Quantitation assay (Thermo Fisher Scientific Q32854). Plasmid libraries were then constructed by restriction digestion and ligation. Entry vectors were digested with BsmBI-v2 (New England Biolabs R0739L), dephosphorylated with rSAP (New England Biolabs M0371L), and gel purified with the QIAquick PCR & Gel Cleanup Kit (Qiagen 28506). Amplified oligo libraries were digested with BsmBI-v2 (New England Biolabs R0739L) and purified with the QIAquick PCR & Gel Cleanup Kit (Qiagen 28506). The digested entry vector and amplified oligo libraries were quantified with the Qubit Broad Range dsDNA Quantitation assay (Thermo Fisher Scientific Q32853). Each 20 µL ligation reaction consisted of 20 femtomole (fmol) of BsmBI-digested amplified oligo pool, 20 fmol of digested backbone, 1X T4 DNA ligase buffer (New England Biolabs M0202S), and 400 U/µL of T4 DNA ligase (New England Biolabs M0202S). We observed that either 0.5:1 or 1:1 molar ratios of insert:backbone were optimal for library uniformity. Excess insert relative to backbone resulted in poor uniformity, with a clear bias determined by the identity of the ligation-adjacent base of iBAR 2 (**Supplementary Figure 4F**). Ligation reactions were incubated at 16 ºC overnight, then heat-inactivated for 10 minutes at 65 ºC. Reactions were purified with a 1.5X ratio of Ampure XP paramagnetic beads (Beckman Coulter A63880) and eluted in 10 µL of water. Purified ligations were electrotransformed by combining 5 µL of ligation and 25 µL of Endura electrocompetent cells (Biosearch Technologies 60242-2) in a 0.1 cm Gene Pulser cuvette (Biorad 1652083) on ice, electroporating on a Gene Pulser Xcell (Biorad 1652662) with the settings 1.8 kV, 600 Ohms, and 10 µF, and recovering for 90 minutes at 30 ºC in 1 mL of Endura recovery media (Biosearch Technologies 60242-2). Cultures were then grown for 16 hours in 50 mL of 2xYT media with 100 µg/mL of carbenicillin, shaking at 225 RPM at 30ºC. Plasmid libraries were then purified with a Plasmid Plus Midi Kit (Qiagen 12943) following the manufacturer's instructions. We observed assembly and transformation efficiencies averaging about 300,000 colony forming units per fmol of digested backbone input.

CROPseq-multi plasmid libraries were prepared for next-generation sequencing using a single-step PCR protocol to minimize amplification bias and PCR template switching. Plasmid libraries were amplified using 1X NEBNext Ultra II Q5 Master Mix (New England Biolabs M0544), 1X EvaGreen dye (Biotium 31000), 500

nM forward and reverse primers (Integrated DNA Technologies), and 100 ng of plasmid template per 25 µL reaction with the following thermal cycling conditions: 98 ºC for 1 min, 10 cycles of (98 ºC for 10 s, 67 ºC for 10 s, 72 ºC for 15 s), and 72 ºC for 1 min. Reactions were purified with a 1X ratio of Ampure XP paramagnetic beads (Beckman Coulter A63880). To assay deletion recombination events, we used primers flanking both sgRNA scaffolds and sequenced on a 300-cycle sequencing kit (e.g. Illumina MS-102-2002). For subsequent library sequencing without quantification of deletion events, we opted for primers that flank only the spacers and iBARs and used custom Illumina read primers (**Supplementary Table 2**). While this strategy does not capture the majority of deletion events, it enables the observation of all variable sequence elements (spacers, iBARs, and the middle tRNA) with 150-cycle sequencing kits (e.g. Illumina MS-102-3001) and positions spacers and iBARs as early as possible within NGS reads, where sequencing error rates are lowest. Code for CROPseq-multi library NGS analysis is available on GitHub.

Cell culture, lentiviral production, and transduction

HEK-293FT (Thermo Fisher Scientific R70007) and doxycycline-inducible-Cas9 A549[57] (a gift from J.T. Neal) cell lines were maintained in DMEM(1X) + GlutaMAX (Gibco 10569010) supplemented with 10% (v/v) heat-inactivated fetal bovine serum (MilliporeSigma F4135), 100 U/mL Penicillin and 100 µg/mL Streptomycin (Gibco 15140122). hTERT-immortalized RPE1 cells with Zim3-dCas9-2A-BFP[20] (a gift from Jonathan Weissman) were cultured in DMEM/F-12 + HEPES (Thermo Fisher Scientific 11330032) supplemented with 10% (v/v) heat-inactivated fetal bovine serum (MilliporeSigma F4135), 100 U/mL Penicillin and 100 µg/mL Streptomycin (Gibco 15140122), 0.01 mg/mL hygromycin B (Thermo Fisher Scientific 10687010). All cell lines were passaged with TrypLE (Thermo Fisher Scientific 12604013).

For lentiviral production, 6-well plates were seeded with 1 million HEK-293FT cells in 2 mL of media per well 24 hours prior to transfection. Lentiviral plasmids MD2.G (Addgene #12259), psPAX (Addgene #12260), and transfer plasmids were transfected at a mass ratio of 2:3:4 with Lipofectamine 3000 (Thermo Fisher Scientific L3000001) following the manufacturer's protocol. The media was replaced 4 hours post-transfection. Lentivirus was harvested 48 hours post-transfection. Media was collected and centrifuged at 400 xg for 5 min at 20 ºC, then the supernatant was collected and filtered with a 0.45 µM syringe-filter (VWR 28143-312). Lentivirus was then stored at -80 ºC in single-use aliquots.

Functional lentiviral titers were determined by transduction in the appropriate cell line (i.e. A549 or RPE1) with a lentiviral dilution series and quantifying survival. For A549 cells, lentiviral transduction was performed by mixing cells in media supplemented with 8 µg/mL of polybrene (Santa Cruz Biotechnology 28728-55-4) with lentivirus and centrifuging in cell culture plates at 1000 xg and 33 ºC for 2 hours. Media was replaced 24 hours post-transduction. Puromycin selection was initiated with the addition of 1 µg/mL puromycin (Thermo Fisher Scientific A1113803) 24 hours post-transduction. Cell Titer Glo (Promega G7573) was used to quantify survival after 48 hours of Puromycin selection, comparing survival against minus-virus/minus-selection positive control and minus-virus/plus-selection negative control. For RPE1 cells, lentiviral transduction was performed by adding lentivirus and 8 µg/mL of polybrene (MilliporeSigma TR-1003) to cells 24-hours after seeding. Media was replaced 24 hours post-transduction with 10 µg/mL puromycin (Thermo Fisher Scientific A1113803). Cell Titer Glo (Promega G7573) was used to quantify survival 4 days post-transduction. Cells were passaged in-well once to accelerate puromycin selection as the RPE1 line is resistant to low puromycin concentrations.

Percent survival was converted to MOI assuming a Poisson distribution for transduction:

$$MOI = -\ln(1-x), \text{ where } x = \text{percent survival.}$$

Dilutions of virus yielding 10%-60% survival were used to calculate titers as high and low extremes are more sensitive to experimental noise. Viral titer (genomic integrations per µL virus) was then determined from MOI:

$$\text{Titer} = (\text{MOI}) \bullet (\text{input number of cells}) / (\mu\text{L virus})$$

CRISPRi essential gene screen

To design guides for each gene target, we selected two guides from a published screen[20] and four guides using CRISPick[4,5]. All steps of the pooled enrichment screen were performed to maintain library representation of at least 1000X across both biological replicates. RPE1 cells expressing Zim3-dCas9-P2A-BFP were transduced with lentivirus at a MOI of 0.1. We added 10 µg/mL puromycin (Thermo Fisher Scientific A1113803) 24-hours post-transduction and harvested cells at days 7 and 14 post-transduction. Genomic DNA was purified with the QIAamp DNA Blood Midi Kit (Qiagen 51183). Libraries were then prepared for next-generation sequencing using a single-step PCR. Genomic DNA libraries were amplified using 1X NEBNext Ultra II Q5 Master Mix (New England Biolabs M0544), 1X EvaGreen dye (Biotium 31000), 500 nM forward and reverse primers (Integrated DNA Technologies), and up to 2.5 µg of genomic DNA template per 50 µL reaction with the following thermal cycling conditions: 98 ºC for 1 min, 24 cycles of (98 ºC for 10 s, 67 ºC for 10 s, 72 ºC for 15 s), and 72 ºC for 1 min. Reactions were pooled and purified with the QIAquick PCR & Gel Cleanup Kit (Qiagen 28506). We then gel purified reactions with the QIAquick PCR & Gel Cleanup Kit (Qiagen 28506). Libraries were quantified with the Qubit High Sensitivity dsDNA Quantitation assay (Thermo Fisher Scientific Q32854) and sequenced on an Illumina MiSeq with a 150-cycle kit (Illumina MS-102-3001) with 75 cycles for each read and 8 cycles for each index. Code for NGS analysis is available on GitHub.

Quantification of endogenous editing by next-generation sequencing

A549 cells were transduced and selected with puromycin for five days as described above. To induce Cas9 expression and editing, transduced cells were cultured for 7 days with 1 µg/mL of doxycycline. Genomic DNA was harvested by discarding culture media and adding lysis buffer consisting of 20 mM Tris pH 8 and 0.1% Triton X-100 (MilliporeSigma T9284) with 60 ng/mL of Proteinase K (New England Biolabs P8107S) added immediately prior to use. Lysate was incubated at 65 ºC for 6 min, then 95 ºC for 2 min, and stored at -20ºC. Endogenous loci were amplified with 1X Q5 High-Fidelity Master Mix (New England Biolabs M0492), 1 M Betaine (MilliporeSigma B0300), 500 nM forward and reverse primers (Integrated DNA Technologies), and gDNA in cell lysate of at least 5,000 cells, and the following thermal cycling protocol: 98 ºC for 1 min, then 30 cycles of (98 ºC for 10 s, 62 ºC for 10 s, and 72 ºC for 20 s), then 72 ºC for 2 min. Amplicons were then barcoded for NGS by combining 2 µL of amplicon from the previous PCR with 1X Q5 High-Fidelity Master Mix (New England Biolabs M0492) and 500 nM forward and reverse indexing primers (Integrated DNA Technologies), and cycling with the following conditions: 98 ºC for 1 min, then 10 cycles of (98 ºC for 10 s, 60 ºC for 10 s, and 72 ºC for 15 s), then 72 ºC for 2 min. Amplicons were then pooled, gel purified with the QIAquick PCR & Gel Cleanup Kit (Qiagen 28506), quantified with the Qubit High Sensitivity dsDNA Quantitation assay (Thermo Fisher Scientific Q32854), and sequenced on an Illumina MiSeq. Genome editing was quantified with CRISPResso2[58].

In situ amplification and sequencing by synthesis

The *in situ* amplification protocol was modified from our previous studies[12,51]. Cells were fixed in 4% (v/v) formaldehyde (Electron Microscopy Sciences 15714) and 0.007% (v/v) glutaraldehyde (Electron

Microscopy Sciences 16120) in 1X PBS (Ambion AM9625) for 30 minutes at room temperature, then washed twice in PBS. We have observed that the use of glutaraldehyde in the primary fixation step can impact some immunofluorescence stains; omission or titration to lower concentrations may offer a balance of detection sensitivity and compatibility with phenotype measurements. Samples were permeabilized in 1X PBS + 0.2% Tween-20 (VWR 100216-360) for 15 minutes at room temperature, then washed twice in 1X PBS + 0.1% Tween-20, henceforth "PBS-T". For phenotypic measurements prior to reverse transcription and cDNA-fixation, we recommend using RiboLock Rnase Inhibitor (Thermo Fisher Scientific EO0384) together with RNase-free reagents to preserve mRNA integrity.

The reverse transcription solution was prepared with the following composition: 1X RevertAid RT buffer (Thermo Fisher Scientific EP0452), 250 µM dNTPs (New England Biolabs N0447L), 1 µM each biotinylated reverse transcription primer (Integrated DNA Technologies), 200 µg/mL molecular biology grade recombinant albumin (rAlbumin) (New England Biolabs B9200S), 0.8 U/µL RiboLock RNase inhibitor (Thermo Fisher Scientific EO0384), and 4.8 U/µL RevertAid H minus Reverse Transcriptase (Thermo Fisher Scientific EP0452). Samples were incubated in reverse transcription solution for 16 hours at 37 ºC. Samples were then washed twice with PBS-T and incubated for 15 minutes with 20 µg/mL Streptavidin (New England Biolabs N7021S) and 100 µg/mL rAlbumin (New England Biolabs B9200S) in 1X PBS. Next, samples were washed twice with PBS-T prior to post-fixation in 3% formaldehyde (Electron Microscopy Sciences 15714) and 0.1% glutaraldehyde (Electron Microscopy Sciences 16120) for 30 minutes at room temperature. After fixation, samples were washed twice with PBS-T and incubated in gapfill and ligation solution at 37 ºC for 5 min, followed by 45 ºC for 90 minutes. The gapfill and ligation solution was composed of 1X Ampligase buffer (Lucigen A3210K), 50 nM dNTPs (New England Biolabs N0447L), 0.1 µM each padlock probe, 200 µg/mL rAlbumin (New England Biolabs B9200S), 0.4 U/µL RNase H (Enzymatics Y9220L), 0.02 U/µL TaqIT polymerase (Enzymatics P7620L), and 0.5 U/µL Ampligase (Lucigen A3210K). Samples were then washed twice with PBS-T and incubated in RCA solution for 16 hours at 30 ºC. RCA solution was composed of 1X Phi29 buffer (Thermo Fisher Scientific EP0091), 5% glycerol (MilliporeSigma G5516), 250 µM dNTPs (New England Biolabs N0447L), 200 µg/mL rAlbumin (New England Biolabs B9200S), and 1 U/µL Phi29 DNA polymerase (Thermo Fisher Scientific EP0091). Following RCA, samples were washed twice in PBS-T and incubated with 1 µM each sequencing primer (Integrated DNA Technologies) in 2X SSC buffer (Ambion AM9763) for 30 minutes at room temperature, followed by two PBS-T washes.

Sequencing by synthesis was then performed as previously described[12,51]. Briefly, samples were incubated in incorporation mix (MiSeq Nano kit v2 reagent 1) (Illumina MS-103-1003) for 5 minutes at 60 ºC on a flat-top thermal cycler, then washed six times with PR2 buffer (Illumina MS-103-1003), followed by 5 heated washes in PR2, 5 min each at 60 ºC. Samples were imaged in 2X SSC + 200 ng/mL DAPI (MilliporeSigma D9542) on a Nikon Ti2 Microscope at 10X magnification. To proceed to the next cycle, samples were incubated in cleavage mix (MiSeq Nano kit v2 reagent 4) (Illumina MS-103-1003) for 6 min at 60 ºC, then washed three times in PR2 followed by three heated PR2 washes of 1 min each at 60 ºC. Samples were then ready to return to the incorporation step for the subsequent sequencing cycle. *In situ* sequencing images were analyzed as previously described[12,51].

Comparison of decoding efficiencies of *in situ* detection approaches

To determine cycling requirements for spacer sequences, we performed 50 simulations at each library size by randomly sampling guides from the Dolcetto genome-wide CRISPRi library[5]. The same approach applied to sgRNAs sampled from CRISPR-KO (Brunello) and CRISPRa (Calabrese) libraries (data not

shown) were indistinguishable from the CRISPRi library. Single barcode encoding was approximated by the step function:

$$\text{number of unique barcodes} = 4^n, \text{ where n is the number of sequencing cycles}$$

For dual barcodes, we assumed that reads could be assigned to the appropriate iBAR position in a sequence-independent manner, such as through the use of a fluorophore-conjugated oligo to label one iBAR. Dual barcode encoding was approximated by the step function:

$$\text{number of unique barcodes} = 4^{2n}, \text{ where n is the number of sequencing cycles}$$

Dual barcode encoding with 95% recombination detection was approximated by the step function:

$$\text{number of unique barcodes} = 4^n \bullet \text{floor}[0.05 \bullet 4^n], \text{ where n is the number of sequencing cycles}$$

Practically, barcodes will be constrained by restriction enzyme sites, homopolymers, and GC content. Barcodes may be restricted further for edit distance to enable error detection and/or error correction, however we have not explored these considerations in this simple approximation.

Acknowledgements

We would like to thank the following individuals for helpful discussions and feedback on the manuscript: Michael Ward, Andrew Bassett, John Doench, Christoph Bock, Rebecca Carlson, Mohamad Najia, Frances Keer, and Owen Andrews, as well as all members of the Blainey Lab. Andrew Bassett engaged throughout the project in helpful discussions about molecular cloning workflows and tRNA-based sgRNA multiplexing systems. John Doench helped make reagents available through the Broad Institute Genetic Perturbation Platform's modular vector assembly platform, Fragmid. We would like to acknowledge Rebecca Dertinger for administrative support and Robert Majovski for manuscript proofreading. R.T.W. is supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 1745302. YQ is supported by the National Institutes of Health (NIH) under grant K00 CA264422 and by funding from the Eric and Wendy Schmidt Center at the Broad Institute of MIT and Harvard.

Author contributions

R.T.W. conceived of the project, designed and performed experiments, analyzed data, and wrote the manuscript. Y.Q. designed and performed experiments and edited the manuscript. P.C.B. supervised the work and edited the manuscript.

Data availability

Data is available upon request.

Code availability

Code is available on GitHub:

- CROPseq-multi library design and sequencing: https://github.com/rtwalton/CROPseq-multi
- DNA barcode design: https://github.com/feldman4/dna-barcodes
- *in situ* sequencing data analysis: https://github.com/feldman4/OpticalPooledScreens

Conflict of Interest

RTW and PCB are inventors on a patent application related to this work. PCB is an inventor on patents relating to optical pooled screening technologies. PCB is a consultant to and/or holds equity in companies that develop or apply biotechnologies: 10X Genomics, General Automation Lab Technologies/Isolation Bio, Celsius Therapeutics, Next Gen Diagnostics, Cache DNA, Concerto Biosciences, Stately, Ramona Optics, Bifrost Biosystems, and Amber Bio. His laboratory has received research funding from Calico Life Sciences, Merck, and Genentech for work related to genetic screening.

References

1. Datlinger, P. *et al.* Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297–301 (2017).
2. Wong, N., Liu, W. & Wang, X. WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system. *Genome Biol.* **16**, 218 (2015).
3. H, X. *et al.* Sequence determinants of improved CRISPR sgRNA design. *Genome Res.* **25**, (2015).
4. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
5. Sanson, K. R. *et al.* Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat. Commun.* **9**, 5416 (2018).
6. Chuai, G. *et al.* DeepCRISPR: optimized CRISPR guide RNA design by deep learning. *Genome Biol.* **19**, 80 (2018).
7. Kim, H. K. *et al.* SpCas9 activity prediction by DeepSpCas9, a deep learning–based model with high generalization performance. *Sci. Adv.* **5**, eaax9249 (2019).
8. Petiwala, S. *et al.* Optimization of Genomewide CRISPR Screens Using AsCas12a and Multi-Guide Arrays. *CRISPR J.* **6**, 75–82 (2023).
9. Replogle, J. M. *et al.* Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat. Biotechnol.* **38**, 954–961 (2020).
10. Liu, J. *et al.* Pooled library screening with multiplexed Cpf1 library. *Nat. Commun.* **10**, 3144 (2019).
11. Yin, J.-A. *et al.* Robust and Versatile Arrayed Libraries for Human Genome-Wide CRISPR Activation, Deletion and Silencing. 2022.05.25.493370 Preprint at https://doi.org/10.1101/2022.05.25.493370 (2023).
12. Feldman, D. *et al.* Optical Pooled Screens in Human Cells. *Cell* **179**, 787-799.e17 (2019).
13. Yao, D. *et al.* Scalable genetic screening for regulatory circuits using compressed Perturb-seq. *Nat. Biotechnol.* 1–14 (2023) doi:10.1038/s41587-023-01964-9.
14. Thompson, N. A. *et al.* Combinatorial CRISPR screen identifies fitness effects of gene paralogues. *Nat. Commun.* **12**, 1302 (2021).

15. Kegel, B. D., Quinn, N., Thompson, N. A., Adams, D. J. & Ryan, C. J. Comprehensive prediction of robust synthetic lethality between paralog pairs in cancer cell lines. *Cell Syst.* **12**, 1144-1159.e6 (2021).
16. Dede, M., McLaughlin, M., Kim, E. & Hart, T. Multiplex enCas12a screens detect functional buffering among paralogs otherwise masked in monogenic Cas9 knockout screens. *Genome Biol.* **21**, 262 (2020).
17. Parrish, P. C. R. *et al.* Discovery of synthetic lethal and tumor suppressor paralog pairs in the human genome. *Cell Rep.* **36**, 109597 (2021).
18. Ryan, C. J., Mehta, I., Kebabci, N. & Adams, D. J. Targeting synthetic lethal paralogs in cancer. *Trends Cancer* **9**, 397–409 (2023).
19. Adamson, B. *et al.* A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* **167**, 1867-1882.e21 (2016).
20. Replogle, J. M. *et al.* Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell* **185**, 2559-2575.e28 (2022).
21. Hill, A. J. *et al.* On the design of CRISPR-based single cell molecular screens. *Nat. Methods* **15**, 271–274 (2018).
22. Feldman, D., Singh, A., Garrity, A. J. & Blainey, P. C. Lentiviral co-packaging mitigates the effects of intermolecular recombination and multiple integrations in pooled genetic screens. 262121 Preprint at https://doi.org/10.1101/262121 (2018).
23. Adamson, B., Norman, T. M., Jost, M. & Weissman, J. S. Approaches to maximize sgRNA-barcode coupling in Perturb-seq screens. 298349 Preprint at https://doi.org/10.1101/298349 (2018).
24. Najm, F. J. *et al.* Orthologous CRISPR–Cas9 enzymes for combinatorial genetic screens. *Nat. Biotechnol.* **36**, 179–189 (2018).
25. Li, R. *et al.* Comparative optimization of combinatorial CRISPR screens. *Nat. Commun.* **13**, 2469 (2022).
26. Hegde, M., Strand, C., Hanna, R. E. & Doench, J. G. Uncoupling of sgRNAs from their associated barcodes during PCR amplification of combinatorial CRISPR screens. *PLoS ONE* **13**, e0197547 (2018).

27. Zetsche, B. *et al.* Cpf1 Is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).
28. Zetsche, B. *et al.* Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nat. Biotechnol.* **35**, 31–34 (2017).
29. DeWeirdt, P. C. *et al.* Optimization of AsCas12a for combinatorial genetic screens in human cells. *Nat. Biotechnol.* **39**, 94–104 (2021).
30. Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* **168**, 20–36 (2017).
31. Moon, S. B., Kim, D. Y., Ko, J.-H. & Kim, Y.-S. Recent advances in the CRISPR genome editing tool set. *Exp. Mol. Med.* **51**, 1–11 (2019).
32. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **38**, 824–844 (2020).
33. Dixit, A. *et al.* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853-1866.e17 (2016).
34. Xie, S., Duan, J., Li, B., Zhou, P. & Hon, G. C. Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells. *Mol. Cell* **66**, 285-299.e5 (2017).
35. Jaitin, D. A. *et al.* Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell* **167**, 1883-1896.e15 (2016).
36. Xie, K., Minkenberg, B. & Yang, Y. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc. Natl. Acad. Sci.* **112**, 3570–3575 (2015).
37. Port, F. & Bullock, S. L. Augmenting CRISPR applications in Drosophila with tRNA-flanked sgRNAs. *Nat. Methods* **13**, 852–854 (2016).
38. Xu, L., Zhao, L., Gao, Y., Xu, J. & Han, R. Empower multiplex cell and tissue-specific CRISPR-mediated gene manipulation with self-cleaving ribozymes and tRNA. *Nucleic Acids Res.* **45**, e28–e28 (2017).
39. Qi, W. *et al.* High-efficiency CRISPR/Cas9 multiplex gene editing using the glycine tRNA-processing system-based strategy in maize. *BMC Biotechnol.* **16**, 58 (2016).
40. Zhang, Y. *et al.* A gRNA-tRNA array for CRISPR-Cas9 based rapid multiplexed genome editing in

Saccharomyces cerevisiae. *Nat. Commun.* **10**, 1053 (2019).

41. Knapp, D. J. H. F. *et al.* Decoupling tRNA promoter and processing activities enables specific Pol-II Cas9 guide RNA expression. *Nat. Commun.* **10**, 1490 (2019).
42. Yuan, Q. & Gao, X. Multiplex base- and prime-editing with drive-and-process CRISPR arrays. *Nat. Commun.* **13**, 2771 (2022).
43. Zhu, S. *et al.* Guide RNAs with embedded barcodes boost CRISPR-pooled screens. *Genome Biol.* **20**, 20 (2019).
44. Urbinati, F. *et al.* Mechanism of Reduction in Titers From Lentivirus Vectors Carrying Large Inserts in the 3′LTR. *Mol. Ther.* **17**, 1527–1536 (2009).
45. McGee, A. V. *et al.* Modular vector assembly enables rapid assessment of emerging CRISPR technologies. *Cell Genomics* **4**, (2024).
46. Schlub, T. E., Smyth, R. P., Grimm, A. J., Mak, J. & Davenport, M. P. Accurately Measuring Recombination between Closely Related HIV-1 Genomes. *PLoS Comput. Biol.* **6**, e1000766 (2010).
47. Meyers, R. M. *et al.* Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).
48. Dempster, J. M. *et al.* Extracting Biological Insights from the Project Achilles Genome-Scale CRISPR Screens in Cancer Cell Lines. 720243 Preprint at https://doi.org/10.1101/720243 (2019).
49. Funk, L. *et al.* The phenotypic landscape of essential human genes. *Cell* **185**, 4634-4653.e22 (2022).
50. Carlson, R. J., Leiken, M. D., Guna, A., Hacohen, N. & Blainey, P. C. A genome-wide optical pooled screen reveals regulators of cellular antiviral responses. *Proc. Natl. Acad. Sci.* **120**, e2210623120 (2023).
51. Feldman, D. *et al.* Pooled genetic perturbation screens with image-based phenotypes. *Nat. Protoc.* 1–37 (2022) doi:10.1038/s41596-021-00653-8.
52. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
53. Kudo, T. *et al.* Highly multiplexed, image-based pooled screens in primary cells and tissues with PerturbView. 2023.12.26.573143 Preprint at https://doi.org/10.1101/2023.12.26.573143 (2023).
54. Fandrey, C. I. *et al.* Cell Type-Agnostic Optical Perturbation Screening Using Nuclear In-Situ

Sequencing (NIS-Seq). 2024.01.18.576210 Preprint at https://doi.org/10.1101/2024.01.18.576210 (2024).

55. Michlits, G. *et al.* CRISPR-UMI: single-cell lineage tracing of pooled CRISPR–Cas9 screens. *Nat. Methods* **14**, 1191–1197 (2017).
56. Schmierer, B. *et al.* CRISPR/Cas9 screening using unique molecular identifiers. *Mol. Syst. Biol.* **13**, 945 (2017).
57. Ramezani, M. *et al.* A genome-wide atlas of human cell morphology. 2023.08.06.552164 Preprint at https://doi.org/10.1101/2023.08.06.552164 (2023).
58. Clement, K. *et al.* CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).
59. Replogle, J. M. *et al.* Maximizing CRISPRi efficacy and accessibility with dual-sgRNA libraries and optimal effectors. *eLife* **11**, e81856 (2022).
60. Sack, L. M., Davoli, T., Xu, Q., Li, M. Z. & Elledge, S. J. Sources of Error in Mammalian Genetic Screens. *G3 GenesGenomesGenetics* **6**, 2781–2790 (2016).
61. Labitigan, R. L. D. *et al.* Mapping variation in the morphological landscape of human cells with optical pooled CRISPRi screening. *eLife* **13**, (2024).
62. Xie, S., Cooley, A., Armendariz, D., Zhou, P. & Hon, G. C. Frequent sgRNA-barcode recombination in single-cell perturbation assays. *PLoS ONE* **13**, e0198635 (2018).

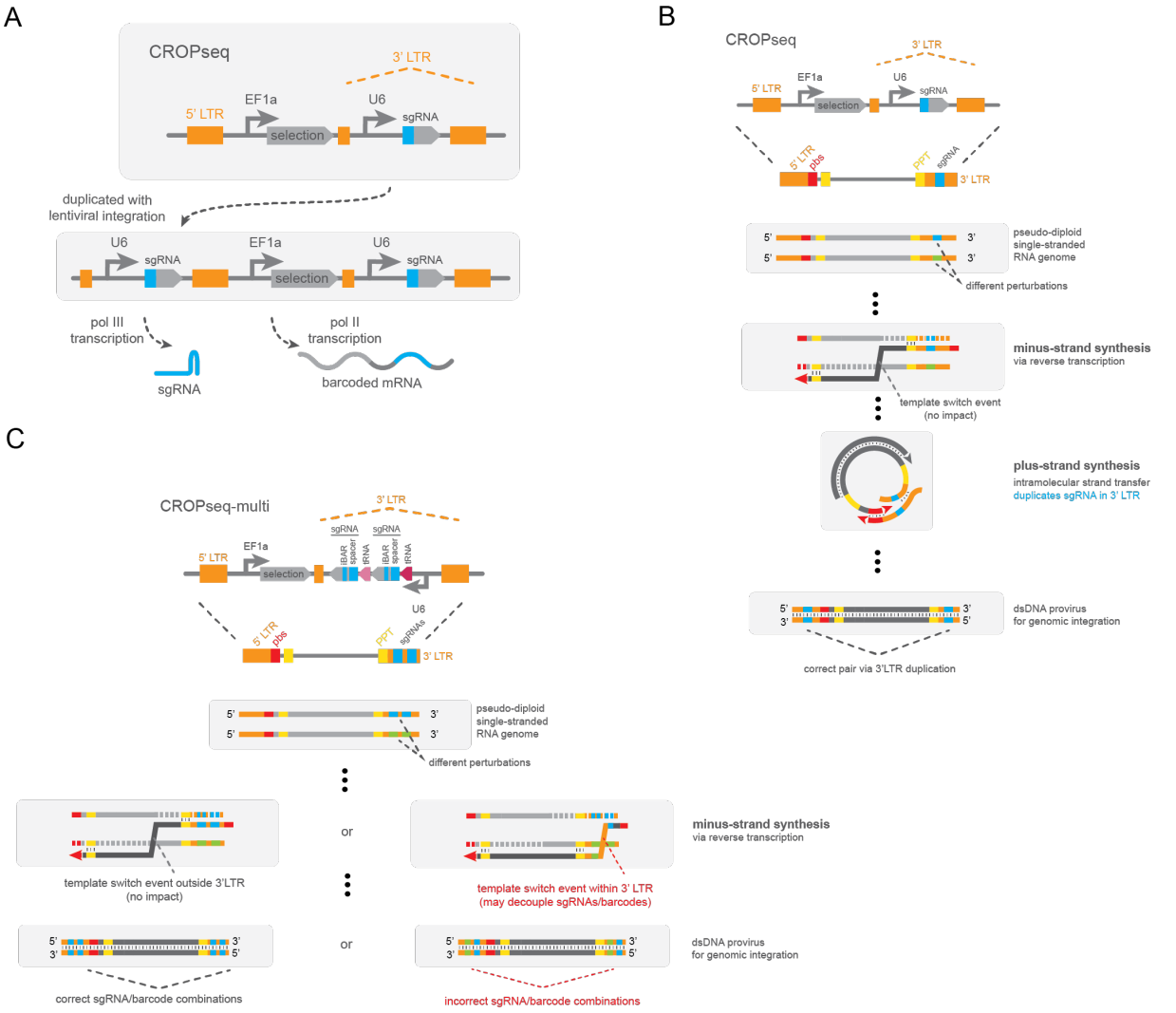
Tables

vector	Cas effector(s)	perturbation identifier	spacer-BC recombination	mRNA barcode	multiplex	spacer- spacer recombination
pLentiGuide and similar	SpCas9	spacer	n/a	no	up to 2*	26%[59]
LentiGuideBC and similar	SpCas9	linked BC	50%[21]	yes	no	n/a
CROPseq	SpCas9	spacer	n/a	yes	no	n/a
Big Papi	SpCas9 and SaCas9	linked BC	5%[26]	no	2	9%[26]
Cas12a crRNA array	Cas12a	spacer	n/a	no	3+	<1%**
CROPseq-multi	SpCas9	linked BC	<1% (this work)	yes	up to 2	8% (this work)

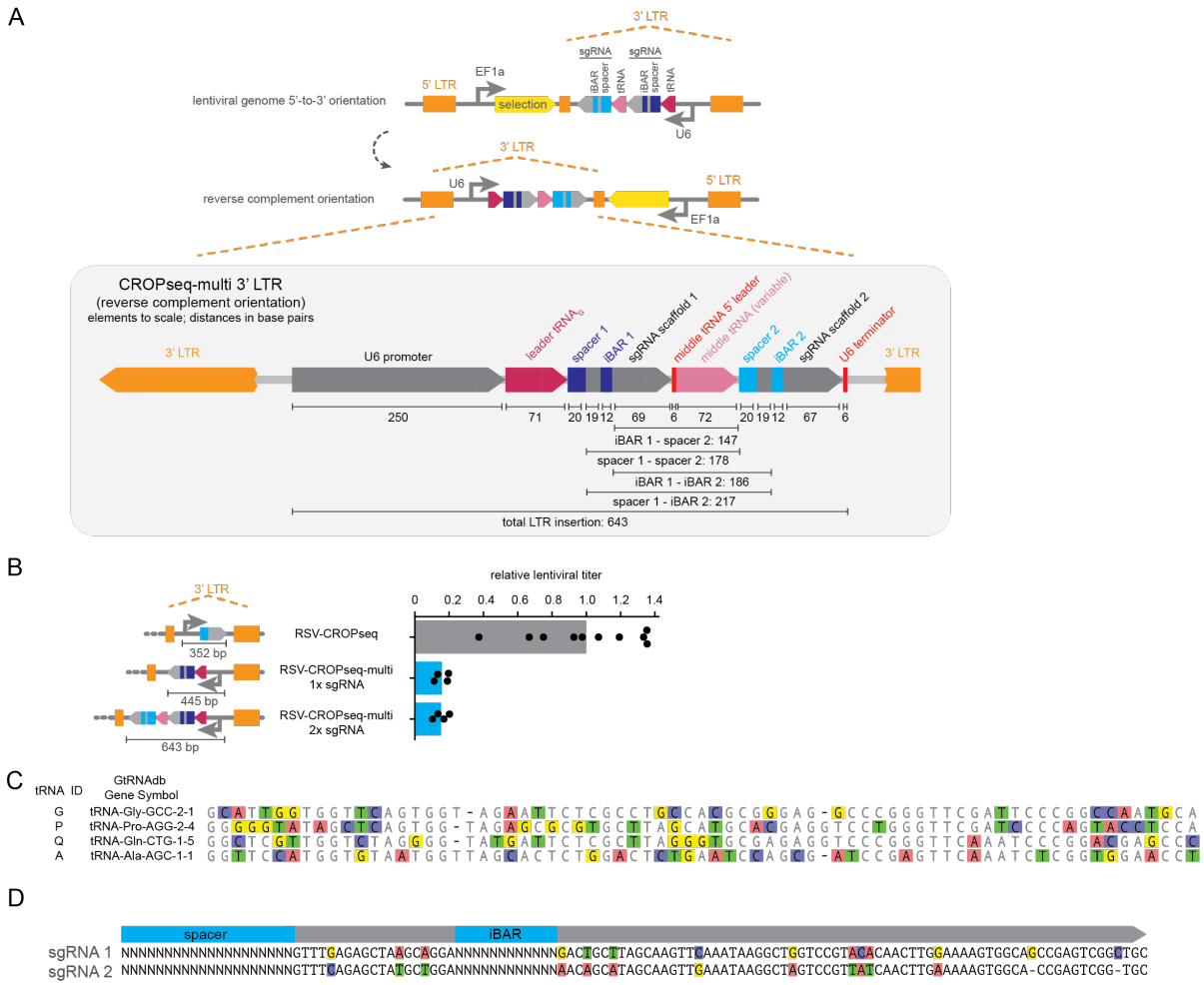
Table 1. Comparison of selected lentiviral gRNA delivery systems for barcoding and multiplexing.
*Up to 2 for pooled library construction; up to 3 have been demonstrated for arrayed library construction[19].
**Predicted based on distance. BC, barcode.

Supplementary Figures

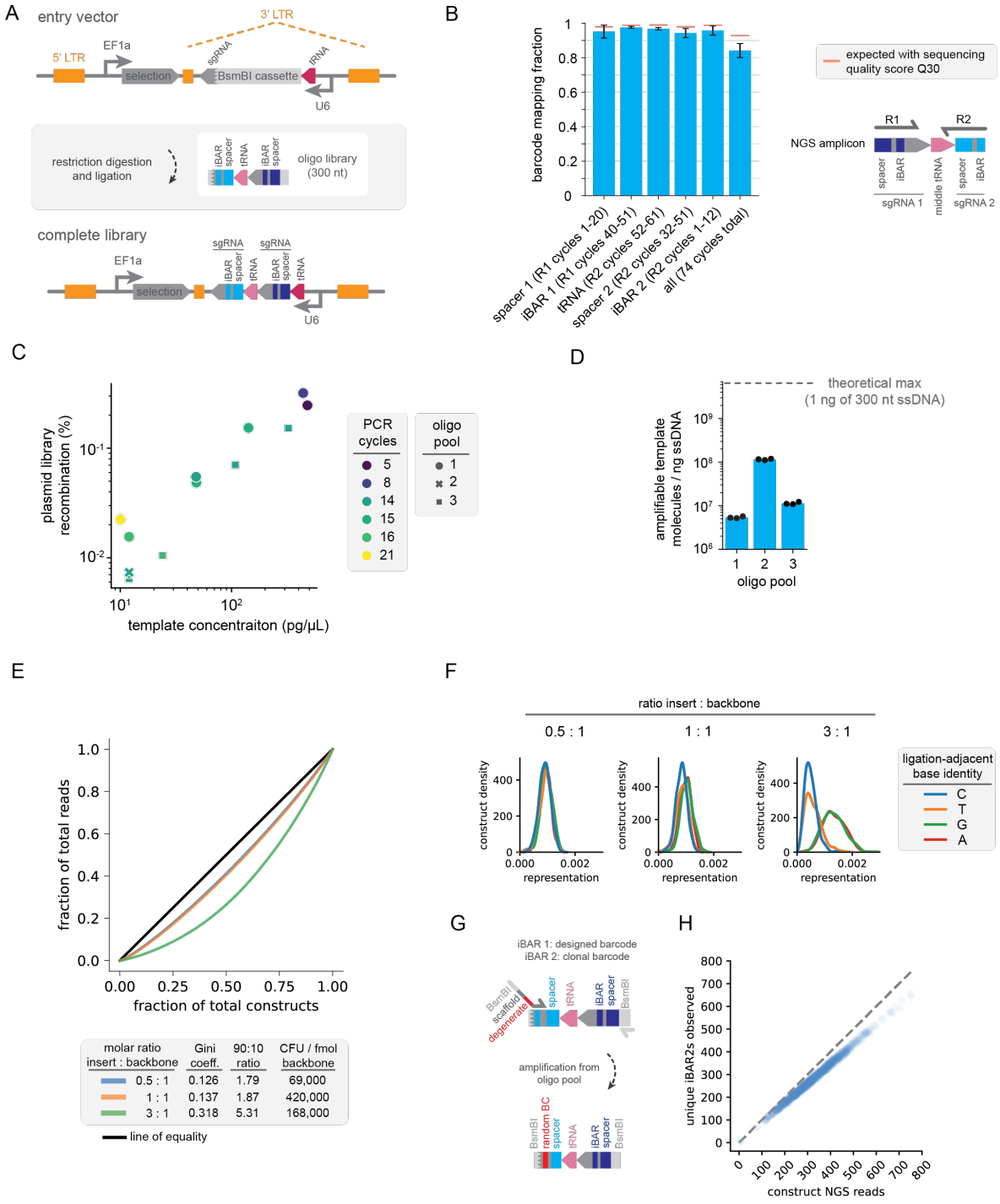
Supplementary Figure 1. (A-B) Schematic representations of a standard single-guide lentiviral vector (A) and derivative multiplexing systems (B). (C-D) Processing of the lentiviral RNA genome into double stranded DNA for genome integration for generic vectors (C) and those with multiple guides or barcode elements (D) to illustrate steps vulnerable to recombination. Illustration inspired by Adamson *et al*.[23] (E) Schematic representations of LentiGuideBC vector design for pairing guide RNAs with mRNA barcodes. LTR, long terminal repeats; pbs, primer binding site; PPT, polypurine tract; cPPT, central polypurine tract.



Supplementary Figure 2. (A) Schematic representation of the CROPseq vector design. (B) Abbreviated schematic of lentiviral RNA genome processing into double stranded DNA for genome integration for CROPseq vectors, highlighting the intramolecular 3' LTR duplication that is not vulnerable to recombination. (C) Abbreviated schematic of lentiviral RNA genome processing into double stranded DNA for genome integration for CROPseq-multi vectors. CROPseq-multi remains vulnerable to recombination during minus strand synthesis, but is robust against recombination during 3' LTR duplication. LTR, long terminal repeats; pbs, primer binding site; PPT, polypurine tract; cPPT, central polypurine tract.

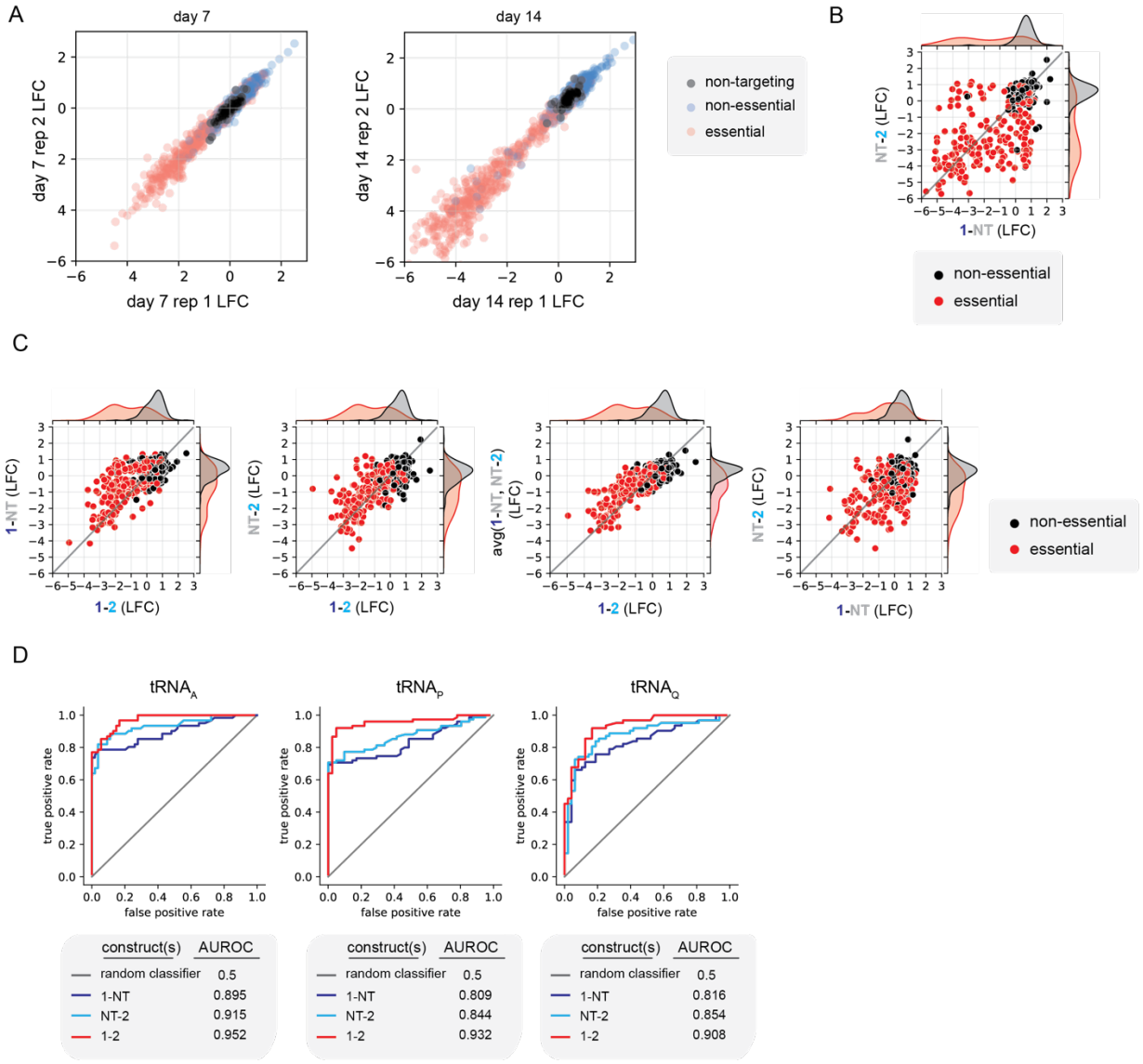


Supplementary Figure 3. (A) Detailed illustration of the CROPseq-multi 3' LTR design. (B) Lentiviral titers of RSV-CROPseq-multi vectors relative to CROPseq. (C) Sequence alignment of orthogonal tRNAs tested in CROPseq-multi. (D) Sequence alignment of orthogonal sgRNA scaffolds used in CROPseq-multi. LTR, long terminal repeats; RSV, Rous sarcoma virus.

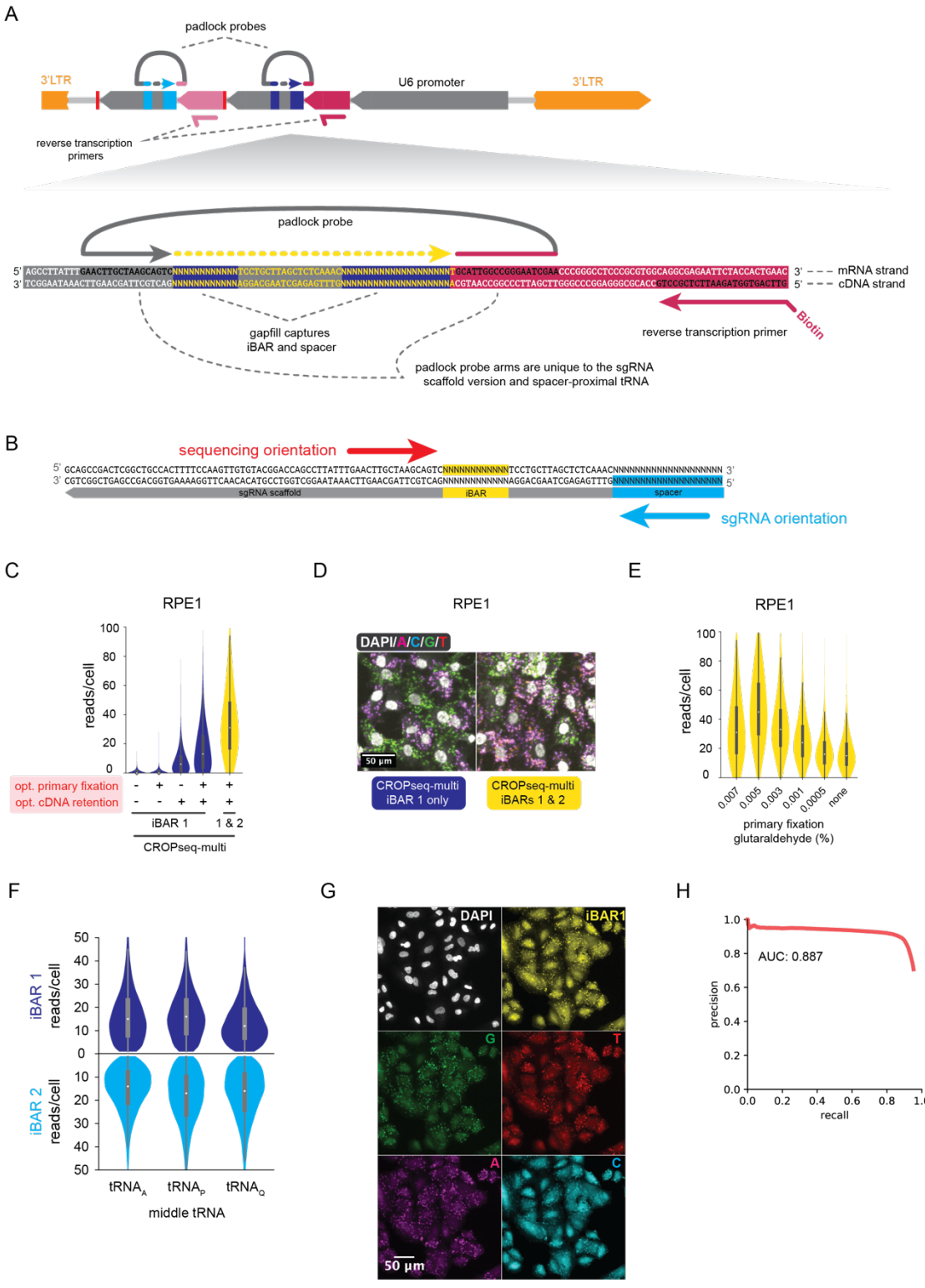


Supplementary Figure 4. (A) Schematic of pooled CROPseq-multi library construction. (B) Mapping rates of individual barcode elements in a CROPseq-multi library, determined by next-generation sequencing. “All” is the mapping rate of all elements (spacer 1, iBAR1, middle tRNA, spacer 2, and iBAR 2) in a read. In addition to oligo synthesis and amplification errors, sequencing error rates may explain a fraction of unmapped barcode elements. Only 10 nucleotides are used to uniquely identify the middle tRNA. (C) Recombination in CROPseq-multi plasmid libraries as a function of effective template concentration and polymerase chain reaction quantification amplification cycle number. (D) Quantitative polymerase chain

reaction quantification of three oligo pool orders. (E) Lorenz plots, Gini coefficients, and 90:10 ratios for 1,080-member CROPseq-multi plasmid libraries built with different assembly conditions. 90:10 ratios are the ratio in abundance of the 90th percentile construct to the 10th percentile construct. (F) Restriction-ligation assembly bias based on the identity of the variable ligation-adjacent base (encoded by iBAR2) is corrected by equimolar or limiting ratios of insert to backbone. (G) Schematic of CROPseq-multi oligo library amplification for clonal barcoding applications. (H) High diversity of iBAR2 sequences for a 1,080-member, clonally-barcoded CROPseq-multi library exceeds next-generation sequencing depth. LTR, long terminal repeats; NGS, next-generation sequencing; ssDNA, single-stranded DNA



Supplementary Figure 5. (A) Scatterplots of log2 fold changes (LFC) for construct abundance across biological replicates at days 7 and 14, colored by gene target as non-targeting control, non-essential, or essential. (B) Scatterplot of LFCs for single-targeting constructs at day 14, colored by gene essentiality. (C) Scatterplots of construct LFCs at day 7 for dual-targeting, single-targeting, and single-targeting averages, colored by gene essentiality. (D) Receiver operator characteristic curves for classification of gene essentiality for dual-targeting and single-targeting constructs and each middle tRNA identity. LFC, log2 fold change; NT, non-targeting; AUROC, area under receiver operator characteristic curve.



Supplementary Figure 6. (A) Illustration of oligonucleotide reagent design for *in situ* detection of CROPseq-multi barcodes. (B) Illustration of the orientation of *in situ* sequencing relative to the sgRNA. (C) Optimization of CROPseq-multi *in situ* detection in RPE1 cells. (D) Representative images of CROPseq-multi *in situ* detection in RPE1 cells with an optimized protocol. (E) Optimization of glutaraldehyde

concentration in the primary fixation step for multiplexed *in situ* detection of CROPseq-multi in RPE1 cells. (F) Comparison of the detection efficiency, in A549 cells, of iBARs 1 and 2 across constructs employing three orthogonal middle tRNAs. (G) Representative image showing the selective labeling of iBAR 1 reads with a fluorescent oligo in a multiplexed detection experiment, together with DAPI-stained nuclei and the four sequencing bases in separate fluorescent channels. (H) Precision-recall curve for assignment of individual reads to either iBAR 1 or iBAR 2 on signal from the iBAR-1-specific probe. LTR, long terminal repeats; AUC: area under curve.

Supplementary Data

Supplementary Table 1

distance (bp)	barcode swapping (%)	system	method	notes	ref.
17	2	Big Papi	NGS	Barcode-barcode distance; PCR recombination subtracted from total recombination in gDNA to estimate lentiviral recombination	26
17	0.5	CROPseq-multi	NGS	spacer-iBAR distance	this work
75	6.4	CROPseq-multi	NGS	iBAR1-spacer2 homologous distance (orthogonal middle tRNA)	this work
82	5	Big Papi	NGS	Spacer-barcode distance; PCR recombination subtracted from total recombination in gDNA to estimate lentiviral recombination	26
96	6	Barcoded ORFs	clone screening		60
108	7.5	CROPseq with linked barcode	NGS, digestion-qPCR		61
147	11.3	CROPseq-multi	NGS	iBAR 1 - spacer 2 distance (same middle tRNA)	this work
181	9	Big Papi	NGS	Spacer-spacer distance; PCR recombination subtracted from total recombination in gDNA to estimate lentiviral recombination; total distance is 193 bp but 12 bp are variable barcodes	26
400	26.33	Serial U6-sgRNA	Direct-capture Perturb-Seq	average of recombination in 3 cell lines	59
720	28.5	Barcoded ORFs	clone screening		60
1700	35.15	LentiGuideBC	clone screening	average of two pooled libraries	22
2000	30	Perturb-Seq	FACS with reporters		23
2400	50	pLGB-scKO	FACS with reporters		21
3000	50	Mosaic-seq	Self-circularization and NGS		62

Reporter barcode swapping rates with lentiviral barcoding systems. NGS, next-generation sequencing; FACS, fluorescence activated cell sorting; qPCR, quantitative polymerase chain reaction; ORF, open reading frame.

Supplementary Table 2 - see separate supplementary files