

# Pooled optical screens in human cells

David Feldman<sup>1,2†</sup>, Avtar Singh<sup>1†</sup>, Jonathan L. Schmid-Burgk<sup>1</sup>, Anja Mezger<sup>1,3,4</sup>, Anthony J. Garrity<sup>1</sup>, Rebecca J. Carlson<sup>1,5</sup>, Feng Zhang<sup>1,6,7,8,9</sup>, Paul C. Blainey<sup>1,6</sup>

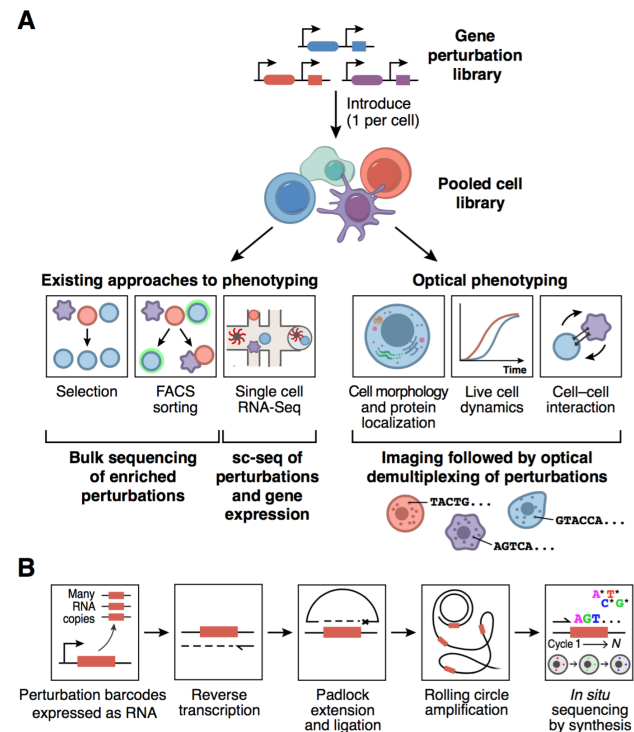
Large-scale genetic screens play a key role in the systematic discovery of genes underlying cellular phenotypes. Pooling of genetic perturbations greatly increases screening throughput, but has so far been limited to screens of enrichments defined by cell fitness and flow cytometry, or to comparatively low-throughput single cell gene expression profiles. Although microscopy is a rich source of spatial and temporal information about mammalian cells, high-content imaging screens have been restricted to much less efficient arrayed formats. Here, we introduce an optical method to link perturbations and their phenotypic outcomes at the single-cell level in a pooled setting. Barcoded perturbations are read out by targeted *in situ* sequencing following image-based phenotyping. We apply this technology to screen a focused set of 952 genes across >3 million cells for involvement in NF- $\kappa$ B activation by imaging the translocation of RelA (p65) to the nucleus, recovering 20 known pathway components and 3 novel candidate positive regulators of IL-1 $\beta$  and TNF $\alpha$ -stimulated immune responses.

## Introduction

Forward genetic screens are a powerful tool for finding genes that cause phenotypes. A variety of methods exist to disrupt genes, introduce exogenous genes, and modulate gene expression in screens of mammalian cells. Many such perturbations can be pooled together and efficiently quantified by next-generation sequencing (NGS) of cell populations. The phenotypic effect of a perturbation can then be defined as an enrichment in cells carrying the perturbation under different conditions (1, 2).

Examples of enrichment-based phenotypes that are compatible with pooled screens include differential cell fitness (e.g., under drug selection) and differential fluorescence of a marker assessed by fluorescence-activated cell sorting (FACS) (e.g. a genetic reporter or immunostained protein) (3). Although such screens have led to important biological discoveries, many complex phenotypes cannot be physically enriched at scale to provide a sample for NGS analysis. As a result, screens of more complex phenotypes have historically been limited to expensive and laborious testing of individual perturbations in arrayed formats. Recently, pooled CRISPR perturbations were integrated with single-cell expression profiling (4–7) to enable screens with a high-dimensional readout of cell state. Alternatively, image-

based screens have been used to access a rich set of spatially and temporally defined phenotypes across diverse processes in eukaryotic cells, including mitosis (8, 9), endocytosis (10), viral infection (11), differentiation (12), metabolism (13), DNA damage (14), autophagy (15) and synaptogenesis (16), but have been restricted to arrayed formats. Pooled screens of image-based phenotypes have thus far only been



**Fig. 1: Pooled optical genetic screens.** (A) In pooled screens, a library of genetic perturbations is introduced, typically at a single copy per target cell. In existing approaches, cellular phenotypes are evaluated by bulk NGS of enriched cell populations or single-cell gene expression profiling. In pooled optical screens, high-content imaging assays are used to extract rich spatiotemporal information from the sample prior to enzymatic amplification and *in situ* detection of RNA barcodes, enabling correlative measurement of phenotype and genotype. (B) Targeted *in situ* sequencing is used to read out RNA barcodes expressed from a single genomic integration. Barcode transcripts are fixed in place, reverse transcribed, and hybridized with single-stranded DNA padlock probes, which bind to common sequences flanking the barcode. The 3' arm of the padlock is extended and ligated, copying the barcode into a circularized ssDNA molecule, which then undergoes rolling circle amplification. The barcode sequence is then read out by multiple rounds of *in situ* sequencing-by-synthesis.

<sup>1</sup>Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. <sup>2</sup>Department of Physics, MIT, Cambridge, Massachusetts, USA.

<sup>3</sup>Department of Genetics, Stanford University, Stanford, California, USA. <sup>4</sup>Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm, Sweden. <sup>5</sup>Department of Health Sciences and Technology, MIT, Cambridge, Massachusetts, USA.

<sup>6</sup>Department of Biological Engineering, MIT, Cambridge, Massachusetts, USA. <sup>7</sup>McGovern Institute for Brain Research at MIT, Cambridge, Massachusetts, USA. <sup>8</sup>Department of Brain and Cognitive Science, MIT, Cambridge, Massachusetts, USA. <sup>9</sup>Howard Hughes Medical Institute, MIT, Cambridge, Massachusetts.

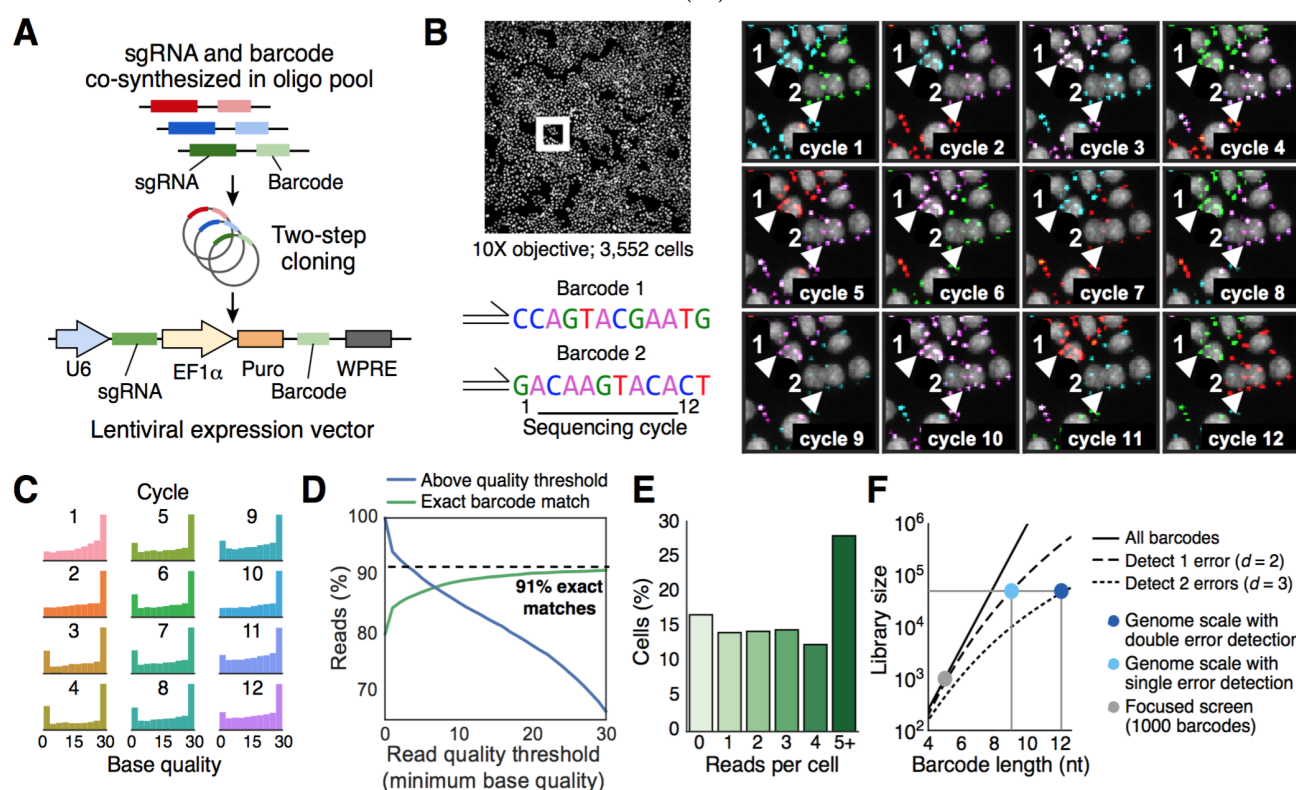
\*Correspondence to: [pblainey@broadinstitute.org](mailto:pblainey@broadinstitute.org) †These authors contributed equally to this work.

demonstrated in bacterial cells (17, 18).

To address this challenge, we developed a pooled genetic screening protocol compatible with a wide range of spatially and temporally resolved optical assays in mammalian cells, greatly expanding the variety of phenotypes amenable to high-throughput forward genetics. Our approach is to determine both the phenotype and the perturbation identity in each cell by microscopy (Fig. 1A), where the perturbations are identified by targeted *in situ* sequencing (19) of associated RNA barcodes. This strategy uses enzymatic amplification and sequencing-by-synthesis (Fig. 1B) to provide high signal levels from a compact, easily synthesized barcode. Our method adapts the existing molecular biology pipeline for pooled genetic screens, permitting information-rich phenotyping of thousands of perturbations across millions of cells in a single sample.

## Results

To apply this approach in the context of CRISPR screening, we encoded the identity of genetic perturbations by cloning guide RNAs (sgRNAs) and associated 12-nt barcodes into lentiGuide-Puro (20) (Addgene #52963), a widely-used sgRNA expression vector (Fig. 2A, Methods). Barcodes and flanking sequences were inserted into the 3' UTR of a Pol II-driven antibiotic resistance gene, a highly expressed mRNA suitable for *in situ* detection, in a design we termed lentiGuide-BC. We used a pooled cloning approach where sgRNAs and barcodes are synthesized in tandem on an oligo array, such that sgRNA-barcode pairings are pre-determined (Fig. 2A). As reported for other lentiviral libraries using paired sequences, we initially observed swapping of barcodes and associated sgRNAs in lentiGuide-BC cells due to reverse transcription-mediated recombination (21–24). To address recombination, we utilized a modified lentiviral packaging protocol that reduces barcode swapping from >28% to <5% (25).



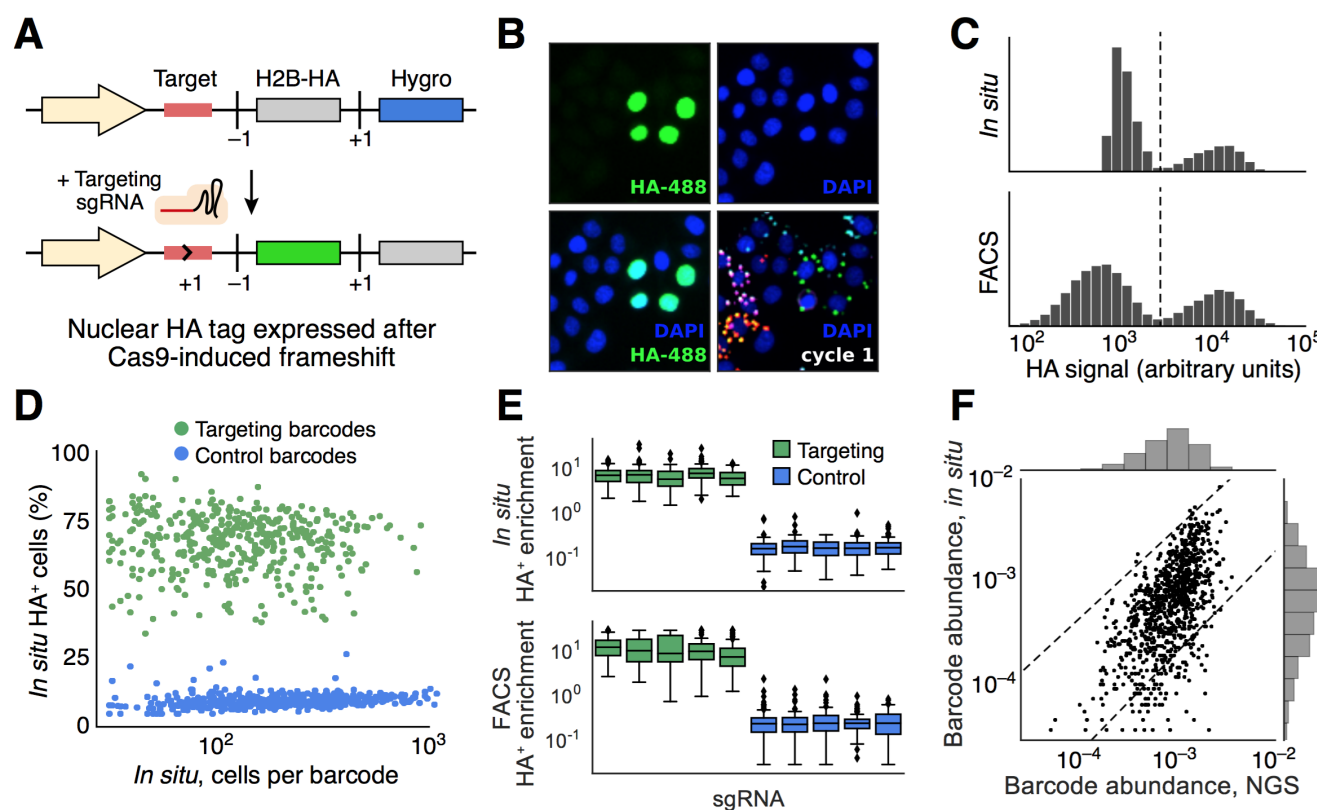
**Fig. 2: Identification of perturbation barcodes by *in situ* sequencing.** (A) A 125-nt oligo pool encoding perturbations (sgRNAs) and associated 12-nt barcodes was cloned into a lentiviral vector via two rounds of Golden Gate assembly. (B) Expressed barcode sequences were read out by padlock detection, rolling circle amplification, and 12 cycles of sequencing-by-synthesis (data shown for lentiGuide-BC). A linear filter (Laplacian-of-Gaussian, kernel width  $\sigma = 1$  pixel) was applied to sequencing channels to enhance spot-like features. (C) Per-base quality score over 12 cycles of *in situ* sequencing, calculated from signal for called base divided by signal for all bases. (D) >80% of barcodes map to 40 designed sequences out of 16.7 million possible 12-nt sequences. (E) Most cells contain multiple barcode reads. (F) The number of possible barcodes scales geometrically with barcode length. Sufficient 12-nt barcodes can be designed to cover a genome-scale perturbation library while maintaining the ability to detect and reject single or double sequencing errors (minimum pairwise Levenshtein distance  $d = 2$  or 3, respectively).

To test *in situ* identification of perturbations, we transduced a library of 40 sgRNA-barcode pairs into HeLa-TetR-Cas9 cells at low multiplicity of infection (MOI). We designed a reverse transcription primer and padlock probe complementary to constant regions flanking the barcode, and amplified the barcode via *in situ* reverse transcription, padlock extension/ligation, and rolling circle amplification (Fig. 1B, Fig. S1 and Table S1). Sequencing *in situ* using a 4-color sequencing-by-synthesis chemistry yielded bright, compact fluorescent spots that retained base specificity over 12 cycles (Fig. 2, B and C). Image segmentation and base calling analysis produced sequence reads with a mapping rate of >85% to the set of known barcodes (>85%) (Fig. 2D).

For optical screening to be scalable, the *in situ* readout step should be able to process millions of cells within a few days, ensuring high coverage per perturbation (typically 100-1,000 cells). This cellular throughput is possible in our system because of the high fluorescence signal-to-noise achieved by

padlock-based barcode amplification and sequencing-by-synthesis. High signal enables accurate sequence data to be obtained at low optical magnification in large fields of view, each containing thousands of cells (Fig. 2B, Movie S1, Tables S2 and S3). Specifically, following optimization of the barcode amplification protocol including the post-fixation and padlock extension/ligation conditions (Fig. S2), *in situ* sequencing signals were readily visible at 10X optical magnification, with one or more exactly mapped reads detected in over 82% of cells transduced with lentiGuide-BC (Fig. 2, B and E).

Following the initial robust detection of 12-nt barcodes, we designed a set of 83,314 barcodes that were used in subsequent screens. As there are 16.7 million possible 12-nt sequences, we selected barcodes with minimum pairwise edit distance of 3, which allowed us to detect and reject up to two errors (insertion/deletion/substitution) arising from oligo synthesis or *in situ* processing (Fig. 2F, Methods) (26).

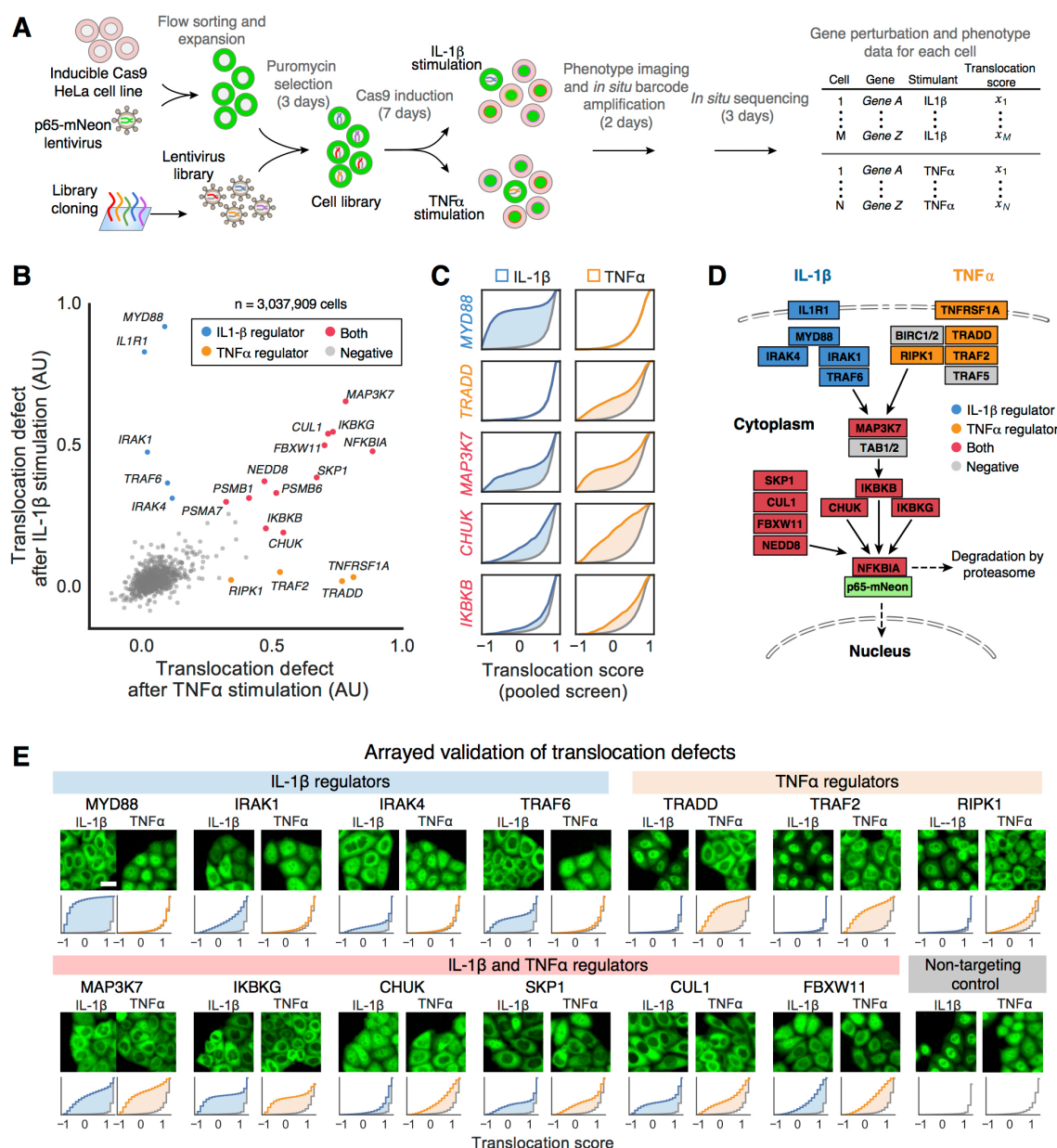


**Fig. 3: Accuracy of phenotype-to-genotype mapping assessed with a fluorescent reporter.** (A) Schematic of a frameshift reporter that converts CRISPR-Cas9-induced indel mutations into a positive fluorescent signal. (B) HeLa-TetR-Cas9 cells were transduced first with the frameshift reporter and subsequently with a library of 972 barcodes, each encoding a targeting or control sgRNA. (C) After Cas9 induction, the frameshift reporter was activated and the resulting nuclear-localized HA epitope tag may be stained with labeled antibody and detected in each cell, either by immunofluorescence or via FACS; corresponding barcode sequences were read out by *in situ* sequencing or NGS. (D) Targeting and control barcodes were well separated by fraction of HA<sup>+</sup> cells. (E) The same cell library was screened by flow sorting cells into HA<sup>+</sup> and HA<sup>-</sup> bins and performing next-generation sequencing of the genomically integrated barcode. (F) Comparison of barcode abundances measured by *in situ* sequencing or NGS ( $R^2 = 0.55$ ). The relative abundance of 95% of barcodes was within 5-fold (indicated by dashed lines).



We next sought to evaluate the overall accuracy of mapping perturbation genotype to cell phenotype *in situ*. In principle, errors could arise from oligo synthesis, library cloning, lentiviral delivery, barcode diffusion during *in situ* processing, barcode readout by *in situ* sequencing, or

incorrect assignment of reads to cells during image processing. To assess the impact of these error sources on CRISPR-Cas9 screening, we built a lentiviral reporter that produces an HA-tagged, nuclear-localized H2B protein after a Cas9-induced +1 frameshift in a target region (Fig. 3A,



**Fig. 4: A screen for regulators of NF-κB signaling.** (A) Workflow for CRISPR-Cas9 knockout-based screening using a fluorescently tagged reporter cell line. Screen hits were identified by the failure of p65-mNeonGreen to translocate following stimulation with IL-1β or TNFα cytokines. (B) Known NF-κB regulators were identified as high-ranking screen hits. Cells were assigned translocation scores based on the pixelwise correlation between mNeonGreen fluorescence and a DAPI nuclear stain. The translocation defect for a gene was defined based on the integrated difference in the distribution of translocation scores relative to non-targeting control sgRNAs across three replicate screens. (C) Cumulative distributions of translocation scores of known NF-κB regulators in response to both cytokines. The shaded areas depict the difference between the translocation score distributions for targeting sgRNAs and non-targeting control sgRNAs (gray). (D) NF-κB pathway map (KEGG HSA04064) color-coded as in (B). (E) Top-ranked genes were validated with individual CRISPR-Cas9 knockouts. Histograms show the cumulative distributions of IL-1β and TNFα-induced translocation scores for each gene knockout compared to wildtype cells (gray).



**Methods**). Cells expressing the reporter can either be screened *in situ* or by FACS. The reporter is highly specific and sensitive, with a mean *in situ* activation across 5 targeting sgRNAs of  $65 \pm 2.7\%$  and background of  $<0.001\%$  in the absence of a targeting sgRNA (**Fig. 3, B and C, Fig. S3**). We transduced cells stably expressing the frameshift reporter with a library containing 972 barcodes redundantly encoding 5 targeting and 5 control sgRNAs (average of 97 barcodes per sgRNA). We then induced Cas9 expression, measured reporter activation by immunofluorescence, and determined barcode sequences by *in situ* sequencing. All barcodes encoding targeting sgRNAs were distinguishable from control sgRNAs by HA<sup>+</sup> fraction. False positive events in which HA<sup>+</sup> cells were assigned control sgRNAs were used to calculate a per-cell misidentification rate of 9.4% (**Fig. 3, D and E, Methods**). We also screened the same cell library via FACS followed by deep sequencing and observed a similar enrichment of targeting sgRNAs (**Fig. 3E**), with similar representation of most barcodes in both contexts (95% within 5-fold abundance, **Fig. 3F**). We achieved comparably robust mapping of CRISPR sgRNAs to the frameshift reporter phenotype with the CROP-seq vector (4), in which the sgRNA sequence is duplicated in the 3' UTR of an antibiotic resistance transcript and may be directly sequenced *in situ* (**Fig. S4**).

Finally, we performed a large screen to identify genes required for activation of NF- $\kappa$ B, an extensively studied family of transcription factors that translocate to the nucleus in response to a host of stimuli (27, 28). We measured nuclear translocation of a p65-mNeonGreen reporter in HeLa-TetR-Cas9 cells following stimulation with either IL-1 $\beta$  or TNF $\alpha$ , cytokines that activate NF- $\kappa$ B via alternate pathways (**Fig. 4A**). We screened a library targeting 952 genes (3,063 sgRNAs) encompassing known NF- $\kappa$ B pathway-related components as well as all genes with GO annotations for ubiquitin ligase and deubiquitination activity, some of which are known to positively and negatively regulate NF- $\kappa$ B activation (29) (**Fig. 4A**). After stimulation with either IL-1 $\beta$  or TNF $\alpha$ , we scored the degree of p65-mNeonGreen translocation in each cell and ranked the perturbations by the deviation in their translocation score distribution from negative control sgRNAs to identify hit genes specific to IL-1 $\beta$  or TNF $\alpha$  as well as genes affecting the response to both cytokines (**Fig. 4, B and C, and Table S5, Methods**). Statistical significance was determined by permutation testing of non-targeting sgRNAs. We collected the full primary screen dataset from a single multi-well plate in which a total of 8,168,177 cells were imaged, with 3,037,909 cells retained for analysis after filtering cells based on reporter expression, nuclear morphology, and exact barcode mapping.

Our screen recovered known pathway components annotated by KEGG (30) for NF- $\kappa$ B activation by IL-1 $\beta$  signaling (5/5

genes), TNF $\alpha$  signaling (4/7 genes) and downstream components (5/7 genes), including the cytokine-specific receptors, adapter proteins, and factors that activate the shared regulator MAP3K7 (**Fig. 4D**) (28). Hits common to both cytokines included MAP3K7 itself and its target, the IKK complex (*CHUK*, *IKBKB*, *IKBKG*), as well as components of the SKP1-CUL1-F-box ubiquitin ligase complex and proteasome subunits, which together lead to degradation of the inhibitory I $\kappa$ B $\alpha$  protein and nuclear translocation of p65. Validation with arrayed CRISPR knockouts confirmed 17/17 top-ranked hits (**Fig. 4E and Fig. S5**), three other known pathway members (*IRAK4*, *RIPK1*, *BIRC2*), and potentially novel hits including *BAP1*, *HCFC1*, and *DCUN1D1*. Phenotype strength was well correlated between the primary screening and validation data (Spearman's  $\rho=0.87$  (IL-1 $\beta$ ),  $\rho=0.76$  (TNF $\alpha$ )), emphasizing the quantitative nature of the primary screen (**Fig. S6**). Within the group of IL-1 $\beta$ -specific screening hits, *BAP1* has been previously described to deubiquitinate *HCFC1* (31), with relevance for controlling metabolism (32), ER-stress signaling (33), cell-cycle progression (34), and viral gene expression (35). However, no involvement in NF- $\kappa$ B signaling has been described to our knowledge. It will be interesting to further elucidate the involvement of these genes in pro-inflammatory signaling.

## Discussion

Pooled optical screens are a novel method for systematic analysis of the genetic components underpinning a broad range of spatially and temporally defined phenotypes. The workflow closely mimics conventional pooled screening and requires no specialized hardware apart from a standard automated epifluorescence microscope. Optical readout of genetic perturbations is compatible with any perturbation that can be identified by a short expressed sequence, including Cas9-mediated gene knockout, repression, and activation, as well as libraries of ORFs, protein variants, and non-coding elements barcoded with expressed tags (36, 37).

Our approach is broadly applicable across many settings. For example, multiple perturbation barcodes can be read out within the same cell (**Fig. S7**), suggesting a straightforward route to studying genetic interactions with microscopy. The potential to integrate optical screening with high-dimensional morphological profiling and *in situ* multiplexed gene expression analysis (19, 38–41) raises the prospect of learning phenotypes from data rather than pre-specifying phenotypes of interest. Existing protocols for *in situ* sequencing in tissue samples (19, 41) highlight the exciting possibility of perturbing cells *in vivo* and measuring the resulting phenotypes within the native spatial context.

## Methods

### Tissue culture

HEK293FT cells (ATCC CRL-1573) were cultured in DMEM with sodium pyruvate and GlutaMAX (Life 10569044) supplemented with heat-inactivated fetal bovine serum (Seradigm 97068-085) and 100 U/mL penicillin-streptomycin (Life Technologies 15140163). All HeLa cell lines were cultured in the same media with serum substituted for 10% tetracycline-screened fetal bovine serum (Hyclone SH30070.03T).

Parental HeLa-TetR-Cas9 cells were a gift from Iain Cheeseman. In order to select an optimal clone for further experiments, single cells were sorted into a 96-well plate (Sony SH800), clonally expanded, and screened for Cas9 activity after 8 days with and without 1 µg/mL doxycycline induction. Cas9 activity was assessed by transducing each clone with pXPR\_011 (Addgene #59702), a reporter vector expressing GFP and an sgRNA targeting GFP, and using FACS to read out efficiency of protein knockdown. Additionally, gene editing was directly assessed by transduced HeLa-TetR-Cas9 clones with a guide targeting TFRC. Genomic DNA was extracted from both uninduced and induced clones by resuspending in cell lysis buffer (10 mM Tris pH 7.5, 1 mM CaCl<sub>2</sub>, 2 mM MgCl<sub>2</sub>, 1 mM EDTA, and 0.2 mg/mL Proteinase K), and heating for 10 minutes at 65°C and 15 minutes at 95°C. The guide target region was amplified by PCR and sequenced on an Illumina MiniSeq. The best clones showed efficient indel generation ( $\geq 97\%$ ) in the presence of doxycycline and minimal cutting ( $\leq 2\%$ ) in its absence.

In preparation for *in situ* analysis, cells were seeded onto glass-bottom plates (6-well: Cellvis P06-1.5H-N, 24-well: Greiner Bio-one 662892, 96-well: Greiner Bio-one 665892) at a density of 50,000 cells/cm<sup>2</sup> and incubated for 2 days to permit proper cell attachment, spreading, and colony formation.

### Lentiviral production

HEK293FT cells were seeded into 15-cm plates or multi-well plates at a density of 100,000 cells/cm<sup>2</sup>. After 20 hours, cells were transfected with pMD2.G (Addgene #12259), psPAX2 (Addgene #12260), and a lentiviral transfer plasmid (2:3:4 ratio by mass) using Lipofectamine 3000 (Thermo Fisher L3000015). Media was exchanged after 4 hours and supplemented with 2 mM caffeine 20 hours post-transfection to increase viral titer. Viral supernatant was harvested 48 hours after transfection and filtered through 0.45 µm PVDF filters (Millipore SLHVR04NL).

### Lentiviral transduction

HeLa-TetR-Cas9 cells were transduced by adding viral supernatant supplemented with polybrene (8 µg/mL) and centrifuging at 1000g for 2 hours at 33°C. At 5 hours post-infection, media was exchanged. At 24 hours post-infection, cells were passaged into media containing selection antibiotic at the following concentrations: 1 µg/mL puromycin (ThermoFisher A1113802), 300 µg/mL hygromycin (Invivogen ant-hg-1), and 300 µg/mL zeocin (ThermoFisher R25001).

For lentiviral transduction of lentiGuide-BC libraries, a carrier plasmid was utilized to minimize recombination between distant genetic elements (e.g., sgRNA and associated barcode). Libraries were packaged following the above protocol, with the library transfer plasmid diluted in integration-deficient carrier vector pR\_LG (1:10 mass ratio of library to carrier, Addgene #112895) (25).

### Library cell line validation

For library transductions, multiplicity of infection was estimated by counting colonies after sparse plating and antibiotic selection. Genomic DNA was also extracted for NGS validation of library representation.

### Next generation sequencing of libraries

Genomic DNA was extracted using an extraction mix as described above. Barcodes and sgRNAs were amplified by PCR from a minimum of 1e6 genomic equivalents per library using NEBNext 2X Master Mix (initial denaturation 5 minutes at 98°C, followed by 28 cycles of annealing for 10 seconds at 65°C, extension for 25 seconds at 72°C, and denaturation for 20 seconds at 98°C).

### Library design and cloning

A set of 12-nt barcodes was designed by selecting 83,314 barcodes from the set of 16.7 million possible 12-nt sequences by filtering for GC content between 25% and 75%, no more than 4 consecutive repeated bases, and minimum substitution and insertion/deletion edit distance (Levenshtein distance) of 3 between any pair of barcodes. Ensuring a minimum edit distance is useful for detecting and correcting errors, which arise mainly from DNA synthesis and *in situ* reads with low signal-to-background ratios. The E-CRISP web tool was used to select sgRNA sequences targeting genes of interest. Barcode-sgRNA pairs were randomly assigned and co-synthesized on a 125-nt 90K oligo array (CustomArray) (table S4). Individual libraries were amplified from the oligo pool by dial-out PCR (42) and cloned into lentiGuide-BC or lentiGuide-BC-CMV (the latter contains the CMV promoter instead of the EF1a promoter) via two steps of Golden Gate assembly using BsmBI and BbsI restriction sites. Libraries

were transformed in electrocompetent cells (Lucigen Endura) and grown in liquid culture for 18 hours at 30°C before extracting plasmid DNA. The sgRNA-barcode association was validated by Sanger sequencing individual colonies from the final library.

## Padlock-based RNA detection

Preparation of targeted RNA amplicons for *in situ* sequencing was adapted from published protocols with modifications to improve molecular detection efficiency and amplification yield (19, 43). Cells were fixed with 4% paraformaldehyde (Electron Microscopy Sciences 15714) for 30 minutes, washed with PBS, and permeabilized with 70% ethanol for 30 minutes. Permeabilization solution was carefully exchanged with PBS-T wash buffer (PBS + 0.05% Tween-20) to minimize sample dehydration. Reverse transcription mix (1x RevertAid RT buffer, 250  $\mu$ M dNTPs, 0.2 mg/mL BSA, 1  $\mu$ M RT primer, 0.8 U/ $\mu$ L Ribolock RNase inhibitor, and 4.8 U/ $\mu$ L RevertAid H minus reverse transcriptase) was added to the sample and incubated for 16 hours at 37°C. Following reverse transcription, cells were washed 5 times with PBS-T and post-fixed with 3% paraformaldehyde and 0.1% glutaraldehyde for 30 minutes at room temperature, then washed with PBS-T 5 times. After this step, cells expressing p65-mNeonGreen were imaged. Samples were thoroughly washed again with PBS-T, incubated in a padlock probe and extension-ligation reaction mix (1x Ampligase buffer, 0.4 U/ $\mu$ L RNase H, 0.2 mg/mL BSA, 10 nM padlock probe, 0.002 U/ $\mu$ L TaqIT polymerase, 0.5 U/ $\mu$ L Ampligase and 50 nM dNTPs) for 5 minutes at 37°C and 90 minutes at 45°C, and then washed 2 times with PBS-T. Circularized padlocks were amplified with rolling circle amplification mix (1x Phi29 buffer, 250  $\mu$ M dNTPs, 0.2 mg/mL BSA, 5% glycerol, and 1 U/ $\mu$ L Phi29 DNA polymerase) at 30°C for either 3 hours or overnight.

## *In situ* sequencing

Rolling circle amplicons were prepared for sequencing by hybridizing a mix containing sequencing primer pSBS\_lentiGuide-BC or pSBS\_CROPseq (1  $\mu$ M primer in 2X SSC + 10% formamide) for 30 minutes at room temperature. Barcodes were read out using sequencing-by-synthesis reagents from the Illumina MiSeq 500 cycle Nano kit (Illumina MS-103-1003). First, samples were washed with incorporation buffer (Nano kit PR2) and incubated for 3 minutes in incorporation mix (Nano kit reagent 1) at 60°C. Samples were then thoroughly washed with PR2 at 60°C (6 washes for 3 minutes each) and placed in 200 ng/mL DAPI in 2x SSC for fluorescence imaging. Following each imaging cycle, dye terminators were removed by incubation for 6 minutes in Illumina cleavage mix (Nano kit reagent 4) at 60°C, and samples were thoroughly washed with PR2.

## Fluorescence microscopy

All images were acquired using a Ti-E Eclipse inverted epifluorescence microscope (Nikon) with automated XYZ stage control and hardware autofocus. An LED light engine (Lumencor Sola SE FISH II) was used for fluorescence illumination and all hardware was controlled using Micromanager software (44). *In situ* sequencing cycles were imaged using a 10X 0.45 NA CFI Plan Apo Lambda objective (Nikon) with the following filters (Semrock) and exposure times for each base: G (excitation 534/20 nm, emission 572/28 nm, dichroic 552 nm, 200 ms); T (excitation 575/25 nm, emission 6158/24 nm, dichroic 596 nm, 200 ms); A (excitation 635/18 nm, emission 680/42 nm, dichroic 652 nm, 200 ms); C (excitation 661/20 nm, emission 732/68 nm, dichroic 695 nm, 800 ms).

## Image analysis

Images of cell phenotype and *in situ* sequencing of perturbations were aligned using cross-correlation of DAPI-stained nuclei. Nuclei were detected using local thresholding and watershed-based segmentation. Cells were typically segmented using local thresholding of cytoplasmic background and assignment of pixels to the nearest nucleus by the fast-marching method. Frameshift reporter and NF- $\kappa$ B translocation phenotypes were quantified by calculating pixel-wise correlations between the nuclear DAPI channel and 488 channel (HA stain or p65-mNeonGreen, respectively).

Sequencing reads were detected by applying a Laplacian-of-Gaussian linear filter (kernel width  $\sigma = 1$  pixel), calculating the per-pixel standard deviation over sequencing cycles, averaging over color channels, and finding local maxima. The base intensity at each cycle was defined as the maximum value in a 3x3 pixel window centered on the read. A linear transformation was then applied to correct for optical cross-talk and intensity differences between color channels. Finally, each base was called according to the channel with maximum corrected intensity, and a per-base quality score was defined as the ratio of intensity for the maximum channel to total intensity for all channels. The output of the sequencing image analysis was a FASTQ file recording each sequencing read along with the identity of the overlapping cell, quality score per base, and spatial location.

Additional details and source code available on Github at <https://github.com/blaineylab/OpticalPooledScreens>.

## Frameshift reporter screen

HeLa-TetR-Cas9 cells were stably transduced at MOI >2 with pL\_FR\_Hygro and selected with hygromycin for 7 days to generate the HeLa-TetR-Cas9-FR cell line. Cells transduced



with the pL\_FR\_Hygro lentiviral vector express an open reading frame consisting of a 50-nt frameshift reporter target sequence, followed by an H2B histone coding sequence with C-terminus HA epitope tag (+1 frameshift), followed by a second H2B sequence with C-terminus myc tag (+0 frameshift) and hygromycin resistance cassette (+0 frameshift). The H2B-HA, H2B-myc, and hygromycin resistance sequences are preceded by self-cleaving 2A peptides in the same reading frame. Before generation of Cas9-mediated indel mutations, cells express the coding sequences with +0 frameshift. Subsequent activation of the reporter by co-expression of Cas9 and a targeting sgRNA leads to mutations in the target sequence, which may alter the downstream reading frame. A frameshift of +1 leads to expression of the H2B-HA protein, which can be visualized by immunofluorescence and detected by microscopy or flow cytometry. Integration of multiple copies of reporter per cell increases the likelihood of generating a +1 frameshift in at least one copy.

HeLa-TetR-Cas9-FR cells were used to screen targeting and control sgRNAs. A barcoded sgRNA perturbation library with 972 barcodes, each encoding one of 5 targeting or 5 control sgRNAs, was synthesized and cloned into lentiGuide-BC-CMV. This library was transduced into HeLa-TetR-Cas9-FR cells at MOI < 0.05 in three replicates, which were independently cultured and screened. Following 4 days of puromycin selection, cells were collected to validate library representation by NGS. Cas9 expression was induced by supplementing the culture media with 1  $\mu$ g/mL doxycycline for 6 days. Cells were then split for screening either via *in situ* sequencing or by FACS.

For *in situ* screening, 500,000 cells were seeded into each well of a glass-bottom 6-well plate (CellVis). After two days of culture, *in situ* padlock detection and sequencing were carried out as above, with the modification that prior to sequencing-by-synthesis, cells were immunostained to detect frameshift reporter activation by blocking and permeabilizing with 3% BSA + 0.5 % Triton X-100 for 5 minutes, incubating in rabbit anti-HA (1:1000 dilution in 3% BSA) for 30 minutes, washing with PBS-T and incubating with goat anti-rabbit F(ab')<sub>2</sub> fragment Alexa 488 (CST 4412S, 1:1000 dilution in 3% BSA) for 30 minutes. Samples were changed into imaging buffer (200 ng/mL DAPI in 2X SSC), and phenotype images were acquired.

FACS screening was carried out by fixing cells with 4% PFA, permeabilizing with 70% ice-cold ethanol, and immunostaining with the same anti-HA primary and secondary antibodies and dilutions used for *in situ* analysis. Cells were sorted into HA<sup>+</sup> and HA<sup>-</sup> populations (Sony SH800) and genomically integrated perturbations were sequenced as described above. The enrichment for each

barcoded perturbation was defined as the ratio of normalized read counts.

To estimate the rate at which cells are assigned an incorrect perturbation barcode, we first assumed all HA<sup>+</sup> cells mapped to a non-targeting control sgRNA (4.7%) were false positive events due to incorrect barcode assignment (supported by the very low false positive rate (<0.001%) of the frameshift reporter itself, measured for a single perturbation in arrayed format). However, as incorrect barcode assignments were equally likely to map an HA<sup>+</sup> cell to a targeting or control sgRNA, we estimated the misidentification rate to be twice as large, or 9.4%.

## NF- $\kappa$ B pooled screen

HeLa-TetR-Cas9 cells were transduced with pR14\_p65-mNeonGreen, a C-terminal fusion of p65 with a bright monomeric green fluorescent protein (Allele Biotechnology). Fluorescent cells were sorted by FACS (Sony SH800) and re-sorted to select for cells with stable expression. This reporter cell line was further transduced with a 4,063-barcode sgRNA library (962 genes targeted (952 detected), 866 barcodes assigned to non-targeting controls) in lentiGuide-BC-CMV. Cells were selected with puromycin for 4 days and library representation was validated by NGS.

Cas9 expression was induced with 1  $\mu$ g/mL doxycycline and cells were seeded onto 6-well glass-bottom plates 2 days prior to translocation experiments. The total time between Cas9 induction and performing the NF- $\kappa$ B activation assay was 7 days. Cells were stimulated with either 30 ng/mL TNF $\alpha$  or 30 ng/mL IL-1 $\beta$  for 40 minutes prior to fixation with 4% paraformaldehyde for 30 minutes and initiation of the *in situ* sequencing protocol. Translocation phenotypes were recorded after the post-fixation step by exchanging for imaging buffer (2X SSC + 200 ng/mL DAPI) and imaging the nuclear DAPI stain and p65-mNeonGreen. After phenotyping, the remainder of the *in situ* sequencing protocol (gap-fill and rolling circle amplification) and 12 bases of sequencing-by-synthesis were completed.

Nuclei of individual cells were segmented by thresholding background-subtracted DAPI signal and separating the resulting regions using the watershed method. Cells with at least one read exactly matching a library barcode were retained for analysis. In order to remove mitotic cells and cells with abnormally high or low reporter expression, cells were further filtered based on nuclear area, maximum DAPI signal, and mean p65-mNeonGreen signal. Pixel-wise DAPI-mNeonGreen correlation within the segmented nuclear region, described above, was used to define the translocation score for each cell as it most clearly separated perturbations against known NF- $\kappa$ B genes from non-targeting controls. The

phenotypic effects of perturbations targeting known NF- $\kappa$ B genes ranged from a large increase in fully untranslocated cells (e.g., *MAP3K7*) to more subtle negative shifts in the distribution of scores (e.g., *IKKB*).

To capture a broad range of effect size, we calculated an sgRNA translocation defect in a given replicate by computing the difference in translocation score distribution compared to non-targeting controls (the shaded area in **Fig. 4, C and E**). We found this metric performed better at separating known genes from controls than the often-used Kolmogorov-Smirnov distance. We defined the gene translocation defect as the second-largest sgRNA translocation defect for sgRNAs targeting that gene. This statistic helps reduce the false positive rate due to clonal effects (integration of an sgRNA into a cell that is defective in translocation) which are independent among sgRNAs and screen replicates, as well as false negatives due to inefficient sgRNAs.

A permutation test was used to calculate p-values for the gene translocation defects. Random subsets of sgRNA translocation defects were sampled from non-targeting controls to build a null distribution (3 sgRNAs per replicate, repeated 100,000 times). The cumulative null distribution was used to determine p-values for the gene translocation defects. Hits at an estimated FDR <10% and <20% were identified using the Benjamini-Hochberg procedure (**Table S5**). KEGG-annotated genes were defined as members of KEGG pathway HSA04064 (NF-kappa B signaling pathway) between IL-1 $\beta$  or TNF $\alpha$  and p65/p60.

## Acknowledgments

We thank Rhiannon Macrae, Iain Cheeseman, Aviv Regev, Fei Chen, Arjun Raj, Nir Hacohen, Sara Jones, Emily Botelho, and members of the Blainey and Zhang labs for critical feedback and discussions. We thank Leslie Gaffney for assistance with figures. We gratefully acknowledge Linyi Gao for the gift of the integration-deficient copackaging plasmid pR<sub>LG</sub>.

## Funding

This work was supported by a Simons Center Seed Grant from MIT, the Broad Institute through startup funding (to P.C.B.) and the BN10 program, and two grants from the National Human Genome Research Institute (HG009283 and HG006193). P.C.B. is supported by a Career Award at the Scientific Interface from the Burroughs Wellcome Fund. F.Z. is a New York Stem Cell Foundation–Robertson Investigator. F.Z. is supported by NIH grants (1R01- HG009761, 1R01-MH110049, and 1DP1-HL141201); the Howard Hughes Medical Institute; the New York Stem Cell, Simons, Paul G. Allen Family, and Vallee Foundations; the Poitras Center for Affective Disorders Research at MIT; the Hock E. Tan and K.

Lisa Yang Center for Autism Research at MIT; J. and P. Poitras; and R. Metcalfe. A.M. is supported by the Swedish Research Council (grant 2015-06403). R.J.C. is supported by a Fannie and John Hertz Foundation Fellowship and an NSF Graduate Research Fellowship.

## Author contributions

D.F. designed the approach with input from all authors. D.F., A.S., A.M., A.J.G., R.J.C. and J.S.B performed experiments. D.F. analyzed data. J.S.B and D.F. designed the NF- $\kappa$ B screen. P.C.B. and F.Z. supervised the research. D.F., A.S., and P.C.B. wrote the manuscript with contributions from all authors.

## Competing interests

The Broad Institute and MIT may seek to commercialize aspects of this work, and related applications for intellectual property have been filed.

## Online software and protocols

Python scripts, including a complete pipeline from example data to sequencing reads, will be made available on Github, along with plasmid sequences and updated protocols at <https://github.com/blaineylab/OpticalPooledScreens>

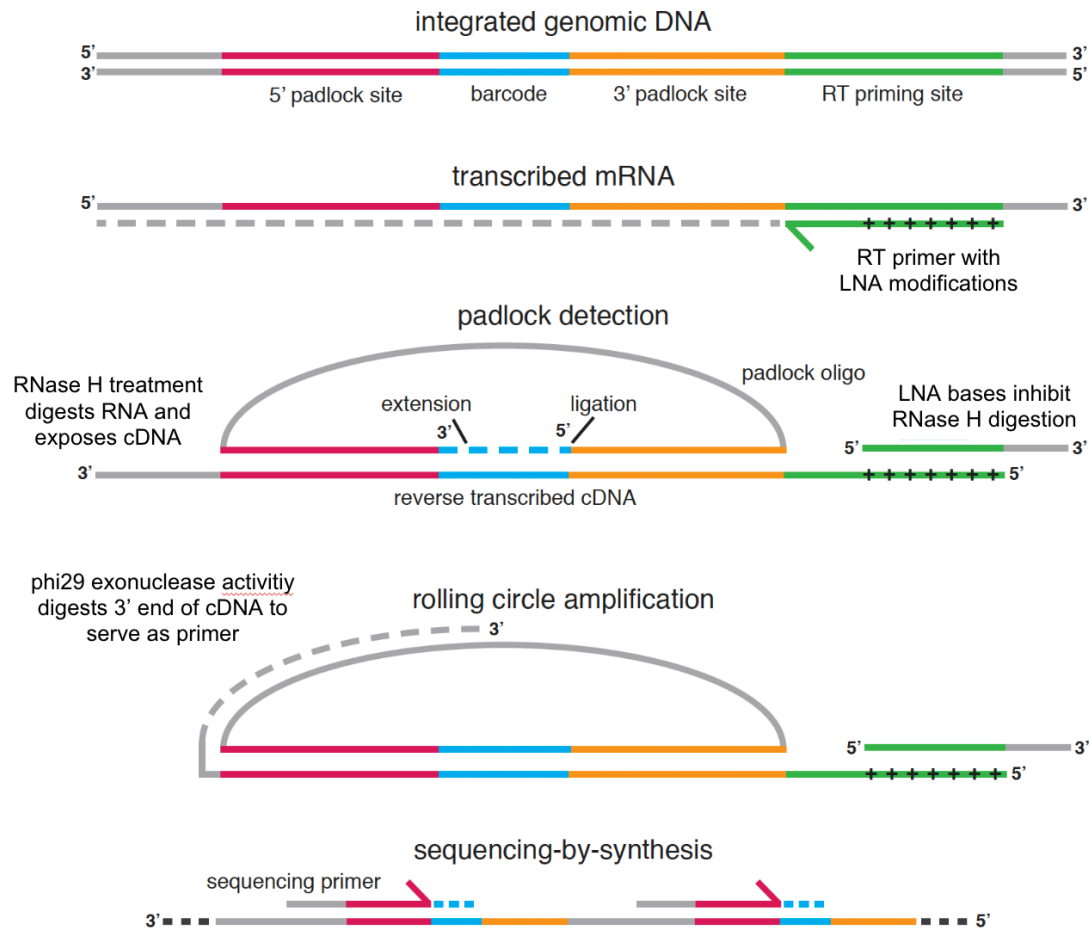
## References

1. Wang, T., J.J. Wei, D.M. Sabatini, and E.S. Lander. 2014. Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science* (80-. ). 343: 80–84.
2. Shalem, O., N.E. Sanjana, E. Hartenian, X. Shi, D.A. Scott, T.S. Mikkelsen, D. Heckl, B.L. Ebert, D.E. Root, J.G. Doench, and F. Zhang. 2014. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* (80-. ). 343: 84–87.
3. Parnas, O., M. Jovanovic, T.M. Eisenhaure, R.H. Herbst, A. Dixit, C.J. Ye, D. Przybylski, R.J. Platt, I. Tirosh, N.E. Sanjana, O. Shalem, R. Satija, R. Raychowdhury, P. Mertins, S.A. Carr, F. Zhang, N. Hacohen, and A. Regev. 2015. A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell*. 162: 675–686.
4. Datlinger, P., A.F. Rendeiro, C. Schmidl, T. Krausgruber, P. Traxler, J. Klughammer, L.C. Schuster, A. Kuchler, D. Alpar, and C. Bock. 2017. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods*. 14: 297–301.
5. Jaitin, D.A., A. Weiner, I. Yofe, D. Lara-Astiaso, H. Keren-Shaul, E. David, T.M. Salame, A. Tanay, A. van Oudenaarden, and I. Amit. 2016. Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell*. 167: 1883–1896.e15.
6. Dixit, A., O. Parnas, B. Li, J. Chen, C.P. Fulco, L. Jerby-Arnon, N.D. Marjanovic, D. Dionne, T. Burks, R. Raychowdhury, B. Adamson, T.M. Norman, E.S. Lander, J.S. Weissman, N. Friedman, and A. Regev. 2016. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell*. 167: 1853–1866.e17.
7. Adamson, B., T.M. Norman, M. Jost, M.Y. Cho, J.K. Nuñez, Y. Chen, J.E. Villalta, L.A. Gilbert, M.A. Horlbeck, M.Y. Hein, R.A. Pak, A.N. Gray, C.A. Gross, A. Dixit, O. Parnas, A. Regev, and J.S. Weissman. 2016. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell*. 167: 1867–1882.e21.
8. Moffat, J., D.A. Grueneberg, X. Yang, S.Y. Kim, A.M. Kloepper, G. Hinkle, B. Piqani, T.M. Eisenhaure, B. Luo, J.K. Grenier, A.E. Carpenter, S.Y. Foo, S.A. Stewart, B.R. Stockwell, N. Hacohen, W.C. Hahn, E.S. Lander, D.M. Sabatini, and D.E. Root. 2006. A Lentiviral RNAi Library for Human and Mouse Genes Applied to an Arrayed Viral High-Content Screen. *Cell*. 124: 1283–1298.
9. Neumann, B., T. Walter, J.-K. Hériché, J. Bulkescher, H. Erfle, C. Conrad, P. Rogers, I. Poser, M. Held, U. Liebel, C. Cetin, F. Sieckmann, G. Pau, R. Kabbe, A. Wünsche, V. Satagopam, M.H.A. Schmitz, C. Chapuis, D.W. Gerlich, R. Schneider, R. Eils, W. Huber, J.-M. Peters, A.A. Hyman, R. Durbin, R. Pepperkok, and J. Ellenberg. 2010. Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature*. 464: 721–727.
10. Collinet, C., M. Stöter, C.R. Bradshaw, N. Samusik, J.C. Rink, D. Kenski, B. Habermann, F. Buchholz, R. Henschel, M.S. Mueller, W.E. Nagel, E. Fava, Y. Kalaidzidis, and M. Zerial. 2010. Systems survey of endocytosis by multiparametric image analysis. *Nature*. 464: 243–249.
11. Karlas, A., N. Machuy, Y. Shin, K.-P. Pleissner, A. Artarini, D. Heuer, D. Becker, H. Khalil, L.A. Ogilvie, S. Hess, A.P. Mäurer, E. Müller, T. Wolff, T. Rudel, and T.F. Meyer. 2010. Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication. *Nature*. 463: 818–822.
12. Chia, N.-Y., Y.-S. Chan, B. Feng, X. Lu, Y.L. Orlov, D. Moreau, P. Kumar, L. Yang, J. Jiang, M.-S. Lau, M. Huss, B.-S. Soh, P. Kraus, P. Li, T. Lufkin, B. Lim, N.D. Clarke, F. Bard, and H.-H. Ng. 2010. A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature*. 468: 316–320.
13. Guo, Y., T.C. Walther, M. Rao, N. Stuurman, G. Goshima, K. Terayama, J.S. Wong, R.D. Vale, P. Walter, and R. V. Farese. 2008. Functional genomic screen reveals genes involved in lipid-droplet formation and utilization. *Nature*. 453: 657–661.
14. Floyd, S.R., M.E. Pacold, Q. Huang, S.M. Clarke, F.C. Lam, I.G. Cannell, B.D. Bryson, J. Rameseder, M.J. Lee, E.J. Blake, A. Fydrych, R. Ho, B.A. Greenberger, G.C. Chen, A. Maffa, A.M. Del Rosario, D.E. Root, A.E. Carpenter, W.C. Hahn, D.M. Sabatini, C.C. Chen, F.M. White, J.E. Bradner, and M.B. Yaffe. 2013. The bromodomain protein Brd4 insulates chromatin from DNA damage signalling. *Nature*. 498: 246–250.
15. Orvedahl, A., R.S. Jr, G. Xiao, A. Ng, Z. Zou, Y. Tang, M. Narimatsu, C. Gilpin, Q. Sun, M. Roth, C. V. Forst, J.L. Wrana, Y.E. Zhang, K. Luby-Phelps, R.J. Xavier, Y. Xie, and B. Levine. 2011. Image-based genome-wide siRNA screen identifies selective autophagy factors. *Nature*. 480: 113–117.



16. Linhoff, M.W., J. Laurén, R.M. Cassidy, F.A. Dobie, H. Takahashi, H.B. Nygaard, M.S. Airaksinen, S.M. Strittmatter, and A.M. Craig. 2009. An Unbiased Expression Screen for Synaptogenic Proteins Identifies the LRRTM Protein Family as Synaptic Organizers. *Neuron*. 61: 734–749.
17. Lawson, M.J., D. Camsund, J. Larsson, Ö. Baltekin, D. Fange, and J. Elf. 2017. In situ genotyping of a pooled strain library after characterizing complex phenotypes. *Mol. Syst. Biol.* 13: 947.
18. Emanuel, G., J.R. Moffitt, and X. Zhuang. 2017. High-throughput, image-based screening of pooled genetic-variant libraries. *Nat. Methods*. 14: 1159–1162.
19. Ke, R., M. Mignardi, A. Pacureanu, J. Svedlund, J. Botling, C. Wählby, and M. Nilsson. 2013. In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods*. 10: 857–860.
20. Sanjana, N.E., O. Shalem, and F. Zhang. 2014. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods*. 11: 783–784.
21. Sack, L.M., T. Davoli, Q. Xu, M.Z. Li, and S.J. Elledge. 2016. Sources of Error in Mammalian Genetic Screens. *G3 (Bethesda)*. 6: 2781–90.
22. Hill, A.J., J.L. McFaline-Figueroa, L.M. Starita, M.J. Gasperini, K.A. Matreyek, J. Packer, D. Jackson, J. Shendure, and C. Trapnell. 2018. On the design of CRISPR-based single-cell molecular screens. *Nat. Methods*. 15: 271–274.
23. Xie, S., A. Cooley, D. Armendariz, P. Zhou, and G.C. Hon. 2018. Frequent sgRNA-barcode recombination in single-cell perturbation assays. 13.
24. Adamson, B., T.M. Norman, M. Jost, and J.S. Weissman. 2018. Approaches to maximize sgRNA-barcode coupling in Perturb-seq screens. *bioRxiv*. : 298349.
25. Feldman, D., A. Singh, A.J. Garrity, and P.C. Blainey. 2018. Lentiviral co-packaging mitigates the effects of intermolecular recombination and multiple integrations in pooled genetic screens. *bioRxiv*. : 262121.
26. Buschmann, T., and L. V. Bystrykh. 2013. Levenshtein error-correcting barcodes for multiplexed DNA sequencing. *BMC Bioinformatics*. 14: 272.
27. Pahl, H.L. 1999. Activators and target genes of Rel/NF- $\kappa$ B transcription factors. *Oncogene*. 18: 6853–6866.
28. Gewurz, B.E., F. Towfic, J.C. Mar, N.P. Shinnars, K. Takasaki, B. Zhao, E.D. Cahir-McFarland, J. Quackenbush, R.J. Xavier, and E. Kieff. 2012. Genome-wide siRNA screen for mediators of NF- $\kappa$ B activation. *Proc. Natl. Acad. Sci.* 109: 2467–2472.
29. Chen, Z.J. 2005. Ubiquitin signalling in the NF-kappaB pathway. *Nat. Cell Biol.* 7: 758–65.
30. Kanehisa, M., and S. Goto. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28: 27–30.
31. Machida, Y.J., Y. Machida, A.A. Vashisht, J.A. Wohlschlegel, and A. Dutta. 2009. The Deubiquitinating Enzyme BAP1 Regulates Cell Growth via Interaction with HCF-1. *J. Biol. Chem.* 284: 34179–34188.
32. Bononi, A., H. Yang, C. Giorgi, S. Patergnani, L. Pellegrini, M. Su, G. Xie, V. Signorato, S. Pastorino, P. Morris, G. Sakamoto, S. Kuchay, G. Gaudino, H.I. Pass, A. Napolitano, P. Pinton, W. Jia, and M. Carbone. 2017. Germline BAP1 mutations induce a Warburg effect. *Cell Death Differ.* 24: 1694–1704.
33. Dai, F., H. Lee, Y. Zhang, L. Zhuang, H. Yao, Y. Xi, Z.-D. Xiao, M.J. You, W. Li, X. Su, and B. Gan. 2017. BAP1 inhibits the ER stress gene regulatory network and modulates metabolic stress response. *Proc. Natl. Acad. Sci.* 114: 3192–3197.
34. Misaghi, S., S. Ottosen, A. Izrael-Tomasevic, D. Arnott, M. Lamkanfi, J. Lee, J. Liu, K. O’Rourke, V.M. Dixit, and A.C. Wilson. 2009. Association of C-Terminal Ubiquitin Hydrolase BRCA1-Associated Protein 1 with Cell Cycle Regulator Host Cell Factor 1. *Mol. Cell. Biol.* 29: 2181–2192.
35. Johnson, K.M., S.S. Mahajan, and A.C. Wilson. 1999. Herpes simplex virus transactivator VP16 discriminates between HCF-1 and a novel family member, HCF-2. *J. Virol.* 73: 3930–40.
36. Yang, X.X., J.S. Boehm, X.X. Yang, K. Salehi-Ashtiani, T. Hao, Y. Shen, R. Lubonja, S.R. Thomas, O. Alkan, T. Bhimdi, T.M. Green, C.M. Johannessen, S.J. Silver, C. Nguyen, R.R. Murray, H. Hieronymus, D. Balcha, C. Fan, C. Lin, L. Ghamsari, M. Vidal, W.C. Hahn, D.E. Hill, and D.E. Root. 2011. A public genome-scale lentiviral expression library of human ORFs. *Nat. Methods*. 8: 659–661.
37. Melnikov, A., A. Murugan, X. Zhang, T. Tesileanu, L. Wang, P. Rogov, S. Feizi, A. Gnirke, C.G. Callan, J.B. Kinney, M. Kellis, E.S. Lander, and T.S. Mikkelsen. 2012. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* 30: 271–277.
38. Lee, J.H., E.R. Daugharthy, J. Scheiman, R. Kalhor, J.L. Yang, T.C. Ferrante, R. Terry, S.S.F. Jeanty, C. Li, R. Amamoto, D.T. Peters, B.M. Turczyk, A.H. Marblestone, S.A. Inverso, A. Bernard, P. Mali, X. Rios, J. Aach, and G.M. Church. 2014. Highly Multiplexed Subcellular RNA Sequencing In Situ. *Science (80-. )*. 343: 1360–

- 1363.
39. Lubeck, E., A.F. Coskun, T. Zhiyentayev, M. Ahmad, and L. Cai. 2014. Single-cell in situ RNA profiling by sequential hybridization. *Nat. Methods*. 11: 360–361.
40. Chen, K.H., A.N. Boettiger, J.R. Moffitt, S. Wang, and X. Zhuang. 2015. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* (80-. ). 348: aaa6090-aaa6090.
41. Wang, X., W.E. Allen, M.A. Wright, E.L. Sylwestrak, N. Samusik, S. Vesuna, K. Evans, C. Liu, C. Ramakrishnan, J. Liu, G.P. Nolan, F.-A. Bava, and K. Deisseroth. 2018. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science*. : eaat5691.
42. Schwartz, J.J., C. Lee, and J. Shendure. 2012. Accurate gene synthesis with tag-directed retrieval of sequence-verified DNA molecules. *Nat. Methods*. 9: 913–915.
43. Larsson, C., I. Grundberg, O. Söderberg, and M. Nilsson. 2010. In situ detection and genotyping of individual mRNA molecules. *Nat. Methods*. 7: 395–397.
44. Edelstein, A., N. Amodaj, K. Hoover, R. Vale, and N. Stuurman. 2010. Computer Control of Microscopes Using µManager. In: *Current Protocols in Molecular Biology*. Hoboken, NJ, USA: John Wiley & Sons, Inc. p. 14.20.1-14.20.17.
45. Chen, X., Y.-C. Sun, G.M. Church, J.H. Lee, and A.M. Zador. 2018. Efficient in situ barcode sequencing using padlock probe-based BaristaSeq. *Nucleic Acids Res*. 46: e22–e22.



**Fig. S1: Workflow for padlock detection and sequencing-by-synthesis.**

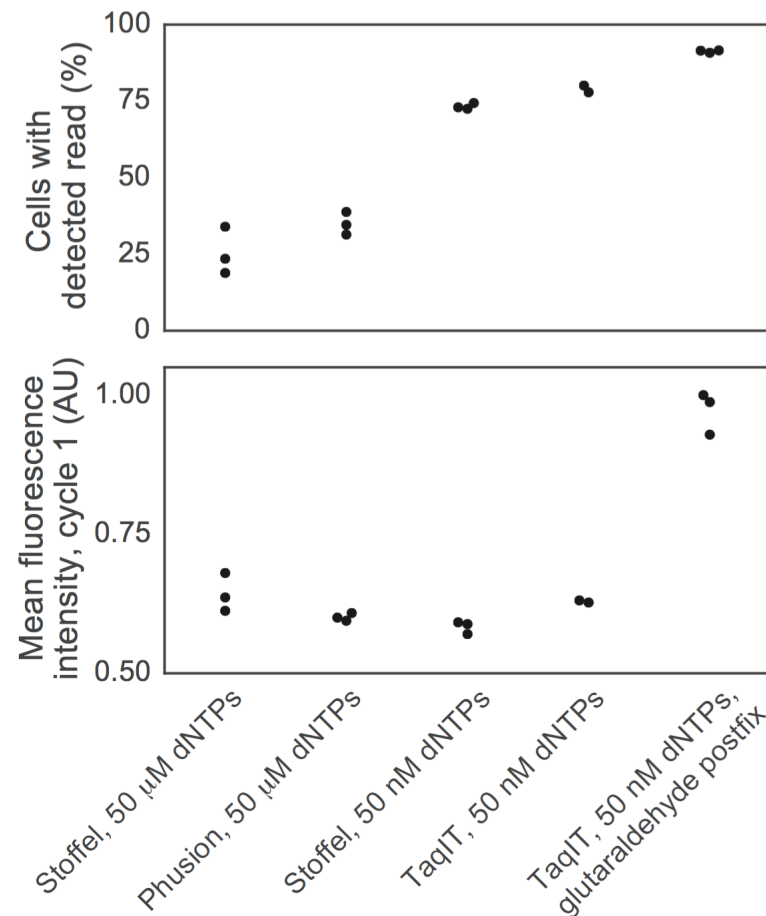
In order to determine the identity of the lentiviral vector integrated in each cell, all cellular RNAs are first fixed in place by formaldehyde treatment. A reverse transcription primer containing locked nucleic acid (LNA) bases is hybridized to the mRNA containing the barcode sequence. Complementary DNA (cDNA) is generated using a reverse transcriptase lacking RNase activity, producing an RNA-DNA hybrid. The cells are then fixed once again ("post-fixed") with a mixture of formaldehyde and glutaraldehyde to improve cDNA retention.

A single reaction mix containing RNase H, a DNA polymerase lacking strand displacement activity, a DNA ligase, and a padlock DNA oligonucleotide is then added. Digestion of the RNA strand exposes the cDNA bases, allowing the padlock to hybridize to the cDNA at sites flanking the barcode. The DNA polymerase extends the padlock, copying the barcode sequence, but does not strand displace the 5' annealed padlock arm. Once extended, the padlock is then ligated into a single-stranded DNA circle. During this step, the cDNA is retained in place via hybridization to the RNA strand at the LNA-modified bases within the RT primer, which inhibit RNase H digestion.

Phi29 polymerase is used to perform rolling circle amplification of the circularized padlock. The 3' exonuclease activity of Phi29 polymerase digests the single-stranded portion of the cDNA strand, generating a primer for rolling circle amplification. The amplified single-stranded DNA product contains tandem repeats of the padlock adaptor sequences and barcode, which can be read out by sequencing-by-synthesis.

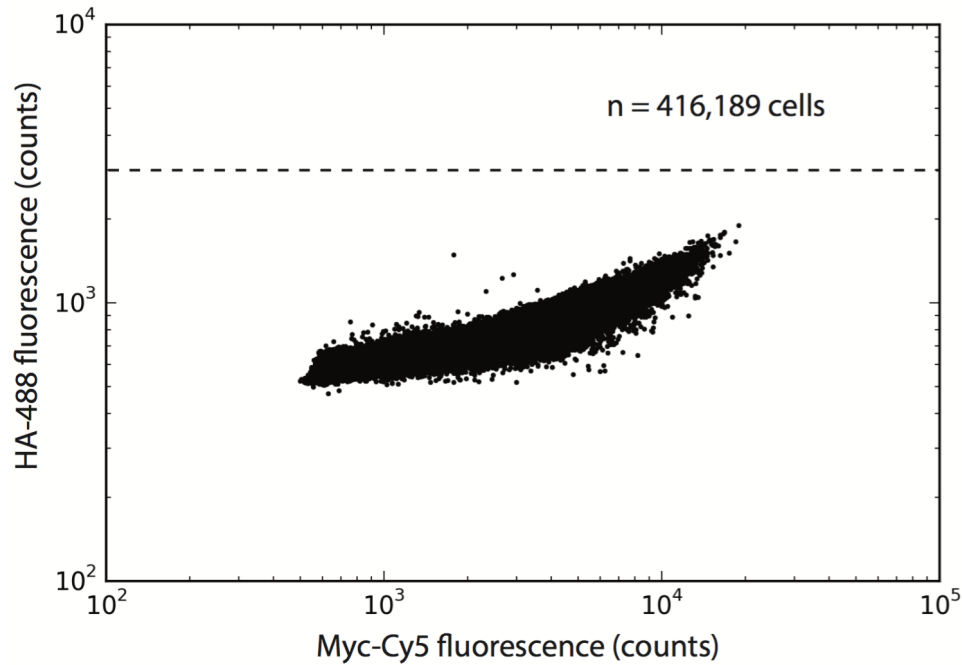
The overall protocol provides a high level of sequence specificity, conferred by hybridization of the RT primer to a unique priming site, hybridization of the padlock to the flanking sites, the preference of the ligase to act only on exactly matched DNA, and sequencing-by-synthesis of the cell-derived barcode sequence itself.





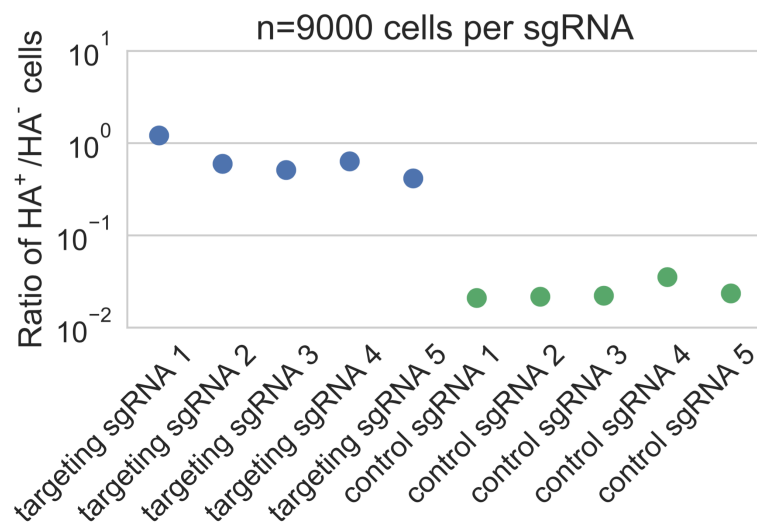
**Fig. S2: Optimization of padlock detection efficiency and amplification yield.**

Padlock detection efficiency was increased more than two-fold compared to literature protocols (19, 45) by optimizing the dNTP concentration and polymerase used for the padlock extension-ligation reaction. A striking improvement in detection efficiency was observed when using Stoffel fragment with a dNTP concentration 1000-fold less than previously published (19). Although Stoffel fragment has been discontinued by its manufacturer, we obtained similar results with another commercially available truncation mutant of *Taq* polymerase (Qiagen TaqIT). While optimizing post-fixation conditions for detection efficiency, we observed that modifying the standard 4% formaldehyde fixative to 3% formaldehyde and 0.1% glutaraldehyde led to a dramatic increase in the yield of overall fluorescence signal from each detected barcode. We presume the improvement was due to an increase in the efficiency of rolling circle amplification, although no specific mechanism was identified. The protocol comparison was performed on a single multi-well plate, using HeLa-TetR-Cas9 cells transduced with lentiGuide-BC. Each data point represents a technical replicate of the *in situ* protocol.



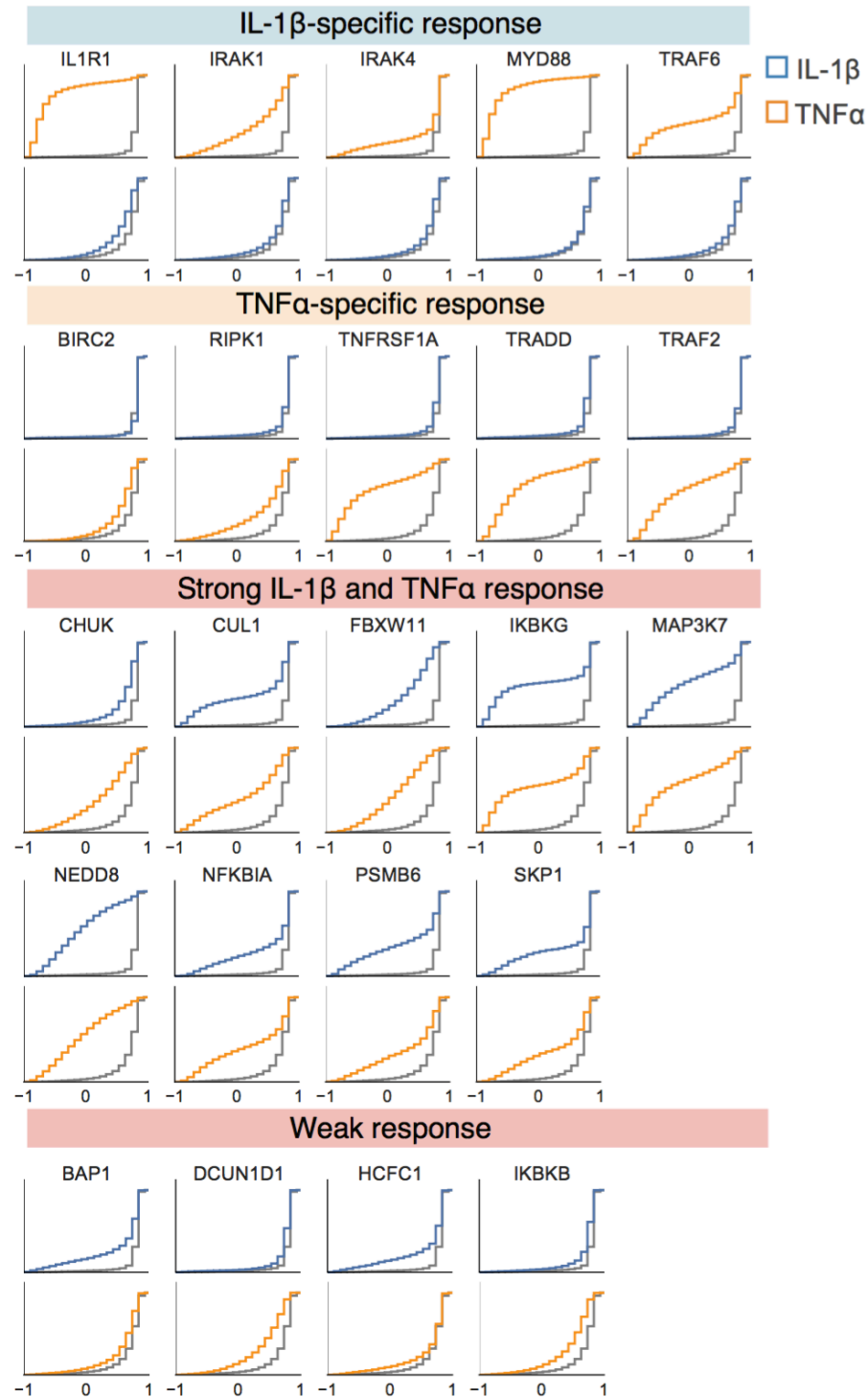
**Fig. S3: Frameshift reporter background.**

The frameshift reporter was read out by microscopy in HeLa-TetR-Cas9-FR cells in the absence of a targeting sgRNA. A *myc* epitope tag in the original, unedited frame was stained to confirm expression levels. The reporter was found to have a very low background, with zero false positives observed among >400,000 cells (dashed line indicates threshold for defining HA<sup>+</sup> cells).



**Fig. S4: Validation of CROPseq-Puro vector in frameshift reporter cells.**

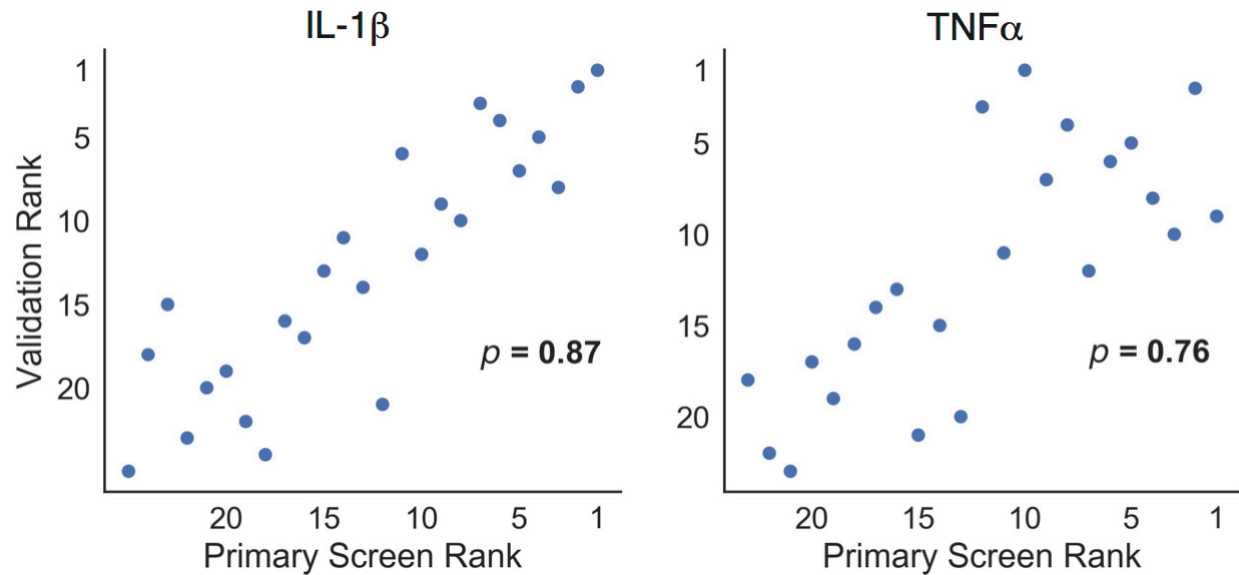
A pool of 5 targeting and 5 control sgRNAs were inserted into CROPseq-Puro and transduced into HeLa-Cas9 frameshift reporter cells at MOI ~10%. After Cas9 induction and frameshift reporter activation, cells were scored for nuclear-localized HA signal. The sgRNA sequence duplicated in the antibiotic resistance cassette was directly sequenced *in situ* over 4 cycles. The ratio of HA<sup>+</sup> cells (19.4% of all cells) to HA<sup>-</sup> cells is shown. The per-read mapping rate was 94% (exact matches to 10/256 possible 4-nt sgRNA prefixes).



**Fig. S5: Arrayed validation of NF- $\kappa$ B screen hits.**

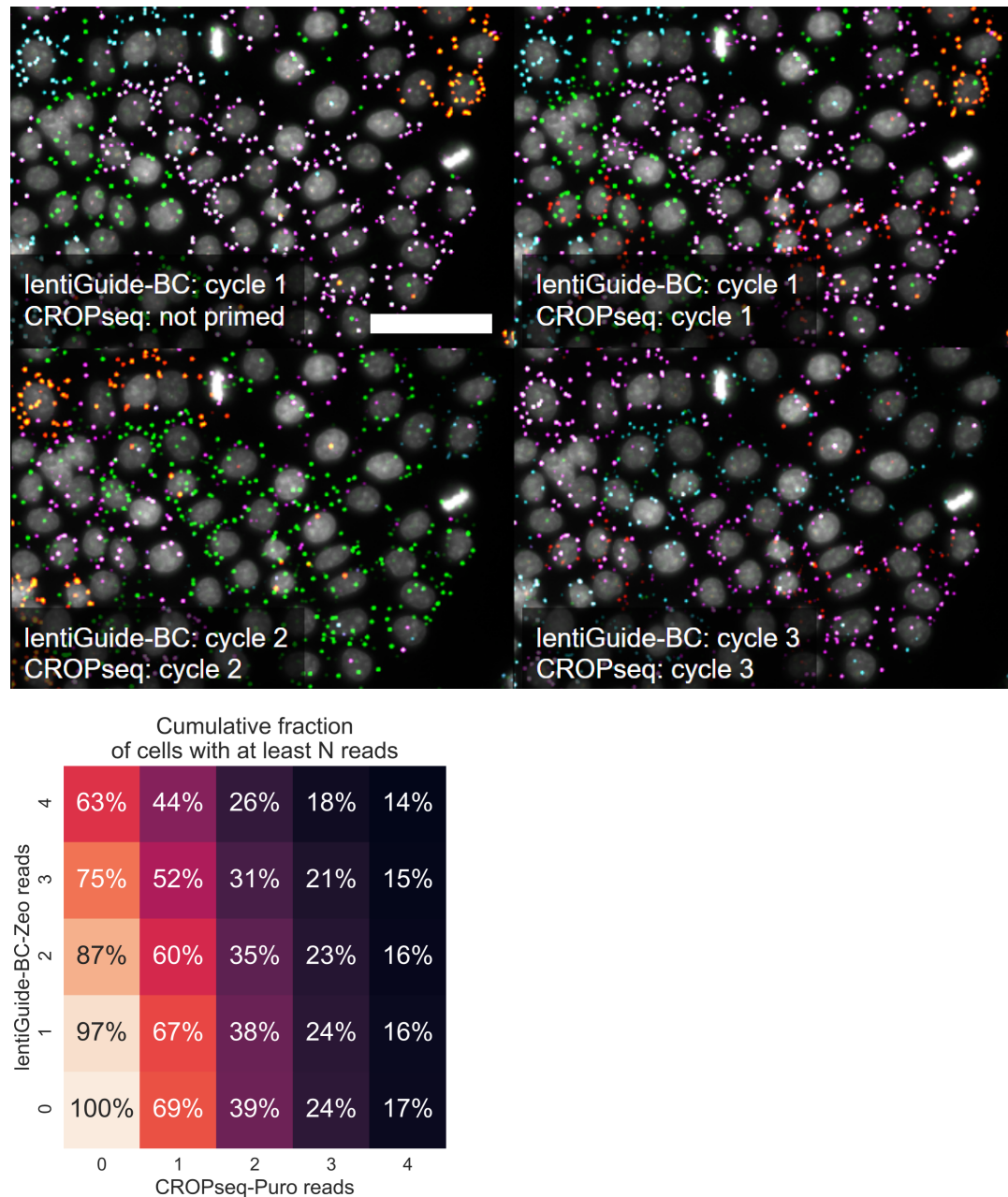
Individual CRISPR knockout gene perturbations were cloned and transduced into HeLa-TetR-Cas9-p65-mNeonGreen cells for validation of primary screen hits. Induction of Cas9, stimulation by IL-1 $\beta$  and TNF $\alpha$ , and translocation score analysis was performed as in the primary screen. Histograms depict the distribution of translocation scores for each knockout.





**Fig. S6: Comparison of primary screen and validation screen rankings.**

For both IL-1 $\beta$  and TNF $\alpha$ , primary screen gene rankings correlate well (Spearman's  $p > 0.75$ ) with rankings in validation screen of single-gene CRISPR-Cas9 knockouts. Proteasome subunits are not shown as they exhibited severe negative fitness effects in arrayed validation experiments, likely biasing the surviving cells to those with incomplete protein knockout.



**Fig. S7: Detection of combinatorial perturbations.**

Multiple perturbations can be delivered via separate lentiviral vectors and detected in the same cell. HeLa-TetR-Cas9 cells were sequentially transduced with lentiGuide-BC (containing a pool of 40 barcodes) and CROPseq-Puro (containing a pool of 10 sgRNAs). The *in situ* padlock detection protocol was the same as for individual vectors, except the RT primers and padlock probes targeting each vector were mixed at equal ratios. Images were acquired at 20X magnification to improve separation of multiple barcode spots per cell (scale bar = 50  $\mu$ m). The first cycle of *in situ* sequencing used only a sequencing primer targeting lentiGuide-BC. Subsequent cycles used sequencing primers targeting both constructs (**top**). The cumulative fraction of cells with at least N reads from each library can be calculated (**bottom**). For example, a total of 35% of all cells imaged had 2 or more reads from each library (heatmap position 2,2).

name	sequence	modifications
pRT_lentiGuide-BC	G+AC+GT+GT+GC+TT+AC+CCAAAGG	LNA bases
pRT_CROPseq	G+AC+TA+GC+CT+TA+TT+TTAACTTGCTAT	LNA bases
pPD_lentiGuide-BC	/5Phos/actggctattcattcgcCTCCTGTTCGA CAGTCAGCCGCATCTGCGTCTATTTAGTGGAGCCC TTGtggtcaatcaacattcc	5' phosphorylation
pPD_CROPseq	/5Phos/gttttagagctagaaatagcCTCCTGTT CGACAGTCAGCCGCATCTGCGTCTATTTAGTGGAG CCCTTGaaggacgaaacaccg	5' phosphorylation
pSBS_lentiGuide-BC	TTCGACAGTCAGCCGCATCTGCGTCTATTTAGTGG AGCCCTTGtggtcaatcaacattcc	
pSBS_CROPseq	TCAGCCGCATCTGCGTCTATTTAGTGGAGCCCTTG aaggacgaaacaccg	

**Table S1: Oligo sequences used for padlock detection and sequencing-by-synthesis.**

Sequences of reverse transcription primer (pRT), padlock probe (pPD), and sequencing-by-synthesis primer (pSBS) used for detecting barcodes in cells transduced with the lentiGuide-BC or CROP-seq vectors. Note that LNA-modified bases in the reverse transcription primer are essential to retain the linkage between the cDNA and fixed RNA, preventing diffusion of cDNA out of the cell (19, 43).

	perturbations (e.g., sgRNAs)	screen replicates	cell coverage	mapped cells	cells (total)	cell density (cells / cm <sup>2</sup> )	sequencing cycle (hours)	# of cycles	total readout time (days)	reagent cost (direct)
NF-κB screen in this study (3 replicates each for IL1b and TNFa)	4,000	6	130	3,120,000	8,320,000	175,000	6	12	3	\$1,485
4,000 perturbation screen with low-density cells	4,000	3	150	1,800,000	4,800,000	25,000	21	12	11	\$5,997
Genome-scale CRISPR screen with 3 replicates	60,000	3	200	36,000,000	96,000,000	175,000	59	12	30	\$17,134

## Table S2: Screening throughput.

The screening approach described in this study uses fully pooled protocols for library construction, cloning, and cell culture. As a result, practical limitations to screen scale come mainly from the time required to image large numbers of cells and the cost of reagents for padlock detection and *in situ* sequencing. Supplementary Table 2 summarizes the scale and throughput of the NF-κB screen performed, as well as two hypothetical screens using either low-density cell culture or a genome-scale perturbation library. The key figure determining both throughput and cost is the total surface area processed, which affects the volume of reagents used and the time per sequencing cycle. The NF-κB screen was read out from a total of ~8 million cells in a single 6-well plate.



<b>step</b>	<b>reagent</b>	<b>cost / 6-well plate (Table S1 row 1)</b>
reverse transcription	RevertAID H minus RT	\$192.42
reverse transcription	RNaseIn	\$55.62
padlock extension/ligation	RnaseH	\$81.60
padlock extension/ligation	TaqIT	\$0.60
padlock extension/ligation	Ampligase	\$97.98
rolling circle amplification	Φ29 polymerase	\$441.00
	<b>padlock detection</b>	<b>\$869.22</b>
sequencing by synthesis (cost per cycle)	Illumina MiSeq v2: incorporation mix, cleavage mix, wash buffer	\$52.50
	<b>total (12 sequencing cycles)</b>	<b>\$1,499.22</b>

**Table S3: Reagent cost estimate for *in situ* barcode readout.**

The major reagent costs (direct) are shown. The cost per 6-well plate corresponds to the NF-κB screen in Fig 4 and the first row of Supplementary Table 2.

### **Table S4: (separate file)**

Libraries of paired barcodes and sgRNAs were amplified from oligo pools and cloned into lentiGuide-BC or lentiGuide-BC-CMV (**Methods**).

### **Table S5: (separate file)**

Each gene was assigned a score for lack of p65-mNeonGreen translocation in response to IL-1 $\beta$  and TNF $\alpha$  stimulation. Gene p-values were calculated by repeatedly sampling sets of three permuted non-targeting sgRNAs to generate gene-level null translocation scores. Genes scored as hits at estimated FDR <10% or <20% were identified by the Benjamini-Hochberg procedure (**Methods**).

### **Movie S1:**

Example data showing a complete 10X magnification field of view with 12 cycles of sequencing, taken from the same dataset used in Fig. 2.