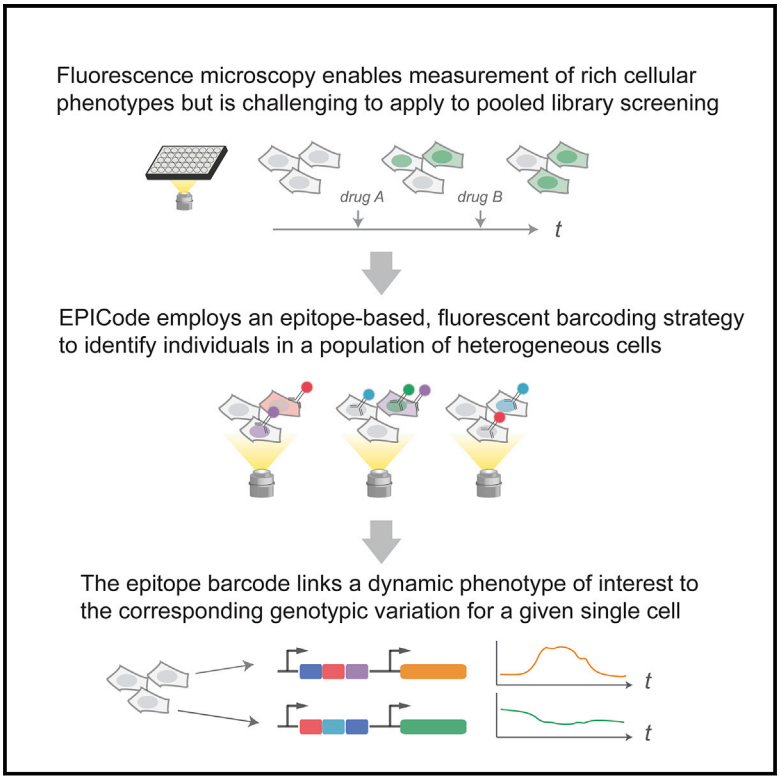
Article

Cell Systems

A multiplexed epitope barcoding strategy that enables dynamic cellular phenotypic screens

Graphical abstract



Highlights

- Combination of epitopes serves as fluorescent barcoding for pooled optical screens
- Differentiating nuclear and cytoplasmic signals enables spatial multiplexing
- The screen was used to optimize the dynamic range of a PKA kinase reporter

Authors

Takamasa Kudo, Keara Lane, Markus W. Covert

Correspondence

mcovert@stanford.edu

In brief

Kudo et al. develop EPICode, an epitope-based barcoding approach for high-throughput screening of dynamic, image-based cellular phenotypes, and use it to optimize the dynamic range of a live-cell fluorescence reporter for PKA kinase activity.

Kudo et al., 2022, Cell Systems *13*, 376–387
May 18, 2022
https://doi.org/10.1016/j.cels.2022.02.006

CellPress

CellPress

Cell Systems

Article

A multiplexed epitope barcoding strategy that enables dynamic cellular phenotypic screens

Takamasa Kudo,[1] Keara Lane,[2] and Markus W. Covert[2,3,*]
[1]Department of Chemical and Systems Biology, Stanford University, Stanford, CA 94305, USA
[2]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA
[3]Lead contact
*Correspondence: mcovert@stanford.edu
https://doi.org/10.1016/j.cels.2022.02.006

SUMMARY

Pooled genetic libraries have improved screening throughput for mapping genotypes to phenotypes. However, selectable phenotypes are limited, restricting screening to outcomes with a low spatiotemporal resolution. Here, we integrated live-cell imaging with pooled library-based screening. To enable intracellular multiplexing, we developed a method called EPICode that uses a combination of short epitopes, which can also appear in various subcellular locations. EPICode thus enables the use of live-cell microscopy to characterize a phenotype of interest over time, including after sequential stimulatory/inhibitory manipulations, and directly connects behavior to the cellular genotype. To test EPICode's capacity against an important milestone—engineering and optimizing dynamic, live-cell reporters—we developed a live-cell PKA kinase translocation reporter with improved sensitivity and specificity. The use of epitopes as fluorescent barcodes introduces a scalable strategy for high-throughput screening broadly applicable to protein engineering and drug discovery settings where image-based phenotyping is desired.

INTRODUCTION

High-throughput screening offers an unbiased, systematic understanding of how various genotypes contribute to a cellular phenotype. In such screens, genetically perturbed cells manifest a selectable phenotype, such as a growth advantage or the expression of a marker that can be used for sorting. Based on the phenotypic readout, variants driving cell population enrichment can be identified by next-generation sequencing. Pooled genetic libraries, enhanced by advances in oligonucleotide synthesis and DNA sequencing, have greatly facilitated high-throughput screening to investigate unknown cell biology and disease mechanisms. Therefore, their use in combination with CRISPR-based genome editing technology for genomic perturbation has become the standard for large-scale genetic screening.

However, such techniques remain limited by the requirement for a screening phenotype that is associated with cell enrichment or depletion. Thus, these methods fall short when the goal is to study complex cellular phenotypes such as dynamic biological events and/or changes in subcellular morphology. Although such rich cellular phenotypes are easily visualized by fluorescence microscopy at a high spatial and temporal resolution, high-throughput screening using microscopic phenotypes is generally not compatible with pooled libraries.

Attempts to integrate pooled library screening and cellular imaging to overcome these limitations are under active development and innovation (Lawson and Elf, 2021). One promising method is *in situ* genotyping, which uses hybridization or sequencing to provide a direct link between a genetic variant and its imaging phenotype (Feldman et al., 2019; Wang et al., 2019). Such methods have opened exciting new avenues in image-based pooled screening and have inspired our and other efforts to reduce the time, effort, and technical overhead required while expanding the range of possible applications beyond what has been previously generated.

In this paper, we present a method called EPICode (epitope-phenotype immunofluorescence barcode), which combines high-throughput pooled screening and imaging-based phenotyping through the use of short epitope combinations that serve as barcodes for each construct in a pooled library. Epitope barcodes were previously used to analyze pooled CRISPR libraries with mass cytometry at a level suitable for focused screens (100–500 genes) (Wroblewska et al., 2018). Here, we expand this concept and apply it to live-cell imaging, vastly expanding the capacity for the multiplexed localization of epitopes within the cell. This level of spatial resolution enables the encoding of barcoding information not only in the fluorescence color but also by its subcellular location. These features make protein-based barcoding a powerful, yet simpler, alternative for *in situ* genotyping.

We previously reported a class of biosensors, called kinase translocation reporters (KTRs), that are genetically encoded fluorescent reporters capable of converting phosphorylation events into nuclear-cytoplasmic translocation (Regot et al., 2014). In addition to a substrate recognition motif and the

fluorescent tag, KTRs also include phosphorylatable domains (a nuclear localization signal [NLS] and a nuclear export signal [NES]) that direct KTRs to the cytoplasm when phosphorylated and KTRs to the nucleus when dephosphorylated. Unlike fluorescence resonance energy transfer (FRET) kinase reporters, KTRs allow multiplexing and are comparably easier to design. Because here we sought to leverage multiplexing to resolve both subcellular and dynamic information, we felt that the optimization of a KTR construct would be a strong challenge and, if successful, would serve as a proof of concept to our approach.

KTRs currently exist for JNK, ERK, p38, AKT, CDK1, CDK4, and protein kinase A (PKA) (Hung et al., 2017; Maryu et al., 2016; Regot et al., 2014; Spencer et al., 2013; Yang et al., 2020). They have been used to measure the dynamics of single or multiple signaling pathways in several biological systems, ranging from cancer cells to zebrafish scales to mouse embryos (De Simone et al., 2021; Pokrass et al., 2020). Although the biology of PKA is of important scientific and therapeutic interest given its role in cell metabolism and neuronal and immune signaling (London et al., 2020; Turnham and Scott, 2016; Yan et al., 2016), the original PKA-KTR is somewhat limited by a lower dynamic range relative to KTRs developed for many other kinases. Because PKA's substrate binding affinity hinges on a few amino acids that surround a phosphorylation site, improving upon its dynamic range has remained an outstanding challenge. Creating an optimized version of PKA-KTR would, therefore, not only demonstrate an important milestone in screening for dynamic, live-cell reporters but would also generate a useful tool of interest in its own right.

Here, we describe the method itself, together with its application in PKA-KTR optimization. As a result of the dynamic phenotyping screen, we were able to retrieve genetic sequences for a new PKA live-cell reporter. Relative to the original construct, the new reporter contained 9 amino acid replacements and exhibited a superior dynamic range. Our findings demonstrate the utility of EPICode, expanding our toolbox of methods that integrate pooled library screening with imaging-based phenotypes.

RESULTS

To enable high-throughput screening of a genetically heterogeneous population that encodes dynamic phenotypes, we create a library of constructs, each of which pairs a genetic variant with an epitope-based barcode (Figure 1). The epitopes used for barcoding are chosen and localized within the cell to maximize the number of variants that can be distinguished under a microscope. The functional variants are generated to screen for a property of interest and can include, for example, the mutagenesis of a gene, a guide RNA, or a cDNA library. The size of the overall construct is a critical feature of the design envelope due to commercial oligo synthesis constraints.

The library of constructs is then analyzed in parallel, using genotyping and dynamic phenotyping. Library sequencing provides a direct link between each genetic variant and its epitope barcode. As the constructs are introduced into cells, dynamic phenotyping is observed under a microscope as cells respond to one or more stimuli of interest. Following these observations, immunofluorescence is used to determine the epitope barcode for each individual cell, which provides a direct link between each dynamic phenotype and its epitope barcode. The epitope barcodes thus link the genetic variant to its corresponding dynamic phenotype.

Barcodes can be identified using sequential multicolor immunofluorescence

To meet our criteria of a high-throughput method, we felt that the number of measurements needed to be at least in the order of tens of thousands. With the epitopes arranged in a triplet combination, 18 epitopes can be used to encode $6^3 = 216$ unique combinations. If a method could detect such barcodes in two cellular locations, it would enable the encoding of 46,656 combinations in total. Thus, the first challenge in developing our workflow was to determine how to visualize up to 18 cellular epitopes in a single sample. The solution would require imaging multiple fluorophores simultaneously, but given the previously noted problems with spectral overlap (Zimmermann, 2005), it would also require the capability for multiple rounds of imaging with the same fluorophores. To achieve our goal of imaging up to 18 different barcodes, the imaging component would, therefore, require resolving six different fluorophores in three imaging cycles.

First, we tested whether using a single fluorophore for three cycles would allow us to identify multiple vector identities *in situ*. We created three lentiviral vectors, each expressing nuclear histone H2B fused with a different combination of a fluorescent protein and epitope: CFP-MAT, iRFP-E2, and YFP-NWS. We then infected HeLa cells in culture with each construct, plated all the infected cells together, and treated the cultures with three cycles of immunofluorescence. For each cycle, we used a red fluorescent CF594-conjugated antibody recognizing one of the three epitopes (Figure 2A) to determine fluorescent intensity. We then computationally aligned the images using the previously described methods (Pluim et al., 2003; Viola and Wells, 1995).

To determine the fluorescence signal generated by each cycle, we calculated the increase in fluorescence intensity between cycles (also correcting for photobleaching, see STAR Methods). For the image corresponding to the cycle of interest, we subtracted the image acquired in the previous cycle (if any) (Figure 2B; Wählby et al., 2002). To this end, we segmented the nuclear region for each cell based on the fluorescent protein signals and then quantified the intensity of fluorescent markers and increase in fluorescent intensity between cycles for each individual cell. For a cell harboring a given fluorescent protein-epitope pair, we observed an increase in the fluorescent signal in the correct imaging cycle, and those fluorescent proteins were only detected in the cells that harbored their conjugate epitopes and not in others (Figure 2C).

To quantify this degree of fluorescent protein-epitope co-occurrence, we used minimum cross-entropy thresholding to classify cells as positive or negative for each fluorescent protein and epitope. When we then computed the Matthews correlation coefficient (MCC) as the degree of co-occurrence, we observed a high coefficient for co-expressed pairs and low scores for other pairs (Figures 2D and S1A). We also confirmed that the associations between pairs were not properly extracted when the raw fluorescent intensities were used (Figure S1B). We concluded from these results that sequential immunofluorescence could be used to decode the identity of multiple vectors using the

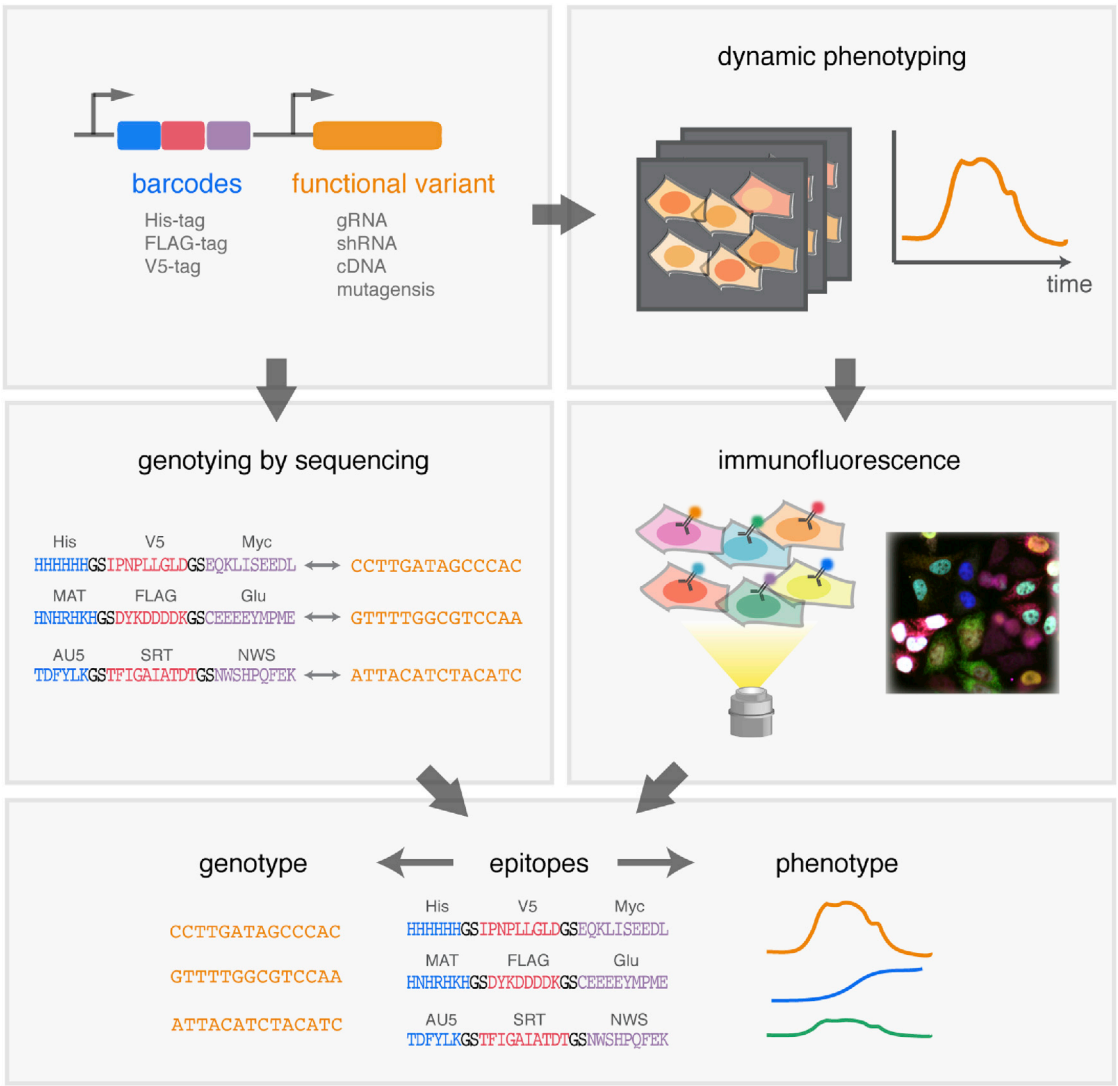


Figure 1. Workflow of image-based pooled screening using EPICode

Each construct in a pooled library carries both epitope barcodes and functional variants (e.g., gRNA, shRNA, cDNA, and mutagenesis). Cells transduced with the library are live-cell imaged for dynamic phenotyping, followed by immunofluorescence to identify the presence of epitopes in each individual cell. Sequencing reveals the pair of epitope barcodes and variant sequences. Phenotypes are then mapped back to genotypes through the epitope barcode.

same antibody-fluorophore conjugates. To remove residual signals from dyes and fluorescent proteins, we incorporated chemical bleaching in subsequent sections (Lin et al., 2015).

Next, we tested whether we could resolve six fluorophores simultaneously in individual cells. We established two cell lines engineered to express H2B fused to a fluorescent protein (monomeric Kusabira-Orange 2 [mKO2] or teal fluorescent protein [TFP]) and three epitopes: mKO2-T7-FLAG-Myc and TFP-HA-NWS-Univ. The cells were then plated in the same well and subjected to staining with the six epitope-specific antibodies, each of which had been primarily conjugated with one of six different dyes: A405, CF430, A488, A532, CF594, and CF680(Figure 2E). We then acquired two sets of six images at different wavelengths, one set before and one after staining. The first set enabled us to use pixel-wise subtraction to remove any existing fluorescent protein signal. To isolate the fluorescence intensity signal from each fluorophore, we applied a linear unmixing algorithm (Figure 2F; STAR Methods). The images were then segmented and evaluated using the MCC to quantify co-occurrence as before (Figures 2G and S1C). The signals indicating the presence of the T7, FLAG, and Myc epitopes were only found in the cells expressing mKO2, and, conversely, the signals indicating the presence of the HA, NWS, and Univ epitopes were only found in the cells expressing mTFP (Figure 2H). Taken together, these experiments demonstrated that these fluorophores could be used to distinguish the presence or absence of these six epitopes.

A chain of multiple epitopes serves as an optical barcode of vector identity

To test for feasibility, we first needed to establish whether an epitope barcode could distinguish individual cells in a

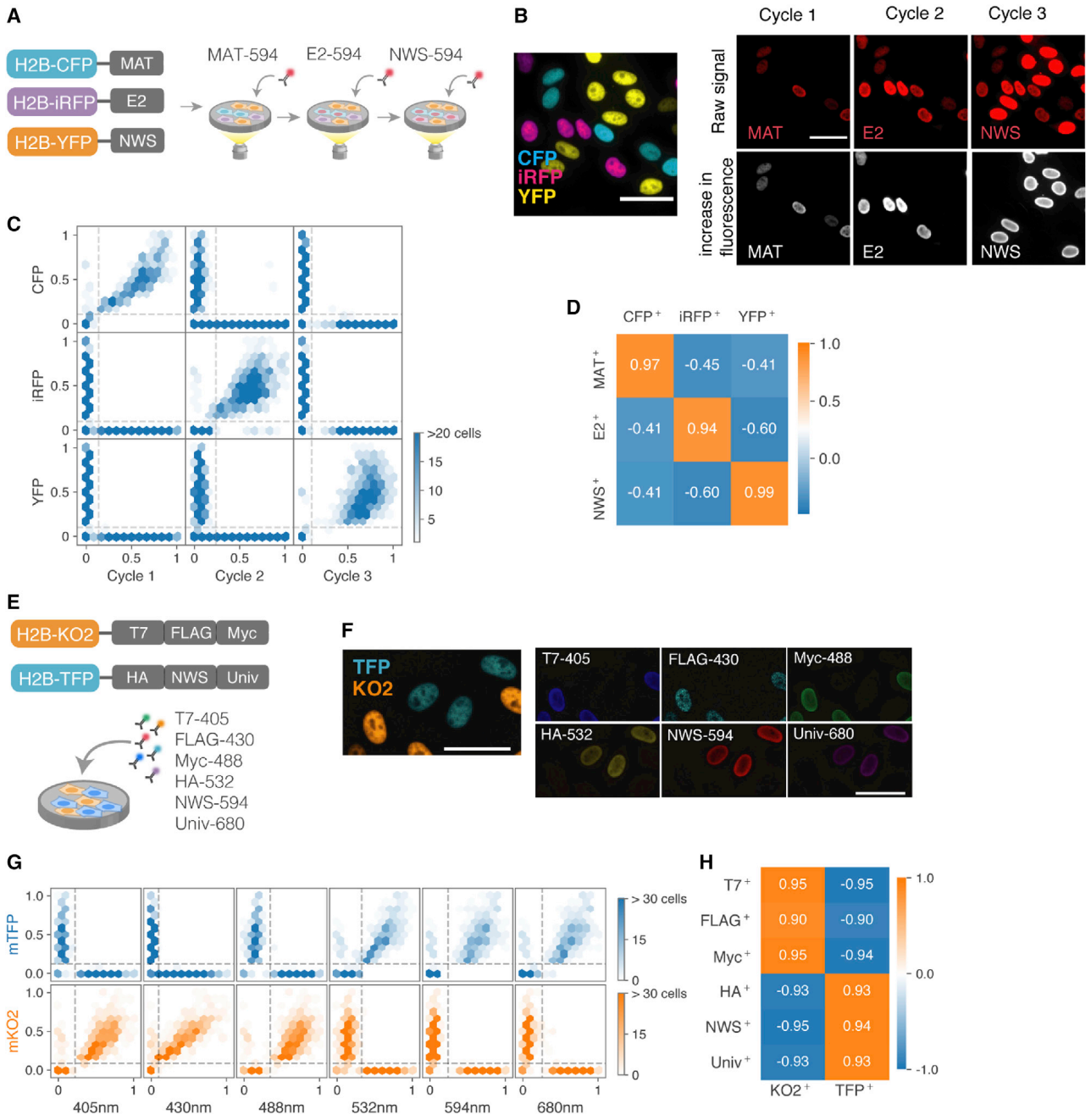


Figure 2. Development of sequential multicolor immunofluorescence

(A) Schematic of sequential immunofluorescence.

(B) Three cell lines were pooled together and subjected to three cycles of immunofluorescence. To the left is a representative image of the fluorescent proteins from 2,028 cells (cyan for H2B-CFP, purple for H2B-iRFP, and yellow for H2B-YFP). To the right, the top three panels represent immunofluorescence signals after each cycle of staining. The bottom three panels represent an increase in fluorescence between cycles.

(C) Association between immunofluorescence signals and fluorescent protein signals. Hexagons are colored according to cell density. Brightness is normalized by the maximum value. Dashed lines indicate threshold values obtained by minimum cross-entropy thresholding.

(D) Co-expressed pairs of fluorescent protein and epitope exhibited a high MCC. The color gradients and values indicate the MCC.

(E) Schematic of six-color immunofluorescence. Antibodies targeting the six different epitopes were primarily conjugated with different fluorophores.

(F) Two cell lines were pooled and subjected to six-color immunofluorescence. The left image is the fluorescent protein signals, representing 2,404 individual cells. The right panels represent the reconstructed immunofluorescence images after pixel-wise subtraction and spectral unmixing. Scale bars, 50 μm.

(G) Association between immunofluorescence signals and fluorescent protein signals. Hexagons are colored according to cell density. Brightness is normalized by the maximum value. Dashed lines indicate threshold values obtained by minimum cross-entropy thresholding.

(H) Six-color immunofluorescence identified the association between the fluorescent proteins and epitope sets. The color gradients and values indicate the MCC.

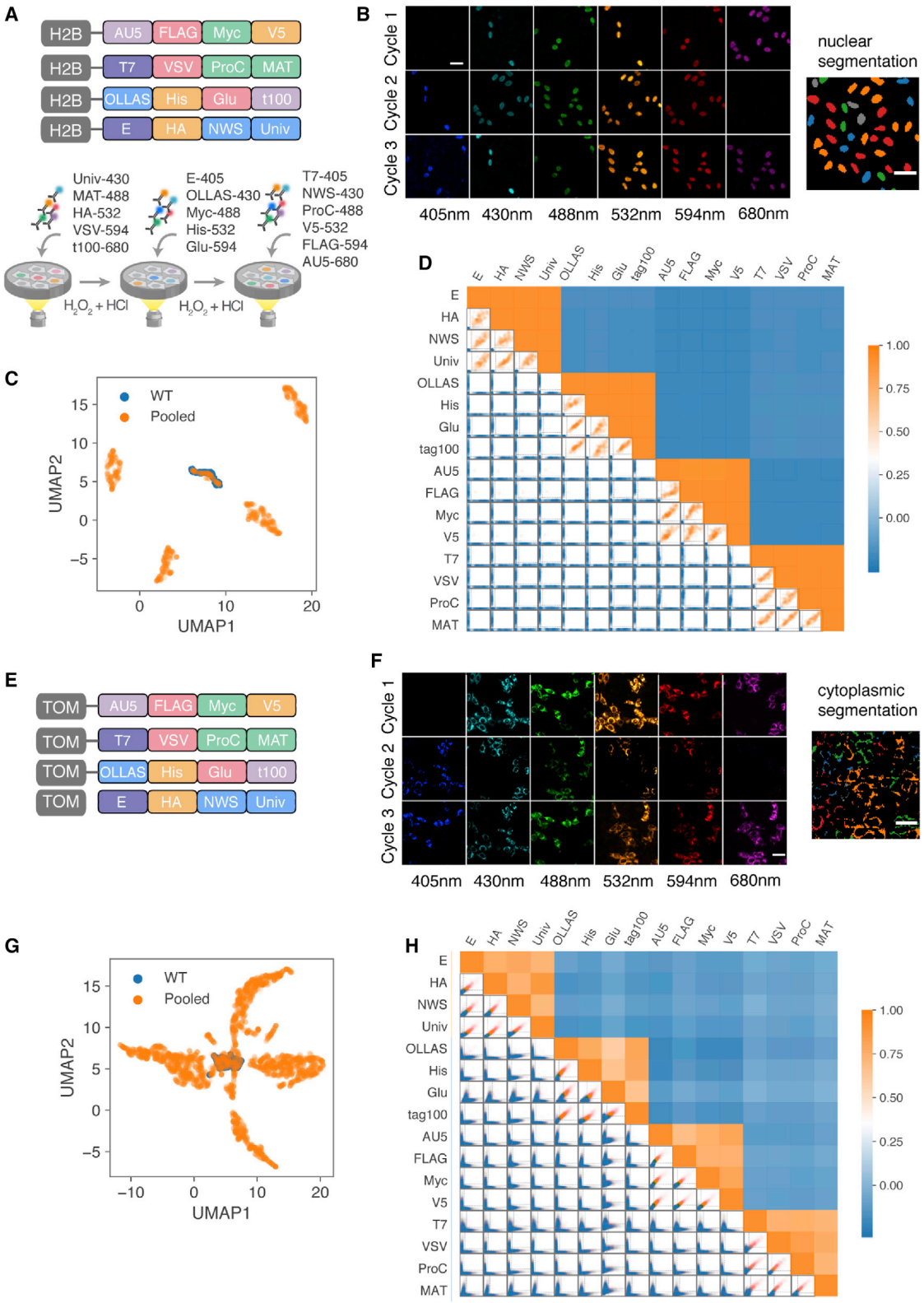


Figure 3. A chain of multiple epitopes serves as an optical barcode of vector identity

(A) Schematic of the nuclear epitope constructs (top) and experimental flow (bottom).

(B) Representative immunofluorescence images and segmentation mask for 1,305 individual cells. The images were processed by pixel-wise subtraction and spectral unmixing. The color of segmented nuclei corresponds to the clusters identified by the clustering applied to the UMAP score. Scale bars, 50 μm.

(legend continued on next page)

population. Therefore, we created constructs in which H2B was fused to a combination of four epitopes, which were chosen based on their compatibility with immunofluorescence and short sequences. For this experiment, we designed four combinations of constructs such that each construct had four of the sixteen epitopes and that no epitope was shared by any construct (Figure 3A). We then incorporated these constructs into lentiviral vectors, which were used to infect the HeLa cells in culture at a low multiplicity of infection (MOI). We then plated infected cells together and performed three cycles of immunofluorescence stain-and-wash steps in which five or six antibody-fluorophore conjugates were applied to fixed cells in each cycle (Figure 3B). When we performed fluorescence microscopy to generate images identifying the presence of any or all of the corresponding six colors of the fluorophores, we found that we could easily distinguish between cells expressing different constructs. The images were then segmented and analyzed computationally as previously described (Figure 3B), generating a quantitation of fluorescence intensity corresponding to each of the sixteen epitopes for each cell.

To derive meaningful clusters from the high-dimensional data, we projected the fluorescence intensity for each nuclear epitope in each individual cell into two-dimensional spaces using uniform manifold approximation and projection (UMAP) (McInnes et al., 2018), a commonly used dimensional reduction method. UMAP analysis separated these data into five distinct clusters (Figure 3C). One of the clusters (central to the figure) corresponded to our wild-type control cells, where epitopes had not been introduced. When we quantified the co-occurrence between epitopes in each individual cell, we were able to determine that all the co-occurrences we observed were due to the co-occurrence in our original genetic constructs (Figures 3D and S1D).

Next, we explored whether we could use non-nuclear epitopes to distinguish individual cells in a population to further increase the multiplexing capacity. To this end, we created four new genetic constructs, each of which fused a four-epitope barcode to mitochondria-specific target translocase of outer mitochondrial membrane 20 (TOMM20) (Huff et al., 2017; McArthur et al., 2018; Figure 3E). As before, we established the cell lines by lentivirally transducing the constructs at a low MOI, plated the cell lines together, and performed three cycles of immunofluorescence. Again, the fluorescence signal was specific to each epitope, but this time, it was present in the cytoplasmic region (Figure 3F). The cytoplasmic regions of each individual cell were then segmented and analyzed, similar to the steps we performed for nuclear epitopes. Once again, the UMAP was able to identify five different clusters, where one of the clusters overlapped with the wild-type cells without epitope expression; however, the separation between clusters was less distinct compared with what we had observed in the nucleus (Figure 3G). We believe this discrepancy reflects the biology of the more stable H2B and the fact that the nucleus is easier to segment computationally from our processed images. Nonetheless, we were able to successfully reconstruct the co-occurrence of epitope pairs that were present in our genetic constructs (Figures 3H and S1E). These results suggested that EPICode could be amenable to spatial multiplexing and resolve proteins localized in both the nucleus and cytoplasm.

Epitope barcodes enable image-based pooled library screening

Given these robust data, we were encouraged to apply our technology in a pooled screen for a dynamic phenotype using live-cell imaging. To trigger translocation events, a genetic construct was engineered to have the docking site for a kinase of interest proximally located to a nuclear localization and nuclear export sequence. Because phosphorylation sites are present within these sequences, kinase binding and phosphorylation decrease the charge on these sites and thereby impact the import and export rate of the construct. A fluorescent protein indicates the localization of the construct over time as a proxy for kinase activity.

We were particularly interested in improving upon the dynamic range of the fluorescent PKA-KTR. Among reported KTRs, the PKA-KTR is limited by a relatively narrow dynamic range of translocation, and increasing this range is technically challenging. Unlike many kinases that possess distal docking sites, the binding affinity to PKA is determined by a few amino acids surrounding a phosphorylation site (Ubersax and Ferrell, 2007). Thus, the incorporation of localization signals around the phosphorylation site risks interfering with PKA binding (Kudo et al., 2018; Taylor and Radzio-Andzelm, 1994). Optimization via sorting and sequencing is also unfeasible since the evaluation of KTRs requires both subcellular and dynamic information.

To apply our high-throughput pooled screening approach to optimize PKA-KTR, we created a lentiviral library with two main parts: the optical barcode and a PKA-KTR that would be mutagenized to create variants (Figures 4A and S2; STAR Methods). In the epitope barcode, H2B and TOMM20 are each fused with a chain of three epitope cassettes and separated by the self-cleaving 2A peptide (STAR Methods; Tables 1 and S1). Each cassette has six epitope candidates to be assembled, in theory, creating 46,656 possible combinations. We designed the epitope barcode region to be <300 bp in order to be compatible with MiSeq Illumina sequencing. Because allowing mutagenesis of the entire KTR would create an unmanageable number of variants, we needed to create the mutagenized KTR region based on rational design. To reduce the overall complexity, we selected ten specific residues that surrounded

(C) UMAP identified five distinct population groups from nuclear epitopes.
(D) The pairwise relationship between nuclear epitopes represented as the scatter plots (lower triangle) and MCC matrix (upper triangle). In the scatter plots, dashed lines indicate threshold values, and the population above both of the thresholds is labeled as orange. Color gradients indicate the F1-score.
(E) Schematic of the mitochondrial epitope constructs (top) and experimental flow (bottom).
(F) Representative immunofluorescence images and segmentation mask for 6,107 individual cells.
(G) UMAP identified five distinct population groups from mitochondrial epitopes.
(H) The pairwise relationship between the mitochondrial epitopes is represented as the scatter plots (lower triangle) and MCC (upper triangle).

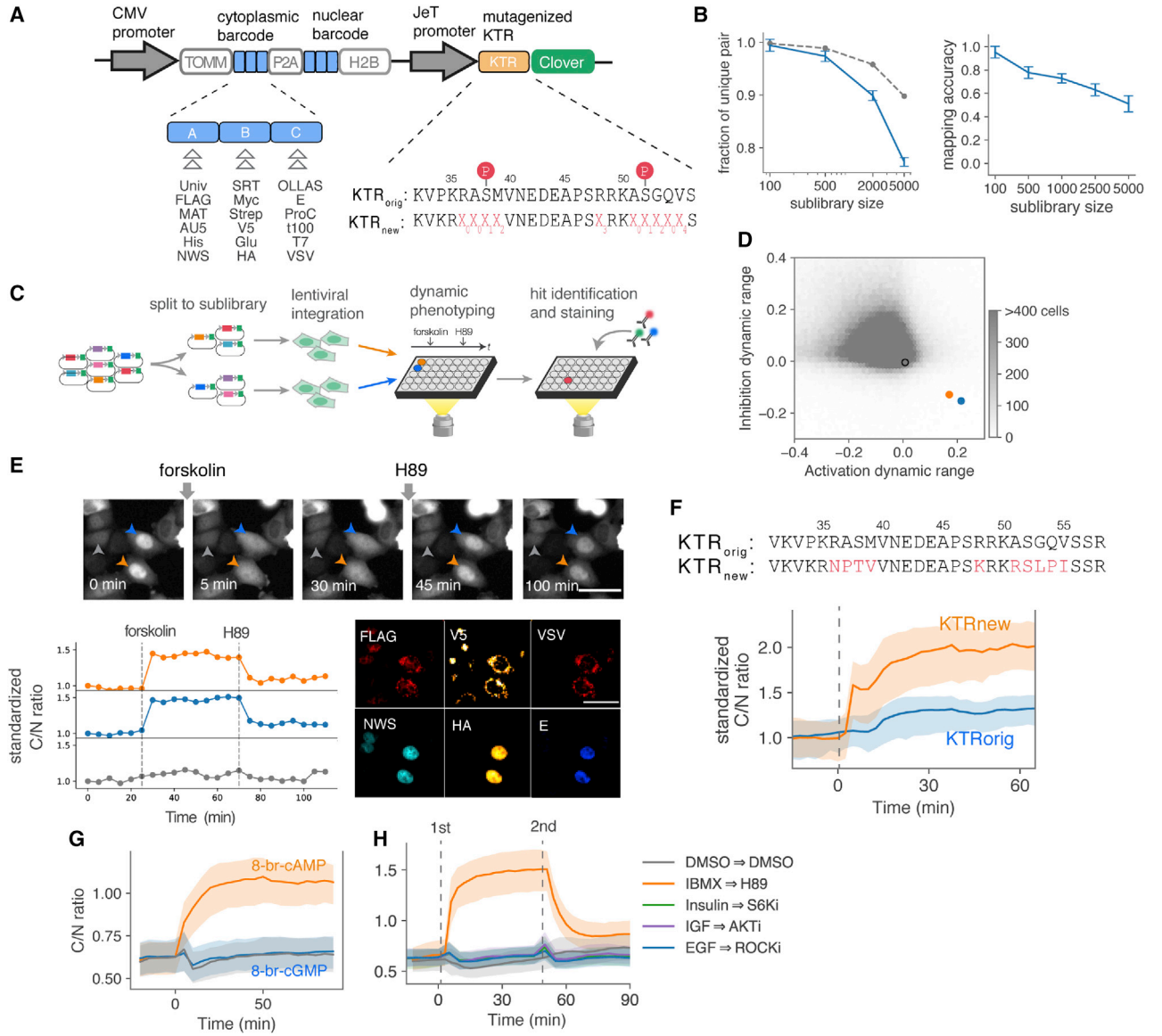


Figure 4. Epitope-based barcoding performed pooled screens for a dynamic phenotype

(A) Design of library constructs. Epitope barcodes and KTR mutants were expressed from the CMV and JeT promoters, respectively. H2B and TOMM20 are each fused with a chain of three epitope cassettes to comprise epitope barcodes. Each cassette contains one of the six epitope candidates. KTR has ten variable amino acids. Phosphorylation sites in the original constructs are displayed as P. Degenerate codons, NNK, WCT, DYK, ARR, and NTN, were used for X_0 to X_4.

(B) (Left panel) Fraction of the unique pair estimated by random sampling (blue line) and theoretical rate (gray line). The mean and standard error of the mean (error bar) were estimated from 100 simulations. (Right panel) Accuracy of mapping estimated by the permutation strategy. The mean and standard error of the mean (error bar) were estimated from 30 permutation steps.

(C) Schematic of screening pipeline.

(D) Reporter dynamic range of the KTR variants upon PKA activation and inhibition. The changes of the C/N ratio after stimulation were plotted against the changes after inhibition. Hexagons are colored according to cell density. A summary of 912,778 cells was plotted. Orange, blue, and dark circles correspond to the cells in (E).

(E) Reporter dynamics and immunofluorescence signals of the translocating mutants. Two cells (orange and blue), but not a neighboring cell (gray), showed cytoplasmic translocation with forskolin treatment (20 μM) and nuclear translocation with H89 treatment (7.5 μM) (image panels). The extracted dynamics of cytoplasmic-to-nuclear ratio (C/N) were shown (bottom left). C/N values were standardized by dividing each time series by the values at the first time point for comparison. The bottom right panels represent immunofluorescence signals. Scale bars, 50 μm. See also Video S1.

(F) A new PKA-KTR exhibits a large dynamic range upon 8-br-cAMP stimulation. The amino acid sequences of the original and new PKA-KTR are shown with red letters indicating the mutation sites. The 3T3 cells expressing PKA-KTR_{orig} (blue) and PKA-KTR_{new} (orange), together with nuclear marker H2B-miRFP670, were stimulated with 8-br-cAMP (0.1 μM). The line represents the mean and the shade represents the 25th to 75th percentile (412 and 385 cells). For comparison, the C/N ratio was standardized by dividing each by the value at the first time point. See also Video S2.

(legend continued on next page)

phosphorylation sites in the original PKA-KTR (PKA-KTR$_{orig}$) (Figures 4A and S2; STAR Methods).

Our construct design included two functional parts, namely epitope barcodes and KTR variants. As a result of combining the two parts, the same epitope combination might be ligated to different KTR variants, leading to a non-unique barcode-KTR variant pair when we perform the screening. To prevent such errors, we bottlenecked our library into smaller sub-libraries such that virtually all of the epitope tags would be unique (Emanuel et al., 2017). To determine the size of the sub-libraries, we sequenced the overall library and found that the epitope frequencies were biased during the library construction process (Figure S2B). Thus, we randomly sampled the sub-library, taking these biases into account, to estimate the rate of non-unique pairs (Figure 4B, left panel). Our analysis shows that the rate of non-unique pairs is <3% at the sub-library of size 500.

Using these smaller sub-libraries instead of the complete library enabled us to narrow down the search space for identifying the combination of epitopes contained within a particular cell. Epitope expression is not expected to be strongly bimodal for all the epitopes because, as mentioned above, epitope frequencies are likely to vary (Figure S2B), and moreover, the use of 2A can lead to a decrease in expression (Liu et al., 2017). Therefore, we decided to move beyond a simple thresholding approach and instead use a likelihood determination to identify the most probable epitope combination for a given construct. Through the use of negative control samples that did not express any epitopes but were stained using immunofluorescence, we were able to obtain fluorescence intensity distributions in epitope-negative samples for each epitope. We then compared the staining results from our test samples with the epitope-negative distributions, obtaining a percentile score (on a scale of 0 to 1) for each epitope stain. These scores were then averaged for the likelihood of expressing certain epitope combinations for each set of six epitopes present in a sub-library. We determined the correct epitope set to be the one with the highest mean percentile score (STAR Methods; Figure S3C).

To evaluate the accuracy of the likelihood approach, we adopted a permutation strategy, beginning with the creation of a minimal pooled library comprising only two vectors with known epitope combinations. We transduced this library to HeLa cells at a low MOI and then plated and stained the cells as described above. To reduce the risk of template switching, lentiviruses were packaged together with a low homology decoy vector as described previously (Figure S3D; STAR Methods; Adamson et al., 2018; Feldman et al., 2018). As the output for each cell, we obtained a vector of size 32, corresponding to the fluorescence intensity for each of the sixteen epitopes in both the nuclear and cytoplasmic regions. Because our likelihood approach requires a set of possible candidates to evaluate the cellular fluorescence intensity against, we can test our method by first combining the stained set of epitopes with a randomly generated set of other candidates and then determine how often the correct epitopes are called from the permuted library. We constructed a virtual sub-library of size N for each permutation step (see the x axis in Figure 4B), quantified the mean percentile score for each candidate in the sub-library, and called the epitope set present in each cell as the set with the highest score. We then asked what fraction of the cells were mapped to the correct epitope set. We found that the estimated accuracy of identification for the epitope set in each cell was ~80% for the sub-libraries containing 500 constructs (Figures 4B, right panel, and S3E). We also found that each epitope contributes to the mapping to a different extent (Figure S3F; see also STAR Methods).

Given these results, we split the pooled libraries into smaller sub-libraries, each containing ~500 constructs, for our PKA-KTR optimization pooled screen. The sub-libraries were transduced to HeLa cells at a low MOI. After antibiotic selection, each cell library was plated in a separate well in a 96-well plate. We then imaged cellular responses to PKA activation via forskolin, treating cells with the PKA inhibitor H89 after 45 min. U-Net (Moen et al., 2019; Ronneberger et al., 2015) was then used to segment the cells from the KTR construct fluorescence and localization (Zikry et al., 2019), and we quantified the C/N ratio over time for analysis of the activation dynamics. In total, we analyzed ~1 million cells carrying >10,000 variants (Figure S3G). For each cell, we summarized the dynamics by quantifying the changes in the C/N ratio after PKA activation and also after PKA inhibition. Overall, the variants that showed translocation were rare because most mutations led to a loss of function of the construct (Figures 4D and S3H). However, in one of the sub-libraries, we observed that two neighboring cells exhibited a strong nuclear-to-cytoplasmic translocation induced by forskolin, followed by cytoplasmic-to-nuclear translocation upon H89 treatment (Figure 4E; Video S1).

We focused on this sub-library for further staining and sequencing analysis. Using immunofluorescence staining, we quantified nuclear and cytoplasmic fluorescent intensity for each epitope as described above (Figure 4E). We also sequenced the sub-library, which enabled us to determine the set of candidate epitopes present in the cells. Using these candidate sets, we were able to compute the mean percentile score for each candidate and thus the most likely set of epitopes expressed by our two cells of interest. For both cells, the most likely set of expressed epitopes was the cytoplasmic FLAG, V5 and VSV, and nuclear NWS, HA and E-tag. Further sequencing analysis revealed that this set of epitopes was linked to mutations in 9 out of 10 positions targeted for PKA-KTR mutagenesis: R36N, A37P, S38T, M39V, R48K, A51R, G53L, Q54P, and V55I. Based on their proximity, epitope expression, and sequences, these two cells are likely sibling products of cell division. Conveniently, this strengthened our confidence that the hit was "real."

(G) 8-br-cGMP did not induce the translocation of the PKA-KTR$_{new}$. Cells were stimulated by 8-br-cAMP (0.25 μM, orange, 838 cells), 8-br-cGMP (0.25 μM, blue, 843 cells), or DMSO (gray, 740 cells).

(H) Dynamics of PKA-KTR$_{new}$ upon various stimuli. Dashed lines represent the timing of the first and second stimulations. The cells were first stimulated with IBMX (250 μM), insulin (10 μg/mL), IGF (25 ng/mL), or EGF (25 ng/mL), and then they were subjected to the inhibitor treatments with H89 (7.5 μM), S6Ki (LY2584702, 0.1 μM), AKTi (GDC-0068, 2 μM), or ROCKi (TC-S 7001, 10 nM). The line represents the mean and the shade represents the 25th to 75th percentile (1,018, 857, 824, 984, 740 cells).

Table 1. Epitopes used in this study, see also STAR Methods

Epitope	Sequence
FLAG	DYKDDDDK
VSV	YTDIEMNRLGK
V5	IPNPLLGLD
His	HHHHHH
Myc	EQKLISEEDL
Protein C	EDQVDPRLIDGK
OLLAS	SGFANELGPRLMGK
HA	YPYDVPDYA
MAT	HNHRHKH
Glu-Glu	CEEEEYMPME
NWS	NWSHPQFEK
AU5	TDFYLK
Universal	HTTPHH
SRT	TFIGAIATDT
tag-100	EETARFQPGYRS
E	GAPVPYPDPLEPR
T7	MASMTGGQQMG
Strep	SAWSHPQFEK
E2	SSTSSDFRDR

The first 18 epitopes were chosen for the screening library. The NWS-tag and Strep-tag were recognized by the same antibody. See also STAR Methods.

To validate the hit, we generated a new reporter, PKA-KTR$_{new}$, which incorporated the mutations determined by our screen. PKA-KTR$_{new}$ and PKA-KTR$_{orig}$ were then introduced into separate 3T3 cell cultures, which were subsequently prepared for live-cell fluorescence microscopy as before. When both cultures were stimulated with a membrane-permeable cAMP analog, 8-br-cAMP, PKA-KTR$_{new}$ exhibited a much larger nuclear-to-cytoplasmic translocation compared with PKA-KTR$_{orig}$, which barely showed any translocation (Figure 4F; Video S2). Induced the localization of PKA-KTR$_{new}$ by 8-br-cAMP was also acutely reversed upon the addition of the H89 inhibitor (Figure S3I). To confirm the new reporter's specificity for PKA, we also exposed PKA-KTR$_{new}$-expressing cells to a panel of stimulants and inhibitors, including those perturbing other AGC kinase families, PKG, S6K, Akt, and ROCK kinase, as well as two additional PKA activators. Both of the PKA activators and none of the off-target perturbants impacted the translocation of PKA-KTR$_{new}$ (Figures 4G, 4H, S3J, and S3K). Thus, by developing a live-cell PKA KTR with improved sensitivity and specificity, we demonstrate a proof of concept for the applicability of our method.

DISCUSSION

In this work, we report on a method for conducting image-based pooled screens by using *in situ* genotyping through epitope barcoding. To retrieve barcode information for up to 18 epitopes, we established robust staining techniques and then expanded the multiplexing capacity by localizing each epitope either in the nucleus or the mitochondria, which is indicative of a cytosolic localization. We applied this approach to optimize the dynamic range of a fluorescent live-cell reporter for PKA kinase activity and successfully produced a new PKA-KTR with high sensitivity and specificity. KTRs were chosen as an application because of their general interest to the signaling community, with ~1,000 constructs that have been distributed around the world so far and ongoing requests for KTRs that currently don't exist, and also because to date, no publication or previously described method has demonstrated the type of screen that is required to generate a new KTR—or by extension, a broad range of other live-cell reporters that measure dynamic behaviors in individual cells.

Our approach enables the quantification of dynamic phenotypes at subcellular resolution (in this case, KTR shuttling between the nucleus and cytoplasm over time). Because we could perturb the same cells at multiple time points and record the responses, we were able to confirm the specificity of the KTR. In particular, the use of forskolin to increase cAMP levels affects a number of signaling pathways;thus, the observed dynamic range could potentially be achieved by kinases other than PKA. Adding the PKA inhibitor H89 to the same cells helped confirm the specificity of the cellular response. Although the H89 inhibitor is imperfect in terms of specificity (and thus further validation was required in this case), the ability to record the response of the same cells for multiple perturbations is an important advantage of our approach.

Sorting-based approaches could in principle also have been applied to the optimization of PKA-KTR. For example, while standard flow cytometry methods cannot follow single cells over time, more recent approaches enable dynamic measurements via photoactivatable proteins followed by standard fluorescence-activated sorting (Kanfer et al., 2021; Lee et al., 2020; Yan et al., 2021) or robotic isolation that directly captures cells for sequencing (Laks et al., 2019; Piatkevich et al., 2018). By contrast, a major advantage of *in situ* genotyping, as described in nucleic acid-based approaches (Feldman et al., 2019; Wang et al., 2019) and also in this study, is that no cellular phenotypes need to be lost: both cells with and without "hits" can be imaged and analyzed together (Lawson and Elf, 2021). This advantage is particularly compatible with recent advances in machine learning (Bedbrook et al., 2019; Chandrasekaran et al., 2021). We focused only on the most promising hits from our study, largely because our library was overrepresented by almost exclusively negative hits. However, in future work and in particular where more variation in screen outcomes is observed, one could imagine comparing all of the sequences with all of the resulting dynamics associated with KTR (or other protein of interest) and using deep learning to derive the underlying relationships between them.

Compared with other *in situ* genotyping methods, the multiplexing capacity of our approach is still relatively limited by, and dependent on, the number of usable antibodies. However, our ability to expand upon multiplexing by spatial localization dramatically increases the method's potential throughput. For example, if four epitope locations could be resolved by imaging, then our method would be able to multiplex >2 billion combinations with our design or >60,000 with just 8 epitopes. In the latter case, the *in situ* genotyping process would be greatly simplified, and with a lower number of epitopes required, the most effective set of epitopes could be chosen to optimize the screen's fidelity. An ideal localization tag needs to

be stable to achieve a high expression of epitopes. One of the promising molecular approaches to achieve more individually resolvable localization sites is the use of self-clustering proteins. Such proteins have been used to express multiple FRET reporters in different intracellular locations (Linghu et al., 2020).

Recent advances in immunostaining may expand the multiplexing capacity of our approach even further by increasing the number of detectable protein targets in our barcodes from 3 to 4 colors to the scale of tens to hundreds. These include cyclic immunofluorescence combined with chemical bleaching (Gut et al., 2018; Lin et al., 2015), mass cytometry imaging with metal-isotope labeled antibodies (Angelo et al., 2014; Giesen et al., 2014), and the use of nucleotide-conjugated antibodies followed by sequencing or hybridization (Goltsev et al., 2018; Kennedy-Darling et al., 2021). All of these approaches should be readily applicable to our method, further simplifying the *in situ* genotyping step.

Most of these antibody-based methods are well established with tissue imaging, which broadens the scope of applications to which our approach could be applied. This possibility, combined with the fact that our approach can easily be extended to any genetic perturbation screen, including CRISPR gRNA and other mutagenesis screens, leads us to anticipate a variety of applications for the method beyond what has been described here.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Generation of cell lines
  - Fluorophore conjugation
  - Immunofluorescence and Live-cell microscopy
  - Image processing
  - Spectral unmixing
  - Segmentation and tracking
  - An estimation of fraction of unique pairs
  - A likelihood approach to map epitope barcodes
  - Matthews correlation coefficient and UMAP visualization
  - Library sequencing
  - Library preparation
  - Library plasmid design and sequencing strategy
  - Epitope barcode oligo design
  - Design of reporter mutagenesis
  - Cross-talk between nuclear and cytoplasmic signals
  - On the rate of template switching
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cels.2022.02.006.

ACKNOWLEDGMENTS

We thank members of the Covert lab for discussions and for providing feedback on the manuscript. We gratefully acknowledge funding from The Paul G. Allen Frontiers Group Allen Discovery Center grant to M.W.C. and the Nakajima Foundation Scholarship to T.K.

AUTHOR CONTRIBUTIONS

Conceptualization, T.K.; methodology, T.K. and K.L.; software, T.K.; validation, T.K.; investigation, T.K.; writing-original draft, T.K. and M.W.C.; writing-review & editing, T.K. and M.W.C.; visualization, T.K.; supervision, M.W.C.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 1, 2021
Revised: September 27, 2021
Accepted: February 25, 2022
Published: March 21, 2022

REFERENCES

Adamson, B., Norman, T.M., Jost, M., and Weissman, J.S. (2018). Approaches to maximize sgRNA-barcode coupling in Perturb-seq screens. Preprint at biorxiv https://www.biorxiv.org/content/10.1101/298349v1.

Angelo, M., Bendall, S.C., Finck, R., Hale, M.B., Hitzman, C., Borowsky, A.D., Levenson, R.M., Lowe, J.B., Liu, S.D., Zhao, S., et al. (2014). Multiplexed ion beam imaging of human breast tumors. Nat. Med. *20*, 436–442.

Bedbrook, C.N., Yang, K.K., Robinson, J.E., Mackey, E.D., Gradinaru, V., and Arnold, F.H. (2019). Machine learning-guided channelrhodopsin engineering enables minimally-invasive optogenetics. Nat. Methods *16*, 1176–1184.

Canver, M.C., Haeussler, M., Bauer, D.E., Orkin, S.H., Sanjana, N.E., Shalem, O., Yuan, G.-C., Zhang, F., Concordet, J.-P., and Pinello, L. (2018). Integrated design, execution, and analysis of arrayed and pooled CRISPR genome-editing experiments. Nat. Protoc. *13*, 946–986.

Chan, T.F., and Vese, L.A. (2001). Active contours without edges. IEEE Trans. Image Process. *10*, 266–277.

Chandrasekaran, S.N., Ceulemans, H., Boyd, J.D., and Carpenter, A.E. (2021). Image-based profiling for drug discovery: due for a machine-learning upgrade? Nat. Rev. Drug Discov. *20*, 145–159.

De Simone, A., Evanitsky, M.N., Hayden, L., Cox, B.D., Wang, J., Tornini, V.A., Ou, J., Chao, A., Poss, K.D., and Di Talia, S. (2021). Control of osteoblast regeneration by a train of Erk activity waves. Nature *590*, 129–133.

Edelstein, A.D., Tsuchida, M.A., Amodaj, N., Pinkard, H., Vale, R.D., and Stuurman, N. (2014). Advanced methods of microscope control using μManager software. J. Biol. Methods *1*.

Emanuel, G., Moffitt, J.R., and Zhuang, X. (2017). High-throughput, image-based screening of pooled genetic-variant libraries. Nat. Methods *14*, 1159–1162.

Feldman, D., Singh, A., Garrity, A.J., and Blainey, P.C. (2018). Lentiviral co-packaging mitigates the effects of intermolecular recombination and multiple integrations in pooled genetic screens. Preprint at biorxiv https://www.biorxiv.org/content/10.1101/262121v1.

Feldman, D., Singh, A., Schmid-Burgk, J.L., Carlson, R.J., Mezger, A., Garrity, A.J., Zhang, F., and Blainey, P.C. (2019). Optical pooled screens in human cells. Cell *179*, 787–799.e17.

Giesen, C., Wang, H.A.O., Schapiro, D., Zivanovic, N., Jacobs, A., Hattendorf, B., Schüffler, P.J., Grolimund, D., Buhmann, J.M., Brandt, S., et al. (2014). Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. Nat. Methods *11*, 417–422.

Goltsev, Y., Samusik, N., Kennedy-Darling, J., Bhate, S., Hale, M., Vazquez, G., Black, S., and Nolan, G.P. (2018). Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. Cell *174*, 968–981.e15.

Gut, G., Herrmann, M.D., and Pelkmans, L. (2018). Multiplexed protein maps link subcellular organization to cellular states. Science *361*, eaar7042.

Huff, J., Bergter, A., Birkenbeil, J., Kleppe, I., Engelmann, R., and Krzic, U. (2017). The new 2D Superresolution mode for ZEISS Airyscan. Nat. Methods *14*, 1223.

Hung, Y.P., Teragawa, C., Kosaisawe, N., Gillies, T.E., Pargett, M., Minguet, M., Distor, K., Rocha-Gregg, B.L., Coloff, J.L., Keibler, M.A., et al. (2017). Akt regulation of glycolysis mediates bioenergetic stability in epithelial cells. Elife *6*, e27293.

Jaqaman, K., Loerke, D., Mettlen, M., Kuwata, H., Grinstein, S., Schmid, S.L., and Danuser, G. (2008). Robust single-particle tracking in live-cell time-lapse sequences. Nat. Methods *5*, 695–702.

Joung, J., Konermann, S., Gootenberg, J.S., Abudayyeh, O.O., Platt, R.J., Brigham, M.D., Sanjana, N.E., and Zhang, F. (2017). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. Nat. Protoc. *12*, 828–863.

Kanfer, G., Sarraf, S.A., Maman, Y., Baldwin, H., Dominguez-Martin, E., Johnson, K.R., Ward, M.E., Kampmann, M., Lippincott-Schwartz, J., and Youle, R.J. (2021). Image-based pooled whole-genome CRISPRi screening for subcellular phenotypes. J. Cell Biol. *220*, e202006180.

Kennedy-Darling, J., Bhate, S.S., Hickey, J.W., Black, S., Barlow, G.L., Vazquez, G., Venkataraaman, V.G., Samusik, N., Goltsev, Y., Schürch, C.M., and Nolan, G.P. (2021). Highly multiplexed tissue imaging using repeated oligonucleotide exchange reaction. Eur. J. Immunol. *51*, 1262–1277.

Kosugi, S., Hasebe, M., Matsumura, N., Takashima, H., Miyamoto-Sato, E., Tomita, M., and Yanagawa, H. (2009). Six classes of nuclear localization signals specific to different binding grooves of importin α. J. Biol. Chem. *284*, 478–485.

Kosugi, S., Hasebe, M., Tomita, M., and Yanagawa, H. (2008). Nuclear export signal consensus sequences defined using a localization-based yeast selection system. Traffic *9*, 2053–2062.

Kraus, B., Ziegler, M., and Wolff, H. (2007). Linear fluorescence unmixing in cell biological research. Mod. Res. Educ. Top. Microsc. *2*, 863–873.

Kudo, T., Jeknić, S., Macklin, D.N., Akhter, S., Hughey, J.J., Regot, S., and Covert, M.W. (2018). Live-cell measurements of kinase activity in single cells using translocation reporters. Nat. Protoc. *13*, 155–169.

Laks, E., McPherson, A., Zahn, H., Lai, D., Steif, A., Brimhall, J., Biele, J., Wang, B., Masud, T., Ting, J., et al. (2019). Clonal decomposition and DNA replication states defined by scaled single-cell genome sequencing. Cell *179*, 1207–1221.e22.

Landini, G. (2006–2014). How to correct background illumination in bright-field microscopy. https://blog.bham.ac.uk/intellimic/background-illumination-correction.

Lawson, M., and Elf, J. (2021). Imaging-based screens of pool-synthesized cell libraries. Nat. Methods *18*, 358–365.

Lee, J., Liu, Z., Suzuki, P.H., Ahrens, J.F., Lai, S., Lu, X., Guan, S., and St-Pierre, F. (2020). Versatile phenotype-activated cell sorting. Sci. Adv. *6*, eabb7438.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv *1303*, 3997v1 [q-GN].

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R., 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, C.H., and Lee, C.K. (1993). Minimum cross entropy thresholding. Pattern Recognit. *26*, 617–625.

Lin, J.-R., Fallahi-Sichani, M., and Sorger, P.K. (2015). Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method. Nat. Commun. *6*, 8390.

Linghu, C., Johnson, S., L, Valdes, P., A, Shemesh, O., A, Park, W.M., Park, D., Piatkevich, K., D, Wassie, A., T, Liu, Y., An, B., et al. (2020). Spatial Multiplexing of Fluorescent Reporters for Imaging Signaling Network Dynamics. Cell *183*, 1682–1698.

Liu, Z., Chen, O., Wall, J.B.J., Zheng, M., Zhou, Y., Wang, L., Vaseghi, H.R., Qian, L., and Liu, J. (2017). Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. Sci. Rep. *7*, 2193.

London, E., Bloyd, M., and Stratakis, C.A. (2020). PKA functions in metabolism and resistance to obesity: lessons from mouse and human studies. J. Endocrinol. *246*, R51–R64.

Maryu, G., Matsuda, M., and Aoki, K. (2016). Multiplexed fluorescence imaging of ERK and Akt activities and cell-cycle progression. Cell Struct. Funct. *41*, 81–92.

Matthews, B.W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim. Biophys. Acta *405*, 442–451.

McArthur, K., Whitehead, L.W., Heddleston, J.M., Li, L., Padman, B.S., Oorschot, V., Geoghegan, N.D., Chappaz, S., Davidson, S., San Chin, H., et al. (2018). BAK/BAX macropores facilitate mitochondrial herniation and mtDNA efflux during apoptosis. Science *359*, eaao6047.

McInnes, L., Healy, J., and Astels, S. (2017). hdbscan: Hierarchical density based clustering. J. Open Source Softw. *2*, 205.

McInnes, L., Healy, J., Saul, N., and Großberger, L. (2018). UMAP: uniform manifold approximation and projection. J. Open Source Software *3*, 861.

Mena, M.A., and Daugherty, P.S. (2005). Automated design of degenerate codon libraries. Protein Eng. Des. Sel. *18*, 559–561.

Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M., and Van Valen, D. (2019). Deep learning for cellular image analysis. Nat. Methods *16*, 1233–1246.

Moore, M.J., Adams, J.A., and Taylor, S.S. (2003). Structural basis for peptide binding in protein kinase A. Role of glutamic acid 203 and tyrosine 204 in the peptide-positioning loop. J. Biol. Chem. *278*, 10613–10618.

Olsen, S.R., and Uhler, M.D. (1991). Inhibition of protein kinase-A by overexpression of the cloned human protein kinase inhibitor. Mol. Endocrinol. *5*, 1246–1256.

Piatkevich, K.D., Jung, E.E., Straub, C., Linghu, C., Park, D., Suk, H.-J., Hochbaum, D.R., Goodwin, D., Pnevmatikakis, E., Pak, N., et al. (2018). A robotic multidimensional directed evolution approach applied to fluorescent voltage reporters. Nat. Chem. Biol. *14*, 352–360.

Pluim, J.P.W., Maintz, J.B.A., and Viergever, M.A. (2003). Mutual-information-based registration of medical images: a survey. IEEE Trans. Med. Imaging *22*, 986–1004.

Pokrass, M.J., Ryan, K.A., Xin, T., Pielstick, B., Timp, W., Greco, V., and Regot, S. (2020). Cell-cycle-dependent ERK signaling dynamics direct fate specification in the mammalian preimplantation embryo. Dev. Cell *55*, 328–340.e5.

Regot, S., Hughey, J.J., Bajar, B.T., Carrasco, S., and Covert, M.W. (2014). High-sensitivity measurements of multiple kinase activities in live single cells. Cell *157*, 1724–1734.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: convolutional networks for biomedical image segmentation. Medical Image Computing and Computer Assisted Intervention MICCAI 2015, Springer *9351*, 234–241.

Sack, L.M., Davoli, T., Xu, Q., Li, M.Z., and Elledge, S.J. (2016). Sources of error in mammalian genetic screens. G3 (Bethesda) *6*, 2781–2790.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.

Spencer, S.L., Cappell, S.D., Tsai, F.-C., Overton, K.W., Wang, C.L., and Meyer, T. (2013). The proliferation-quiescence decision is controlled by a bifurcation in CDK2 activity at mitotic exit. Cell *155*, 369–383.

Tareen, A., and Kinney, J.B. (2020). Logomaker: beautiful sequence logos in Python. Bioinformatics *36*, 2272–2274.

Taylor, S.S., and Radzio-Andzelm, E. (1994). Three protein kinase structures define a common motif. Structure *2*, 345–355.

Tornøe, J., Kusk, P., Johansen, T.E., and Jensen, P.R. (2002). Generation of a synthetic mammalian promoter library by modification of sequences spacing transcription factor binding sites. Gene *297*, 21–32.

Turnham, R.E., and Scott, J.D. (2016). Protein kinase A catalytic subunit isoform PRKACA; history, function and physiology. Gene *577*, 101–108.

Ubersax, J.A., and Ferrell, J.E., Jr. (2007). Mechanisms of specificity in protein phosphorylation. Nat. Rev. Mol. Cell Biol. *8*, 530–541.

Viola, P., and Wells, W.M. (1995). Alignment by maximization of mutual information. Proceedings of the IEEE International Conference on Computer Vision.

Wählby, C., Erlandsson, F., Bengtsson, E., and Zetterberg, A. (2002). Sequential immunofluorescence staining and image analysis for detection of large numbers of antigens in individual cell nuclei. Cytometry *47*, 32–41.

Wang, C., Lu, T., Emanuel, G., Babcock, H.P., and Zhuang, X. (2019). Imaging-based pooled CRISPR screening reveals regulators of lncRNA localization. Proc. Natl. Acad. Sci. USA *116*, 10842–10851.

Wroblewska, A., Dhainaut, M., Ben-Zvi, B., Rose, S.A., Park, E.S., Amir, E.-A.D., Bektesevic, A., Baccarini, A., Merad, M., Rahman, A.H., and Brown, B.D. (2018). Protein barcodes enable high-dimensional single-cell CRISPR screens. Cell *175*, 1141–1155.e16.

Yan, K., Gao, L.-N., Cui, Y.-L., Zhang, Y., and Zhou, X. (2016). The cyclic AMP signaling pathway: exploring targets for successful drug discovery (Review). Mol. Med. Rep. *13*, 3715–3723.

Yan, X., Stuurman, N., Ribeiro, S.A., Tanenbaum, M.E., Horlbeck, M.A., Liem, C.R., Jost, M., Weissman, J.S., and Vale, R.D. (2021). High-content imaging-based pooled CRISPR screens in mammalian cells. J. Cell Biol. *220*, e202008158.

Yang, H.W., Cappell, S.D., Jaimovich, A., Liu, C., Chung, M., Daigh, L.H., Pack, L.R., Fan, Y., Regot, S., Covert, M., et al. (2020). Stress-mediated exit to quiescence restricted by increasing persistence in CDK4/6 activation. Elife *9*, e44571.

Zikry, T.M., Kedziora, K.M., Kosorok, M.R., and Purvis, J.E. (2019). In and out of the nucleus: CNN based segmentation of cell nuclei from images of a translocating sensor. In Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (Learning) (Association for Computing Machinery), pp. 1–4.

Zimmermann, T. (2005). Spectral Imaging and linear unmixing in light microscopy. Adv. Biochem. Eng. Biotechnol. *95*, 245–265.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti Universal, rabbit pAb	GenScript	A00680-40; RRID:AB_915564
anti NWSHPQFEK Tag, mouse mAb	GenScript	A01732-100; RRID:AB_2622218
anti SRT-tag, rabbit pAb	GenScript	A00725-40; RRID:AB_915560
anti tag-100-tag, rabbit pAb	GenScript	A00677-40; RRID:AB_915550
anti E-tag antibody	abcam	ab95868; RRID:AB_10679906
Alexa Fluor 594 anti DYKDDDDK tag, rat mAb	BioLegend	637313
Myc Tag (9B11), mouse mAb (Alexa Fluor 488 Conjugate)	Cell Signaling Technology	2279S; RRID:AB_2151849
anti T7-tag, goat pAb	abcam	ab9138; RRID:AB_307038
Anti AU5, goat pAb	Novus Biologicals	NB600-461; RRID:AB_10002033
Anti V5 tag, mouse mAb	Genscript	A01724-40; RRID:AB_2622216
Anti HA tag, mouse mAb	Genscript	A01244-100; RRID:AB_1289306
Anti Glu-Glu tag, rabbit pAb	Millipore Sigma	AB3788; RRID:AB_91589
Anti VSV-G-tag, rabbit pAb	Genscript	A00199-40; RRID:AB_915517
Anti E2 tag, mouse mAb	abcam	ab977; RRID:AB_296610
Anti OLLAS-tag, rabbit pAb	Genscript	A01658-40; RRID:AB_2622186
Anti His-tag, mouse mAb	Cell Signaling Technology	12698S; RRID:AB_2744546
Anti Protein C-tag, rabbit pAb	Genscript	A01774-100; RRID:AB_2744686
Anti MAT tag, mouse mAb	Millipore Sigma	M6693-200UG; RRID:AB_1078089
Bacterial and virus strains		
ElectroMAX Stbl4 Competent Cells	ThermoFisher Scientific	11635018
One Shot Stbl3 Chemically Competent E. coli	ThermoFisher Scientific	C737303
ElectroMAX DH5α-E Competent Cells	ThermoFisher Scientific	12034013
Chemicals, peptides, and recombinant proteins		
Alexa Fluor 488 NHS Ester (A488)	Thermo Fisher Scientific	A20000
Alexa Fluor 532 NHS Ester (A532)	Thermo Fisher Scientific	A20001
Alexa Fluor 405 NHS Ester (A405)	Thermo Fisher Scientific	A30000
CF430 CF Dye Succinimidyl Ester	Biotium	92117
CF680R Dye Succinimidyl Esters	Biotium	92107
CF594 CF Dye Succinimidyl Esters	Biotium	92132
Puromycin	Invivogen	ant-pr-1
Blasticidin	Invivogen	ant-bl-1
Forskolin	Sigma-Aldrich	F6886
IBMX	Cayman Chemical	13347
8-br-cAMP	Tocris Bioscience	1140
8-br-cGMP	Cayman Chemical	14431
H89	Sigma-Aldrich	B1427
TC-S 7001	Cayman Chemical	19129
GDC-0068	Cayman Chemical	18412
LY2584702	Cayman Chemical	15320
EGF	Sigma-Aldrich	SRP3196
Insulin	Sigma-Aldrich	I-1882
Fibronectin	Sigma-Aldrich	F0895
FluoroBrite DMEM	Life Technologies	A1896702

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
Amicon Pro Purification System (50 kDa cutoff)	Amicon	ACS505012
QIAquick Gel Extraction Kit	Qiagen	28704
DNA Clean & Concentrator-5	Zymo research	D4004
TaKaRa DNA Ligation Kit Long	Clontech	6024
PrimeSTAR MAX DNA Polymerase	Clontech	R045B
PureLink HiPure Plasmid Midiprep Kit	ThermoFisher Scientific	K210004
QIAprep Spin Miniprep Kit	Qiagen	27104
Experimental models: Cell lines		
Human: Lenti-X 293T cells	Clontech	632180
Human: HeLa S3 cells	ATCC	RRID:CVCL_0058
Mouse: 3T3 cells	ATCC	RRID:CVCL_0594
Oligonucleotides		
oPKAmut (Template for PKA mutagenesis, degeneracy follows IUPAC code) gtcagcatCtgTgattcac tagttaaggttaagcgaNNKNN KWCTDYKgtcaacgaagatgaagc accctccARRcgaaagNNKW CTDYKNNKNTNagcTCGaggt tagaAcgtctgacc	This paper	N/A
oamppool0 (for amplifying the annealed pool) GAGTGTGATCAACTAGTGGCGCG	This paper	N/A
oamppool1 (for amplifying the annealed pool) CCATACCGGTCTCGAGGCGATC	This paper	N/A
oN2K_seq0 (to prep amplicon) CAAGCAGAAGACGGCATACG AGATTAAGGCGAGTCTCG TGGGCTCGGAGATGTGT ATAAGAGACAGNNNNNNNN GCTCCTCGCCCTTGCTCAC	This paper	N/A
oN2K_seq1 (to prep amplicon) AATGATACGGCGACCACCGAGATCTACACTT ATGCGATCGTCGGCAGCGTCAGATGTG TATAAGAGACAGNNNNNN NNGTGTCCGAGGGTACTAAGGCCATCAC	This paper	N/A
oB2N_seq0 (to prep amplicon) CAAGCAGAAGACGGCATACGAGATTC GACGTCGTCTCGTGGGCTCGGAGATG TGTATAAGAGACAGNNNNNNNNTCGG AACGGCGCACGCTGATTGG	This paper	N/A
oB2N_seq1 (to prep amplicon, lower case corresponds to Nextera i5 index) AATGATACGGCGACCA CCGAGATCTACACtatcctctTCGTCG GCAGCGTCAGATGTGTATAAG	This paper	N/A
o9mer0 (t1203, to include NNNs to PKAintermediate) TCGGTACCC CATACTCTTCC	This paper	N/A
o9mer1 (t1204, to include NNNs to PKAintermediate) NNNNNNNNNGGGCGGAGTATGGGC	This paper	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
oPKA_0 (t925, to amplify PKA degenerated oligonucleotides) GTCAGCATCTGTGATTCACTAGTTAAGG	This paper	N/A
oPKA_1 (t926, to amplify PKA degenerated oligonucleotides) GTCAGACGTTCTAACCTCGAGC	This paper	N/A
Recombinant DNA		
In-frame kanamycin backbone	This paper	Addgene Plasmid 177149
In-frame ampicillin PKAKTR backbone	This paper	Addgene Plasmid 177197
pLenti EPICode screen backbone	This paper	Addgene Plasmid 177198
pENTR PKAKTRn-Clover	This paper	Addgene Plasmid 177199
pLenti PKAKTRn-Clover	This paper	Addgene Plasmid 177200
pMD2.G	Didier Trono (https://tronolab.epfl.ch)	Addgene #12259
psPAX2	Didier Trono (https://tronolab.epfl.ch)	Addgene #12260
Software and algorithms		
CellTK	(Kudo et al., 2018)	https://github.com/braysia/CellTK (https://doi.org/10.5281/zenodo.5940197)
Micro-manager	(Edelstein et al., 2014)	https://micro-manager.org/
ImageJ	(Schneider et al., 2012)	https://imagej.nih.gov/ij/
Samtools	(Li et al., 2009)	http://samtools.sourceforge.net/
Burrows-Wheeler Aligner	(Li, 2013)	http://bio-bwa.sourceforge.net/
Pysam	Andreas Heger, Kevin Jacobs and Pysam contributors	https://github.com/pysam-developers/pysam
UMAP	(McInnes et al., 2018)	https://github.com/lmcinnes/umap
HDBSCAN	(McInnes et al., 2017)	https://github.com/scikit-learn-contrib/hdbscan
Logomaker	(Tareen and Kinney, 2020)	https://github.com/jbkinney/logomaker
Deposited data		
Sequencing and image analysis	This paper	https://github.com/braysia/epitope_barcoding (Database: https://doi.org/10.5281/zenodo.5904975)
CellUNet	This paper	https://github.com/braysia/cellunet (Database: https://doi.org/10.5281/zenodo.5935232)
Other		
Nunc Microwell 96-Well Optical-Bottom Plates with Coverglass Base	Thermo Fisher Scientific	164588

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Markus Covert (mcovert@stanford.edu).

Materials availability

Plasmids generated in this study have been deposited at Addgene.

Data and code availability

- All data reported in this paper will be shared by the lead contact upon request.
- Code and sample data are deposited at Github (https://github.com/braysia/epitope_barcoding; https://doi.org/10.5281/zenodo.5940197), and information for analysis can be found in this paper's supplemental information.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

HeLa S3 (RRID:CVCL_0058), 3T3 (RRID:CVCL_0594) and Lenti-X 293T cells were maintained in DMEM supplemented with 10% FBS (Omega Scientific), 2mM L-Glutamine (Life Technologies), and 1X Penicillin/Streptomycin (Life Technologies) at 37°C, 5% CO2.

METHOD DETAILS

Generation of cell lines

Lentivirus was generated with lipofectamine transfection of second-generation viral packaging plasmids and lentiviral construct into Lenti-X 293T cells. To infect cells at low MOI, a non-integrating lentiviral vector was mixed with the construct of interest at 9:1 ratio (Feldman et al., 2018). The volume of the supernatant used for infection was adjusted to make MOI below 10%, which was confirmed by flow cytometry or fluorescent microscopy. 24 h after infection, cells were selected using puromycin (2 μg/ml) for 24 h or blasticidin (1.5 μg/ml) for 96 h.

Fluorophore conjugation

To conjugate fluorophores with a small amount of antibody, we developed the conjugation protocol using the Amicon Pro Purification System. Briefly, 50 μl of antibody was loaded to the Amicon Ultra 0.5 device (50 kDa cutoff). If needed, the remaining antibody was added from the top to the exchange device. The device was then attached to the rest of the system. The tube was loaded with 1.5 ml PBS and centrifuged for 10 min at 4,500 x g for PBS wash. 100 μM succinimidyl ester of fluorophore dyes in 1 ml of 50 mM phosphate buffer (pH 8.5) was then loaded and spun for 10 min at 4,500 x g. After the spin, 50 μl of the phosphate buffer was added from the top to prevent drying of the frit. The tube was then left at room temperature in the dark for 1 h for the conjugation reaction. After the reaction, the filter was washed with 1.5 mL PBS three times. To prevent contamination, the tube was further washed two times with PBS containing 0.04 % sodium azide. Finally, the conjugated antibody was recovered by a reverse spin and stored at 4 °C. After each spin, we reattached the filter and centrifuge for an extra 5 min when we visually confirmed the remaining solution in the exchange device.

Immunofluorescence and Live-cell microscopy

Cells were plated at 10,000 cells per well in a fibronectin-coated 96-well plate and left to adhere overnight. The next day, cells were fixed with 4% paraformaldehyde in PBS for 15 min followed by washing three times with PBS. Permeabilization was carried out with 0.1% Triton X-100 in PBS for 15 min, followed by 2 h blocking with 10% donkey serum (Jackson Laboratory). Cells were then incubated with the primarily conjugated antibody for 30 min, followed by three-time washes with PBS. If needed, cells were then stained with 50 ng/mL Hoechst 33342 for 15 min followed by PBS wash. To chemically bleach fluorescent dyes, the sample was treated with 3% H2O2 and 20mM HCl in PBS for 30 min followed by PBS wash three times. As the bleaching step often caused air bubbles in the sample, we vacuumed the plate in a desiccator for 5 min if bubbles were visually confirmed.

Imaging was conducted using a Nikon Eclipse Ti fluorescence microscope with 20x/0.75 numerical aperture objective. 3x3 binning images were acquired by Andor Neo 5.5 sCMOS camera. We chose 3x3 binning to facilitate analysis time but less binning may be ideal for certain phenotypes or cell types.The filter sets corresponding to A405, CF430, A488, A532, CF594 and CF680 were 350/50-455/50 (emission-excitation), 430/24-470/24, 490/20-525/36, 540/15-579/34, 572/35-632/60 and 645/30-705/72. To perform the pixel-wise subtraction later, images were acquired before and after each cycle of staining.

For live-cell microscopy, cells were plated at 10,000 cells per well in a 96-well plate a day prior to imaging. Six hours prior to imaging, the growth media was replaced by the imaging media (Fluorobrite + 10 mM HEPES pH=7.0 (Sigma) + 1% FBS + 2 mM L-glutamine). Time-lapse microscopy was performed with the Nikon Eclipse Ti fluorescence microscope equipped with temperature (37°C) and environmental control (5% CO2) and automated by Micro-manager. Our setup allowed us to image 384 positions using a 5 min interval. Each live-cell experiment took ~2 hours, screening 6-8 sublibraries with 3 to 4 replicate wells. We repeated the experiments to screen approximately 16,000 variants (32 sub-libraries). For the screening, cells were immediately fixed after imaging for subsequent immunofluorescence.

Image processing

Imaging data were analyzed using CellTK with a custom pipeline. Illumination correction and background subtraction were first applied to all the images (Landini, 2006). For the image registration, images between cycles were aligned by computing the jitters which maximize the mutual information between the images. Due to the risk of local optimum, we repeated the optimization processes from a few different initial positions and adopted the jitters that produced the highest mutual information.

The pixel-wise subtraction was then applied between cycles. To account for signal reduction due to the photobleaching and washing, we estimated the contribution of photobleaching by calculating how much signal decreased using a control sample where cells were stained with all the six fluorophores in the first cycle. The images in a previous cycle were divided by this weight and subtracted from the images in the current cycle. We then downscaled the subtracted images to half size by bilinear interpolation to reduce the data size and processing time.

The spectral unmixing was then applied to these images. The reference spectra were experimentally obtained in a separate experiment by imaging fluorophores diluted in PBS.

Spectral unmixing

After background subtraction, illumination correction, and pixel-wise subtraction, spectral unmixing was applied to fluorescent images (Kraus et al., 2007; Maryu et al., 2016). For each pixel position, we define the following:

$$S = \begin{bmatrix} w_0^T \\ \vdots \\ w_N^T \end{bmatrix} \cdot F^T = W \cdot F^T$$

where S is a vector of observed intensities from each channel, w_i represents the spectral contribution of fluorophore i to each channel and F represents the abundances of each fluorophore. Assuming a linear relationship between signal intensity and exposure time, we further extended W to allow each experiment to use flexible exposure time by

$$W = W_{\beta_0} \cdot (\beta \cdot 1^T)$$

where W_{β_0} is the reference spectra experimentally obtained under the exposure setting β_0, β represents the relative exposure time to β_0 for each channel, and **1** represents a column vector of ones. Each w_i is normalized such that $w_i/\max(w_i)$. The contributions from each fluorophore were then obtained by solving for F with non-negative least squares.

$$\underset{F}{argmin} \|W \cdot F^T - S\|_2, F \geq 0$$

Segmentation and tracking

Imaging data were analyzed using CellTK with a custom pipeline. When nuclear markers (Hoechst staining or fluorescent protein fused with H2B) were available, they were segmented by adaptive thresholding followed by an adaptation of the Morphological Active Contours without Edges algorithm (Chan and Vese, 2001) to round out the nuclei. The segmented objects were then subjected to watershed separation. Tracking of cells throughout the time course was performed by using the linear assignment problem (LAP) (Jaqaman et al., 2008) along with adaptive watershed separation and the nearest neighbor algorithm for cells that were not detected primarily by LAP. After nuclei were tracked, cytoplasmic masks were created by the dilation of nuclei a set number of pixels. These masks were further refined by eliminating background pixels, where foreground regions were detected by applying adaptive thresholding to fluorescent protein signals or unspecific signals due to antibody staining.

We avoided the usage of a fluorescent nuclear marker for the screen, and thus nuclear regions were segmented based on PKAKTR fluorescence using a deep learning approach. To automate the training data curation, we prepared a separate sample of Hoechst-stained cells transduced with a pooled library and performed the computational segmentation to obtain the ground truth images. The U-Net model was then trained to perform nuclear segmentation using PKAKTR fluorescence (Zikry et al., 2019) (https://github.com/braysia/cellunet). The trained model was finally applied to segment the screening images. Tracking and cytoplasmic segmentation were performed as described above.

The C/N ratio was computed by dividing median cytoplasmic values (C) by median nuclear values (N) for each individual cell. To screen for improved translocation of the KTR, we used a metric that identified cells with consistently high activation and inhibition as a result of our treatment with forskolin and H89, respectively. The imaging time course can be divided into three regions of baseline, activated and inhibited KTR activity (Figure S3H). To account for the variation due to segmentation error, floating objects, or cell division, we chose to use the minimum difference of C/N ratio between the baseline and activated regions, as well as between the activated and inhibited regions. Thus, the activation dynamic range is the difference between the minimum C/N ratio after forskolin stimulation and the maximum C/N ratio before stimulation, and the inhibition dynamic range is the difference between the minimum C/N ratio after forskolin stimulation and the maximum C/N ratio after H89 treatment. Cells that exhibited a minimum dynamic range equal to 0.15 of the C/N ratio for both activation and inhibition were chosen as positive hit candidates. Even with this stringent filter, some of the positive hit candidates may not be true positives - for example, when the nuclear envelope breaks down prior to cell division, or when two cells overlap in the image. As a result, we also visually inspected the (~10-20) positive outcomes that passed the minimum dynamic range filter. The positions at the last time frame from the live-cell assay were then associated with immunofluorescence data by linking cells through the nearest neighbor algorithm.

An estimation of fraction of unique pairs

To account for biases of epitope frequencies in our library, we first measured the probabilities of epitope appearance from library sequencing data (for 36 epitopes in nuclei and cytoplasm combined). For each simulation, we generated a random sample based on these probabilities and calculated the fraction of unique pairs.

The theoretical fraction of unique pairs, without taking these biases into account, was calculated by the following:

$$r(n) = \frac{n \cdot \left(\frac{S-1}{S}\right)^{n-1}}{n}$$

where n is the size of the sub-library and S is the total multiplexing capacity, 46,656.

A likelihood approach to map epitope barcodes

Likelihood approach was used to map immunofluorescence profiles to a possible epitope combination. Median nuclear and cytoplasmic intensities (N and C) for each epitope were extracted from each cell. These values were then converted into percentile ranks on a scale of 0 to 1 relative to the distribution of control samples. When $N < 2C$, we forced the percentile to be 0.5 for nuclear epitope in order to account for the possible cross-talk from TOMM20 to nuclear segmentation. The mean percentile score was then calculated for each set of epitope combinations in a library of interest, where the percentile ranks of the corresponding epitopes and positions were averaged together.

To estimate the accuracy of this mapping, we generated a virtual library of different sizes, in addition to the known epitope barcodes that were transduced. We chose two constructs from the library, TOMM20-NWS-V5-OLLAS-2A-His-V5-E-H2B and TOMM20-FLAG-Glu-T7-2A-AU5-HA-VSV, to generate a minimal sub-library. The combination of epitopes in the virtual library followed our library design, where six epitope candidates were assigned to each epitope cassette. For each permutation step, mean percentile scores were calculated for each epitope combination in the library, and the fraction of cells that showed the highest scores for the transduced constructs were recorded. To account for cells that have no or low expression of epitopes, we filtered the cells with the highest percentile score of less than 0.8 (Figure S3E).

To estimate the contribution from each epitope, we stained the cells expressing TOMM20-FLAG-Myc-VSV-2A-His-HA-OLLAS-H2B and computed the accuracy while forcing each epitope's percentile score to be 0.5 during the permutation step. The misidentification rate refers to 1-accuracy.

Matthews correlation coefficient and UMAP visualization

Given median intensities for each epitope or fluorescent protein, cells were classified as positive or negative by Minimum Cross-Entropy thresholding (Li and Lee, 1993). For a pair of epitopes (or a fluorescent protein and epitope), we computed the Matthews correlation coefficient, where we assumed double positive as true positive (Matthews, 1975).

Immunofluorescence profiles were visualized by UMAP implemented using the UMAP package. The output from UMAP was further clustered using Hierarchical Density-based spatial clustering of applications with noise (HDBSCAN), and the cluster information was then used to group colors of the representative segmentation masks.

Library sequencing

We prepared two amplicons to associate epitope barcodes and KTR variants; one that pairs the epitopes and 9-mer sequences (N9), and the other that pairs the N9 and KTR variants. A sub-library of interest was amplified by 22 cycles of PCR. The PCR products were then purified using Qiagen QIAquick Gel Extraction Kit, followed by further purification using Zymo DNA Clean & Concentrator-5. These amplicons were pooled and sequenced using illumina MiSeq sequencing. Sequences of epitope barcodes and KTR variants were then associated via N9 sequences. The sequences were analyzed by a custom script based on Pysam. See https://github.com/braysia/epitope_barcoding for code and sample data.

Library preparation

To create a lentiviral library, we first made separate intermediate libraries for both the epitope barcodes and PKA-KTR variants (Figure S2A). For cloning of epitope barcodes, an oligonucleotide pool (Agilent, G7220A) was synthesized according to the sequences in Table 1. This oligo pool was extended by 1 cycle of PCR, followed by another 30 cycles of PCR using primers that bound to the ends of the annealed fragments. The products were then purified using high-resolution 2% MetaPhor Agarose gel. The purified product was digested and ligated into pKL787 using AsiSI and AscI with kanamycin selection. For mutagenesis of KTR, the degenerated long oligonucleotide was amplified by PCR. The product was then purified, digested and ligated into pKL770 using SpeI and XhoI. After carbenicillin selection, the library was then amplified with primers containing random nucleotides (N9) and self-ligated. Finally, these two intermediate libraries were ligated into the lentiviral backbone. The epitope barcodes were ligated using AscI and AsiSI, and KTR variants were ligated using KpnI and SbfI.

To create a sublibrary, 1 ng of the library was chemically transformed to Stbl3 cells and plated on an agar plate followed by scraping for DNA preparation. The plating volume was adjusted to produce approximately 500 colonies. The intermediate plasmids and lentiviral plasmids were transformed with ElectroMax DH5α-E and ElectroMAX Stbl4 cells, respectively. The general transformation and DNA preparation procedures followed the previously described protocols (Canver et al., 2018; Joung et al., 2017). For each gel extraction, extraction using Qiagen QIAquick Gel Extraction Kit, followed by further purification using Zymo DNA Clean & Concentrator-5. Plasmids were extracted using PureLink Midiprep kit or QIAprep Spin Miniprep Kit.

Library plasmid design and sequencing strategy

Each construct in our lentiviral pooled library expresses two open reading frames, the epitope barcodes fused to TOMM20 and H2B, and the mutagenized PKA-KTR fused to mClover. To create this library, we first made a separate library for the barcode and PKA-KTR regions and then ligated them together for incorporation into a lentiviral vector.

Cloning of the barcode region is a challenging task, as oligo pool synthesis tends to contain some errors. To reduce the impact of these errors, we took several steps to enrich the products which were the expected size. First, after the oligonucleotide pool was annealed and extended by PCR, we performed additional PCR amplification using primers that bound to the ends of the annealed fragments. These amplified fragments were then purified by size, using a high-resolution agarose gel. We then digested and ligated

these products into an intermediate plasmid, which is designed to fuse an antibiotic resistance gene to the barcode if the constructs are ligated in-frame. Any out-of-frame ligation results in introducing a stop codon before the antibiotic resistance gene, reducing the number of ligation products which contain oligo insertion or deletion errors. The mutagenized PKA-KTR library was also created using this intermediate plasmid approach. See Figure S2 for the overview of cloning.

One major consideration for the final ligation product was that it be compatible with illumina sequencing. This implies two constraints: (1) the read length must be less than 300 bp, and (2) the insert size (of the paired-end reads) must be less than 1 kb. Regarding the read length, we designed the entire barcode region, including the 2A peptide, to be 276 bp length, as described above. One feature of our strategy was to avoid the introduction of extra primer binding sites in the amplicon. We therefore designed the linker sequences between the TOMM20 and barcodes such that they encode a polypeptide (SSAASDVYKRQ) - but also function as an illumina sequencing primer binding site. With regard to the insert size, we chose to use the JeT promoter for its small size (< 190 bp) (Tornøe et al., 2002). Even with these reductions, the insert size from the barcode regions to the PKA-KTR mutations was slightly over 1 kb in our design (1,059 nt without sequencing primers). We thus introduced random 9-mer (N9) into the intermediate plasmid, upstream of the JeT promoter. This enabled us to divide the region into two amplicons, one encoding the barcodes and 9-mer, and the other encoding the same 9-mer together with PKA-KTR variants. We could then computationally assemble our constructs via the 9-mer sequences.

To evaluate the quality of our library, we sequenced the library to determine whether or not there would be a bias in terms of the representation of individual epitopes (Figure S2B). We found that 29 out of the 32 epitopes were present in the library at a frequency between 0.01 and 0.05; the outliers were the cytoplasmic tag100 and nuclear ProC epitopes, which were present at a low frequency, and the cytoplasmic FLAG epitope, which was present at a higher frequency. > 40% of our library successfully carried six epitope combinations (Figure S2C).

Epitope barcode oligo design

Each construct in our barcode library carries two sets of three epitopes, for six epitopes total; one set will localize to the nucleus while the other resides in the mitochondria. There are six candidates for each epitope position in the construct (with the same eighteen epitope candidates for each set), as shown in Table 1. Two amino acids, glycine and serine, were used as a linker between the epitopes, either in GS or SG order. Although the constructs each carry a different combination of epitopes, it was important that they be the same length so that the entire library would be amenable to size selection using gel electrophoresis. This concern, in addition to minimizing the overall construct length for sequencing, were our top priorities in library design. Thus, we designed the construct as a series of three cassettes with sizes of 9, 10 and 12 amino acids, respectively. The epitope candidates were then sorted by their length to determine which cassette they could be incorporated into. For epitopes with length shorter than their cognate cassette, an A or G was added to the sequence. Conversely, the long epitopes, OLLAS-tag and E-tag, are 14 and 13 amino acids long, respectively. To include this tag, we recognized that the first two amino acids of OLLAS-tag were serine and glycine, and the first amino acid of E-tag was glycine; and thus by including the SG-ordered linker the same amino acids could be used for both the epitope and the linker. Finally, although the NWS and Strep epitopes are slightly different in their sequences, they are actually recognized by the same antibody. We therefore recommend that any new library construction consider switching one of these epitopes for another, for example S1-tag, E2-tag or AU1-tag.

For each construct, the two sets of epitopes are separated by the self-cleaving 2A peptide. The total length of the barcode region, including the 2A peptide, is 92 amino acids (276 nt). The current state-of-the-art for oligonucleotide synthesis is that oligos must be of size < 200 nt to maintain sequence fidelity. Thus, we were required to synthesize two sets of oligonucleotides, one carrying the first three epitopes and half of the 2A sequence, and the other carrying the other half of 2A and the latter three epitopes. Finally, the nature of the constructs opens the possibility of epitope sequences binding between the two sets of epitopes in trans. To reduce the likelihood of this binding (and subsequent lentiviral template switching), we used one set of codons for half of the oligos (e.g., the oligos encoding the nuclear epitopes), and another set for the other half (containing the cytoplasmic epitopes). The oligo sequences were available as Table S1.

Design of reporter mutagenesis

In designing our PKA-KTR sequence optimization strategy, one of the most important considerations was the sequence search space. In particular, since our method enables a simultaneous observation for on the order of tens of thousands of mutant behaviors, we wanted the search space to be small enough to be tractable, while still enabling the identification of optimized mutants. Thus, we used rational design to identify the amino acid positions for which mutagenesis was most likely to lead to an improved dynamic range.

First, a number of structural studies have suggested two critical features of PKA kinase substrates: (1) an R at the -3 position relative to a phosphorylation site, (2) the preference for a large hydrophobic residue at the +1 position (Moore et al., 2003). To fulfill the first requirement, we removed P34 from our construct, which shifted an R to the -3 position with respect to the first phosphorylation site. For the second requirement, we noted that a degenerate codon in DNA sequences (DYK) can be used to produce hydrophobic residues, including any of the following: AFILMSTV (Mena and Daugherty, 2005). Accordingly, our constructs carried a DYK codon at the +1 position.

Our next phase of design involved preserving a functional NLS and NES. Specifically, it has been shown that basic amino acid residues are required before and after the 10-12 amino acid core of KTR's bipartite NLS (Kosugi et al., 2009). The original construct already had basic amino acid residues (KR) preceding the core. To ensure that basic amino acids appeared in the construct after the

core, we mutated R48, which is located 12 amino acids after the preceding KR, to allow for either R or K, using the degenerate codon ARR. An RK sequence followed this codon, which we retained to make this area more basic. With regard to the NES sequence, a hydrophobic amino acid residue was required at position 55. Thus, we integrated the codon NTN, encoding a strong hydrophobic residue, into our design. The DYK codon at position 53 (mentioned above) is also thought to contribute to the efficiency of the NES (Kosugi et al., 2008).

We allowed either S or T as the amino acids to be phosphorylated, using the WCT codon, and the positions surrounding the phosphorylation sites (36, 37, 51 and 54) to be any amino acid, using the codon NNK, since amino acids nearby phosphorylation sites are likely to contribute to both kinase specificity and localization efficiency (Regot et al., 2014). In total, the sequence space produced by this design strategy produces more than 400 million possible sequences.

For completeness, we originally suspected that some of the amino acids in our original sequence should be modified based on mutagenesis and binding studies of protein kinase inhibitor (PKI), an enzyme that binds to PKA kinase (Olsen and Uhler, 1991). These studies indicated several amino acids contributing to the binding of affinity of PKI, including F at -11 and Y at -14. We incorporated these mutations into our design by removing 25V and replacing 28C to F in the KTR_{orig}. However, after completing our study it does not appear that these changes had an impact on the output of our final reporter (Figure S3I).

We visualized the overall variability in the library's KTR sequence space and confirmed the mutations were introduced as we designed (Figure S2D). The logo was generated by Logomaker (Tareen and Kinney, 2020).

Cross-talk between nuclear and cytoplasmic signals

TOMM20 signals occasionally overlap with the nuclear region in an image, due to the fact that the images can capture fluorescence from mitochondria above or below the nuclear envelope (Figure S3A). To evaluate the extent of such overlap, we performed immunofluorescence on cells expressing TOMM20-V5-2A-SRT-H2B (V5 and SRT are two epitope tags from our set) and quantified the presence of SRT and V5 signals in both the cytoplasmic and nuclear regions. The SRT signal detected in the cytoplasm was less than 5% of the nuclear signal (Figure S3B, top left). However, the V5 signal detected in nuclei was ~30% of the cytoplasmic signal on average (Figure S3B, bottom left), indicating that overlap of the mitochondrial signal on the nucleus needed to be accounted for in our analysis. To correct for the overlap effect, we noticed that the same epitope tended to have a stronger signal when expressed in the nucleus than in the cytoplasm. We quantified this difference by staining cells expressing TOMM20-V5-2A-V5-H2B and determining the V5 signals in both the cytoplasmic and nuclear regions. In this case, the V5 signal was > 2.5 fold brighter in cell nuclei than in the cytoplasm (Figure S3B, right). Based on these results, we applied a threshold to our nuclear epitope calling procedure, in which the presence of a given nuclear epitope is considered verified only when the nuclear intensity for that epitope is greater than 2-fold of the cytoplasmic intensity. Otherwise, the percentile score for our likelihood approach was forced to be 0.5, reflecting that we did not have confidence in the nuclear presence of that epitope tag.

On the rate of template switching

The template switching of lentivirus has been a common issue for making pooled libraries. To reduce the risk of recombination between variants and barcodes, we adapted a non-integrating lentiviral vector as a decoy strategy as previously described (Adamson et al., 2018; Feldman et al., 2018). These groups reduced the frequency of template switching from 30% to ~7.5% (4-fold reduction using 2kb distance) and from 37% to 1.5% (25-fold using 1.7kb distance), respectively (Adamson et al., 2018; Feldman et al., 2018). The frequency of template switching increases as the distance between the barcodes increases. Our design minimized the fragment size to approximately 1 kb by having the self-cleaving peptide and short promoter. However, Sack et al observed the swap rate to be ~28% for 0.72kb distance (Sack et al., 2016). Thus, we expect our design to have a similar frequency of template switching to groups who did not employ a decoy strategy (~30%). To validate the efficacy of the decoy strategy, we first created two constructs, CMV-Clover-H2B-Ruby and CMV-H2B-miRFP. The constructs were created together in lentivirus, which opens the possibility of recombination between the constructs due to homology caused by the common H2B presence in each. Thus, after transducing these constructs to HeLa cells and imaging for and analyzing the Clover and Ruby signals, we were able to quantify a recombination product, CMV-H2B-Ruby, as the frequency of cells expressing Ruby but not Clover. When using a decoy lentiviral vector (1:1:8 ratio for CMV-Clover-H2B-Ruby, CMV-H2B-miRFP and the decoy lentiviral vector), we observed ~4-fold reduction of the recombined product (Figure S5D), similar to previous observations (Adamson et al., 2018). Based on these results, we estimated the rate of template switching to be less than 10%, which is sufficient for screening. The tradeoff between throughput and the error that comes from recombination requires consideration in experimental design. In our case, the preparation of infected pool replicates with the same sub-library may help resolving this issue. As a random pairing of barcodes and variants led to bottlenecking and limited our throughput, co-synthesizing variants and barcodes may also be an attractive option in the future – especially as synthesis costs continue to fall – to further expand the throughput.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis was carried out in Python and a description is available in the main text, method details and associated figure legends.