8-2024

Genetically-Encoded Optical Barcodes for Single-Cell Analysis

Daniel Pritko
dpritko@clemson.edu

GENETICALLY-ENCODED OPTICAL BARCODES FOR SINGLE-CELL ANALYSIS

A Thesis Presented

To the Graduate School of

Clemson University

In Partial Fulfillment

Of the Requirements for the Degree

Master of Science

Professional Communication

By

Daniel Pritko

August 2024

Accepted By:

Dr. Marc Birtwistle, Committee Chair

Dr. Mark Blenner

Dr. David Karig

Dr. Adam Melvin

Table of Contents

LIST OF FIGURES

Figures

LIST OF TABLES

Tables

Abstract

Cell barcodes are capable of being used to answer many different biological questions. They have been used to track the lineage of cells to identifying the function of a gene. While there are multiple different methods to creating cell barcodes, they are limited in their scalability and application. In a previous publication we propose and computationally prove an optical single-cell barcoding method that bridges fast and scalable readouts with the benefits of genetic encoding. In this approach fluorescent proteins (fps) are combined to form fluorescent barcodes that can then be analyzed using a spectral flow cytometer. Here, we test the experimental viability of this barcoding approach at a small scale. We construct ~150 barcodes in a pooled format and then verify the makeup of the pool to validate the construction method. Spectral flow cytometry is then used to determine whether individual fluorescent proteins can be identified within single-cells, finding that most fluorescent proteins are identifiable. The fp identification results also help to verify the use of spectral flow cytometry as an analysis method capable of analyzing the barcoding approach. Current experiments are being performed to assess barcode identification within single-cells. Through these experiments we aim to show that this barcoding method is able to pair fast and scalable readouts through the use of spectral flow cytometry with genetic encoding.

Introduction

Cell barcoding is a technique used to label single-cells and is fundamental to answering a wide variety of biological questions[1–12]. Applications for cell barcoding include identifying cell type and/or state of a single cell or cell in a population (e.g. understanding the tumor heterogeneity of breast cancer), tracking the lineage of a cell (e.g. to understand cell fate decisions), or identifying genetic perturbations made within a cell (e.g. genetic / genetic interaction screening)[12,13]. One of the main benefits of cell barcodes is the diversity of barcodes that can be created, increasing the number of cells that can be distinctly identified in a single study. For example, one study used cell barcodes, that were constructed of nucleotide sequences, attached to specific gRNA to understand the function of ~9,000 genes in a single experiment[14].

There are two main current methods for constructing cell barcodes, either combining fluorophores (small molecules or fluorescent proteins) or generating sequences of nucleic acids, each having benefits and drawbacks depending on the questions at hand[1–4,6,7,14–24]. Ways of generating fluorescent barcodes include combining fluorescent proteins or small molecule fluorophores at different intensities, and attaching multiple fluorophores to a single antibody[1,3,4,15]. One study using fluorescent cell barcodes combined small molecule fluorophores at different concentrations to generate unique fluorescence intensities for each sample, allowing cells from each sample to be identified by flow cytometry when mixed together[1,3]. While the speed of analysis for fluorescent barcoding (flow cytometry and/or high-content imaging) is beneficial for high-throughput studies, these approaches have limitations. Small molecule fluorophore approaches are useful for understanding the characteristics of a cell or cell population

but cannot be used for studies where the barcode identifies a specific genetic perturbation since they are not able to be genetically encoded [1,3,15]. Fluorescent protein (fp) barcodes can be genetically encoded; however current approaches only have implementation for tracking cell lineage and lack high levels of diversity (~100 different barcodes)[4].

Nucleic acid-based barcodes are random base sequences[14,17–24]. A major benefit of this barcoding technique, aside from genetic-encoding compatibility, is the diversity of barcodes that can be generated. For example, cell barcodes consisting of 12 bases were used to identify ~50,000 gRNA in a study looking at the combinatorial repression effects of transcriptional enhancers[19]. As opposed to using flow cytometry or cell imaging for barcode identification, modern approaches predominantly use single-cell RNA-seq. Single cell RNA-seq enables in-depth analysis of each cell's transcriptome however, it is destructive to the cell, slower, and more expensive when compared to other analysis methods. These properties prevent observation of the cell after barcode identification and hinder large-scale experiments (e.g. genome-wide genetic interaction screens)[1,13].

In a previous paper [25], we proposed and computationally established proof-of-principle for a single-cell barcoding method that bridges fast and scalable fluorescence readouts with the large diversity and perturbation potential offered by nucleic acid genetic encoding. This barcoding method is based on Multiplexing using Spectral Imaging and Combinatorics (MuSIC)[15,25–27], which uses multiple fluorophores to generate probes and/or barcodes with unique spectral emission shapes through the use of Förster resonance energy transfer (FRET)[28]. While implementation of MuSIC is the eventual

goal of this cell barcoding approach, in this paper we expand upon previous computational work to demonstrate experimental viability of this barcoding approach at a small scale. We first clone and verify a library of ~150 barcodes from 18 fluorescent proteins. These individual probes and barcodes are then tested for their deconvolvability and stability of their emission spectral shape. Finally, we plan experiments to test barcode identification performance at the single-cell level. Overall the experiments performed in this paper aim to establish this approach to fluorescent barcoding method as a way of combining fast and scalable readouts with genetic encoding capabilities.

Results

Barcodes in Single Cells

While previous work explored the concept of a barcode and their application to high-dimensional single cell analysis[25], experimental demonstration remained. To answer the question of whether barcodes could be constructed, characterized, and then reliably identified in single cells, we targeted a simple application consisting of two fluorescent protein (fp) barcodes. While this application is a small fraction of the theoretical potential, we reasoned it nevertheless provides a sizable library diversity for some biological applications and would support larger scale efforts if successful.

Specifically, we selected 18 fps spanning UV to IR spectral properties (EBFP2[29], mTagBFP2[30], mT-Sapphire[31], mAmetrine[31], mCerulean3[32], LSSmOrange[33], mBeRFP[34], mTFP1[35], EGFP[36], CyOFP1[37], mClover3[38], mVenus[39], mPapaya[40], mOrange2[41], mRuby3[38], mKate2[42], mCardinal[43], miRFP670[44]). These are called “probes”, in this case

simply individual fps (**Fig. 1A**). A barcode is a combination of two probes (although in future work it could consist of more than two probes). Given 18 fps and 2 probe barcodes, 154 unique barcodes can be generated. It is important to note here we only count combinations, where order does not matter, as opposed to permutations, where order matters. That is because in the targeted application of single-cell analysis of fp expression through fluorescence emission spectral measurements, the order of the barcode would not be distinguishable. Furthermore, we consider combinations without replacement, that is, we do not consider a barcode consisting of two of the same fp. While these may exist in the eventual construction, we suspect inclusion of such barcodes will deteriorate detection reliability (versus eliminating false negatives when one fp is not detected when actually present).

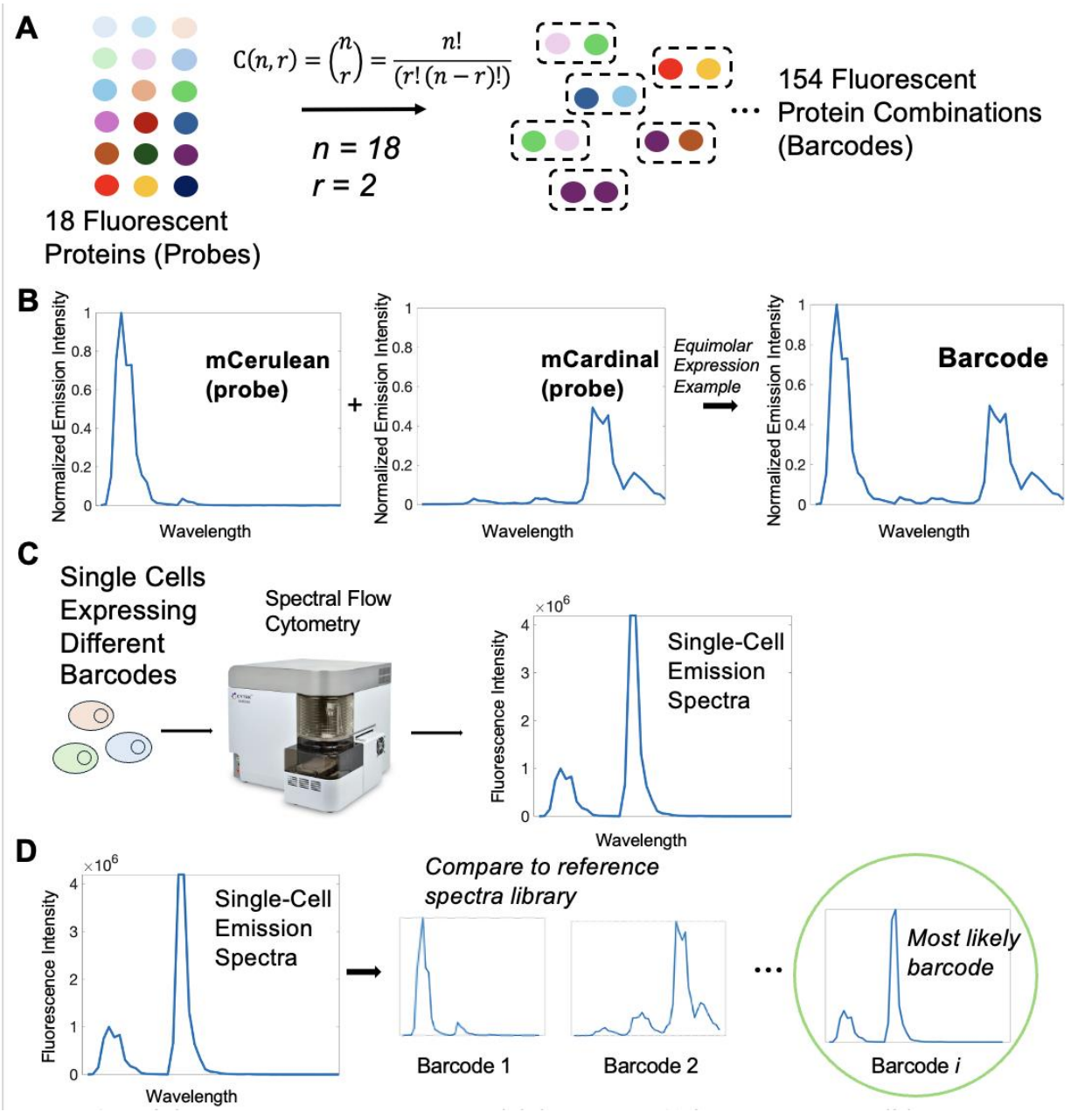

Figure 1. Barcodes and Their Analysis. (**A**) Starting with 18 fluorescent proteins (fp) and combining two fps to make a single barcode, 154 unique barcodes can be generated. Order of fps does not affect the barcode, hence the combination versus permutation formula is used. (**B**) Barcodes have a unique fluorescence spectral emission shape that is representative of the two fluorescent proteins used to construct the barcode. Fluorescent protein emission spectra were generated through spectral flow cytometry of fp positive cells. All values plots were normalized to the maximum value of the barcode spectra (**C**) Spectral flow cytometry outputs the spectral emission shape of each individual barcode-expressing cell that passes through the instrument. (**D**) The emission spectra for each cell can then be compared to the reference spectra library (containing spectra for each barcode) to identify which barcode the cell mostly likely possesses.

As an example, consider a cell expressing an mCerulean and mCardinal barcode (**Fig. 1B**). The individual fp spectral signatures are invariant to whether a barcode is

mCerulean / mCardinal or mCardinal / mCerulean. The combination of these individual spectra gives the barcode spectra, balanced by their relative expression levels (in this example equimolar), which in principle are unique from all the other barcodes. This uniqueness, of course, depends on the spectral emission detector properties. If different single cells in a population are expressing different barcodes, full spectrum flow cytometry can be used to read out single-cell fluorescence emission spectra (**Fig. 1C**). These measured emission spectra can be compared to known references, from which the most likely barcode can be identified (**Fig. 1D**). This conceptual setup forms the basis of the experimental testing of the approach in what follows.

Construction of a Barcode Library

As mentioned above, barcodes are constructed by combining two fluorescent proteins. To accomplish this, we generated two separate pools of probes (**Fig. 2A**). Each pool has a unique backbone, either pReceiver (pR) or pMulti (pM). Every fluorescent protein was cloned into each backbone. To subsequently generate barcodes, we took a pooled construction approach (**Fig. 2B**). Each backbone contains two BbsI cutting sites that were used to combine the probes into a barcode. For efficiency of design, barcode expression is driven by a single promoter in pReceiver, and the barcode elements are separated by 2A peptide sequences. Each of the 18 plasmids in both pools were digested individually, allowing for the purification of each to control relative copy number prior to ligation. The pR and pM pools were combined and a pooled ligation was performed forming a pooled barcode library. Initial attempts at cloning the barcode pool had difficulties, with chemical transformations showing that the negative control (ligation with no pM added to the reaction) produced colonies when no colonies were expected

(**Fig. S1A)**. To determine whether the results were caused by incomplete pR digestion or self-ligation of the pR plasmids, we performed BbsI digestion on the pR pool with no subsequent ligation experiment and chemically transformed the results (**Fig. S1B)**. Since no colonies were formed after chemical transformation, it was determined that self-ligation of the pR plasmids was the cause of the negative control results. To prevent further self-ligation in the generation of the barcode pool, we applied calf intestinal alkaline phosphatase (CIP) to the pR pool after digestion.

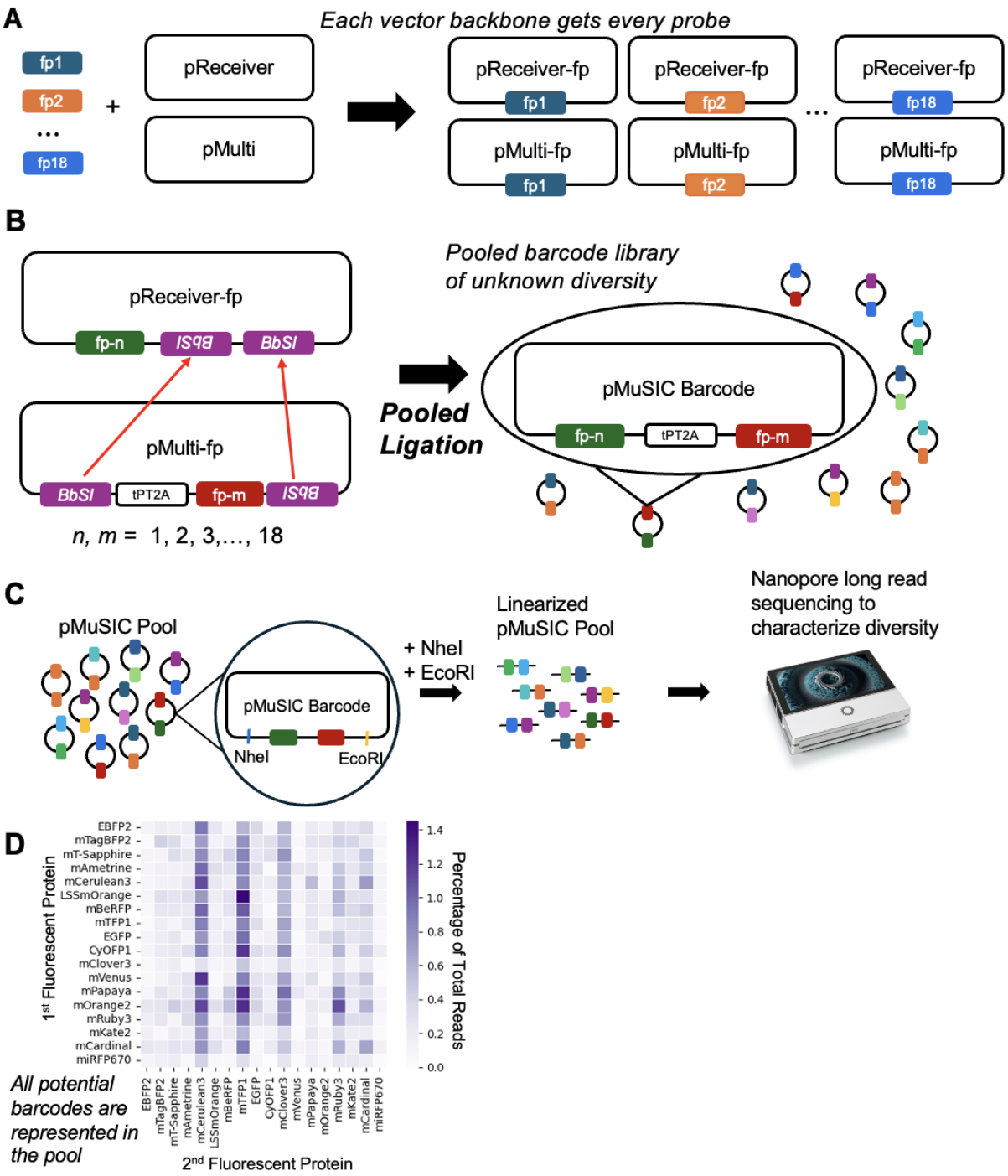

Figure 2. Barcode Library Construction and Validation. (**A**) Each of the 18 fluorescent proteins (fps) were cloned into the pReceiver and pMulti backbones to construct barcode plasmids. (**B**) The fps from pMulti were inserted into the pReceiver backbone by a pooled ligation using *BbsI* sites to generate a barcode library. (**C**) The barcodes were double-digested with *NheI* and *EcoRI* and then sequenced using a nanopore long read sequencer (MK-1C). This avoids library amplification errors by PCR because of high DNA-homology shared among those fps. (**D**) Results of nanopore sequencing of the barcode pool was analyzed and unmixed. The color in the heatmap denotes the percentage of reads attributed to a particular fp in either the 1st or 2nd position of the barcode.

With the barcode pool having been constructed in a pooled ligation, the diversity, i.e. the proportion of each barcode type, was unknown. To this end, sequencing was performed on the pool (**Fig. 2C**). When considering the type of sequencing to use, we had to address two main concerns. One was that we had to know the correct pairing of the fluorescent proteins of each barcode. Instead of trying to pair together reads from short read sequencing, we decided that long read sequencing would be best for this purpose. The next concern was that some fluorescent proteins have high sequence homology, which can lead to complications if PCR is used. Nanopore sequencing became apparent as the best option for our purpose as it would allow us to perform long read sequencing while also eliminating the need for PCR that is required for most other deep sequencing technologies. Linearization of the pool was done using MfeI and NheI to cut the plasmid around the target sequence containing both of the probes. To analyze the sequencing results, we determined which fluorescent protein was in the first position, and which was in the second (see Methods for alignment analyses). While there appears to be bias towards mCerulean, mTFP1, and mClover3 in the second position (we think most likely due to systematic variation at the pooling stage), all barcode combinations were found to be present, as indicated by color spread throughout the heatmap visualization (**Fig. 2D**). These results validate the cloning process used to generate the barcode pool and provide a plasmid library ready for testing the approach in cells.

Unmixing of pR Probes and Generating Reference Spectra

Before testing the barcode identification performance, we wanted to ensure that each of the 18 fluorescent proteins were deconvolvable from one another, while also generating

the reference spectra for each fp needed for the identification of barcodes. To test whether the 18 fps could be unmixed, we individually transfected each of the 18 pR plasmids (**Fig. 3A**) and assayed cells with a spectral flow cytometer. The spectral data for each sample was exported from the flow cytometer and was analyzed to perform the unmixing of the 18 samples (see Methods). Initial experiments had trouble getting accurate emission spectra for each fluorescent protein samples due to issues keeping fluorescence intensities on-scale (**Fig. S2**). Initial attempts to address these issues looked at reducing the amount of each plasmid that was transfected and the cells that were initially seeded for transfection. While these experiments were able to show decreased levels of fluorescence intensity, many of the experiments resulted in low levels of transfection efficiency (as low as ~100 positive cells). Therefore in addition to the reducing the plasmid concentrations and number of cells seeded, the flow cytometer settings were adjusted to optimize sample analysis. With experimental and mechanical settings optimized, two independent experiments were performed to determine whether the emission spectra were stable and reproducible, with results showing near uniform concordance **(Fig. S3).**

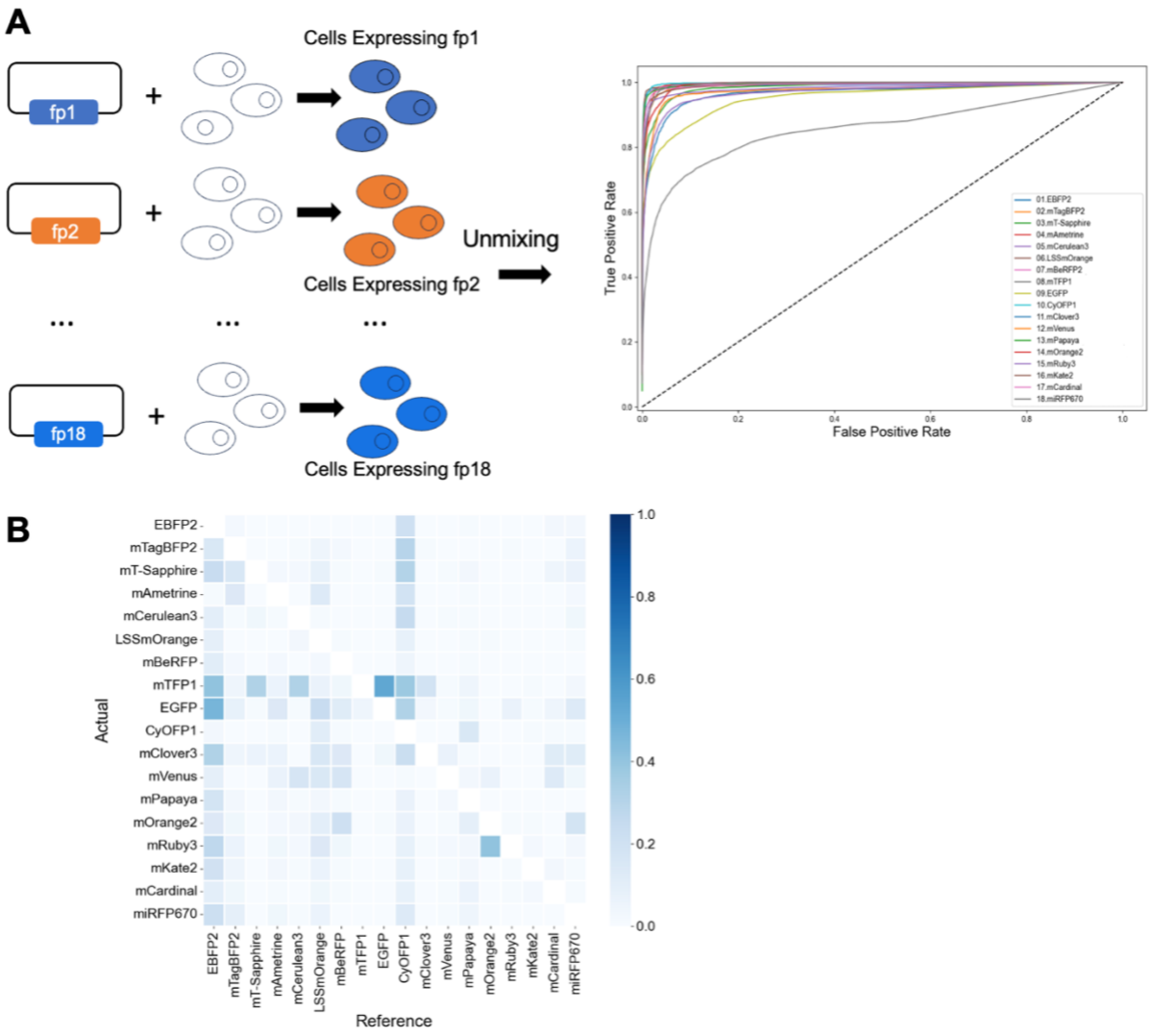

Figure 3. Single-Cell Identification of Fluorescent Proteins. (**A**) Single fluorescent proteins are transfected and then cells analyzed to assess unmixing performance. ROC Curve is generated based on positively gated cells unmixed using reference emission spectra. (**B**) False positive rate heatmap for each of the 18 fp samples based on unmixing using reference emission spectra

As a first metric of performance, we generated an ROC curve for each of the 18 positive fp samples (**Fig. 3A**). Each of the samples were manually gated for positive cells to determine the total number of positive cells so that true positive rates (TPRs) and false positive rates (FPRs) could be determined (see Methods). There results were then plotted where the best performing fps would be expected to reach a TPR of 1 before any false positives were identified. Analysis of the samples showed that fps such as

mTFP1 and EGFP performed the worst while mAmetrine and CyOFP1 performed the best.

To better understand the cause of the difference in performance of the unmixing, we calculated the false positive rates for each fp in each of the 18 samples (**Fig. 3B**). The false positive rates were found for each of the fps at their ideal threshold, which was determined as the threshold correlating to point closest to 100% TPR and 0% FPR on the ROC curve. By analyzing the FPRs we were able to determine which fps the unmixing analysis had trouble differentiating when analyzing a specific fp. From this analysis we can see that the sample positive for mTFP1 had high false positive rates (>0.4) for ~5 fps including EGFP and EBFP2, meaning that >40% of cells expressing mTFP1 were misidentified as expressing EGFP or EBFP2 leading to the low identification performance seen in the ROC curve analysis. Further analysis showed that while CyOFP1 was rarely misidentified as other fps (1 fp with FPR ≥ 0.2), it was misidentified at a high rate in other sampled (8 samples with FPR ≥ 0.2) leading to lower identification performance of these samples.

Identification of Barcodes in Single Cells

To assess the validity of barcode identification in single cells, we decided to isolate individual barcodes so precisely controlled experiments could be performed (**Fig. 4A).** To do this, we performed repeated single-colony pickup following transformation of the pooled library, and then determined which barcodes were found by sequencing. From 111 colonies we obtained 78 unique barcodes for subsequent testing.

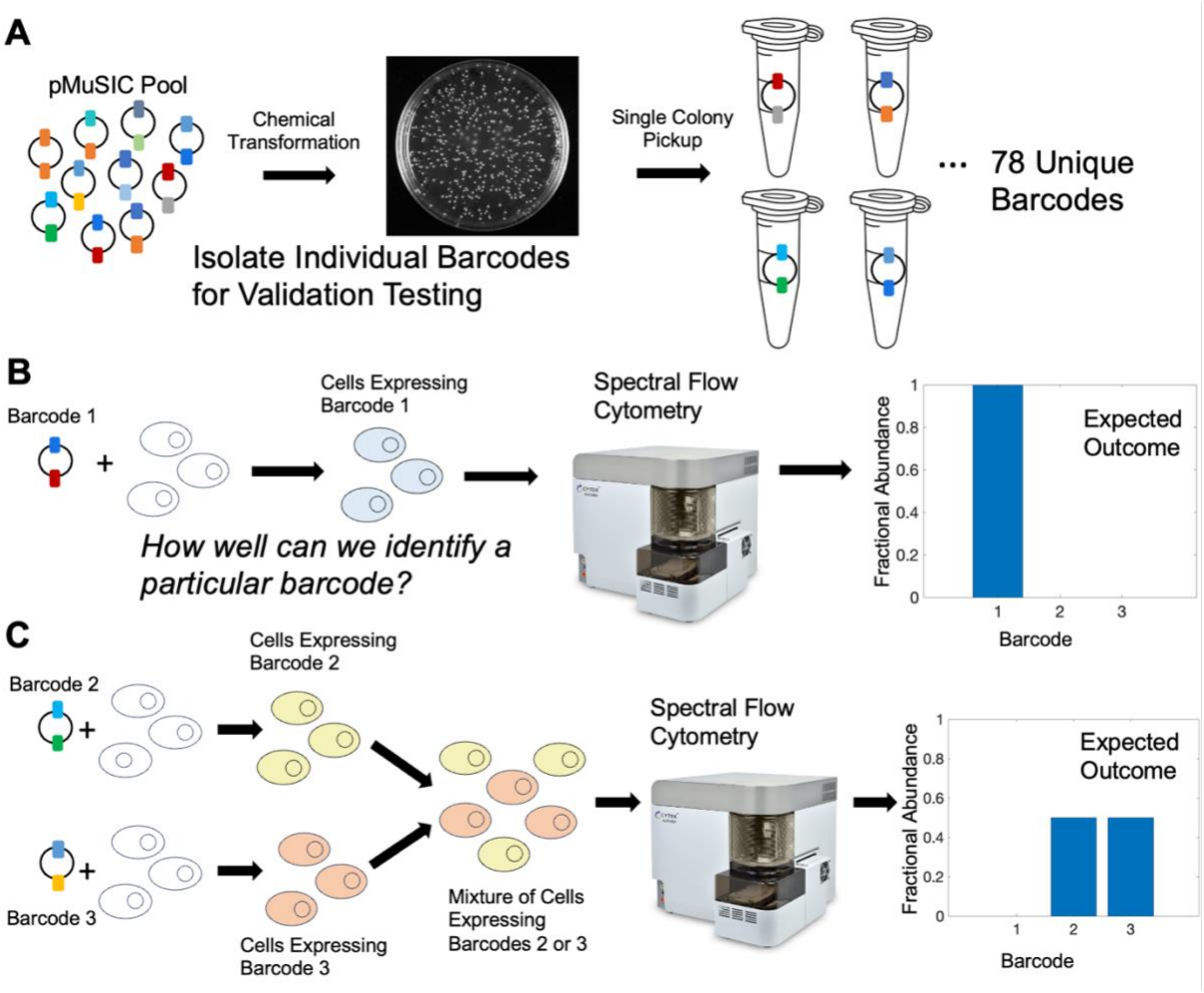

Figure 4. Single-Cell Identification of Barcodes. (**A**) After chemical transformation of the barcode pool, single colonies are picked to obtain individual barcodes. (**B**) Single barcodes are transfected and then analyzed to assess barcode identification performance. (**C**) Single pMuSIC barcodes are transfected into populations of cells, then the cell populations are mixed and analyzed using a spectral flow cytometer to assess barcode identification performance in mixtures.

To assess barcode identification performance we performed two experiments. In the first we transfected a single barcode into a population of cells, then analyzed the cells using a full spectrum flow cytometer (**Fig. 4B**). Analysis of this experiment should show that all positive cells analyzed in each sample contained the known barcode and no other barcodes were identified. In the second experiment we repeated the transfection of single barcodes into populations of cells, cell expressing different barcodes were then combined at known proportions to generate mixed barcode samples (**Fig. 4C**). Results

of this experiment would be expected to show that the percentage of cells expressing each barcode is the same as the proportion that was initially mixed. Analysis of both of these experiments is currently being performed.

Conclusion

While cell barcodes have been used for many years, there are still gaps in current barcoding methods that allow for high throughput experiments with the benefit of genetic encoding. Here were introduce a new barcoding technique that combines fluorescent proteins to generate barcodes that are identifiable at the single-cell level. By combining 18 fluorescent proteins we are able to generate 154 unique barcode combinations. The pooled cloning of these barcodes was validated showing that while there is some bias in the cloning process, all barcode combinations are present. Most of the fluorescent proteins were then shown to be identifiable at the when individually transfected into a population of cells. Further investigations into the difficulties identifying mTFP1 showed that the unmixing often confused it for other fps including EGFP and EBFP2. Finally, experiments to test the identification of barcodes in single-cells are currently being performed and analyzed.

Early experiments aimed at analyzing the fluorescence emissions of individual fluorescent proteins showed that some of the fluorescent proteins fluorescence intensities were off scale resulting in differences in emission spectra from repeated experiments while others remained on-scale. These results proved that further analysis into experimental conditions and the understanding of the flow cytometer settings

needed to be done. Through experimentation, optimal experimental and mechanical settings were able to be determined. While we were able to solve these problems for the purposes of fp identification, more challenges resulting from the differences in intensities between the fps may appear when trying to identify barcodes.

Based on the results of the unmixing of the individual fps it can be reasoned that mTFP1 is hindering the overall unmixing process and will most likely continue to prove difficult in future experiment. Therefore we hypothesize, that getting rid of mTFP1 from the fp pool will allow for better barcode identification. Along with the difficulties identifying mTFP1 in single cells, it can also be seen that both EBFP2 and CyOFP1 have high false positive rates in most of the samples. This is also problematic and leads to lower overall fp identification in all samples. Further studies should look into the extent of complications that the inclusion of these fluorescent proteins causes.

Analysis of the barcode identification experiments were initially performed using flow cytometry analysis software. While this software was able to correctly identify individual fps, the analysis of barcoding experiments showed that the software had difficulties identifying barcodes in single cells (**Fig. S4**). Ideally barcodes would be identified by identifying the two fps that were positive in a cell, however this software identified many of the fps within a single cell preventing barcodes from being determined. Therefore, we are in the process of establishing a coding workflow that allows for the identifies the two fps that are most likely contained within each cell. While it is not ideal to have to develop a way to identify barcodes within a cell it does allow for the identification to be specific to the use case. In this case we are able to identify the two fps that are most likely contained within a cell, instead of identifying all fps that could possibly explain the cell's

emission spectra. While this analysis method has not been verified in terms of barcode identification, a modified workflow was used for the unmixing of the single fps showing its viability.

Future Directions

If the barcode identification experiments produce positive results, the next steps for this project have already been planned to a reasonable extent. As mentioned previously, one of the goals of this barcoding method is to increase the scalability of current methods. Therefore, by incorporating MuSIC into the creation of the probes and barcodes, the number of barcodes produced will be able to be scaled exponentially from what it is currently. Plans for this expansion include developing MuSIC probes consisting of 2- or 3-way combination of fluorescent proteins. With more probes, the number of barcodes increases dramatically allowing for 2- or 3-way combination of the MuSIC probes into MuSIC barcodes. For example, if 3 fps are used to construct MuSIC probes the number of probe combinations becomes ~800. With 800 MuSIC probes, the total number of MuSIC barcodes that can be generated using of 3-way combination of probes is on the order of 10^7.

Other directions include implementation of the barcode method to experiments such as genetic interaction screen or lineage tracing. As part of my time in the Birtwistle Lab, I spent a lot of time planning the future use of this barcoding technique for genetic interaction screening. A pooled approach linking gRNA to individual barcodes was developed utilizing cloning tools such as GoldenGate Cloning. After generating a library

of gRNA linked to barcodes, plans for embedding the library in pairwise combinations was developed in order to systematically induce all 2-way genetic interactions (GIs) possible within an organism were developed. One of the challenges that arose during this planning was finding a way to stably express two barcodes, and therefore gRNA, within a single cell. While transfection was used in the experiments performed in this paper, transfection does not enable stable transfection and the number of plasmids delivered to a cell cannot be controlled. Therefore most applications use viral transduction[17–24]. However viral transduction is not applicable for our approach as the plasmids are either too large for the viral particle (AAV) or other complications such as template switching may occur in viruses that would encapsulate two barcodes (retroviruses and lentiviruses)[45]. Therefore we developed a systematic site-specific recombination method that allows for only two barcodes to be embedded within a single cell (**Fig. S5**).

Another possibility for the next step of this project is to test the viability of the approach through application before expanding the size of the barcode library. For example, a genetic interaction screen measuring the 2-way genetic interaction of 8 genes (~28 GIs) on viability could be perfomed using the current barcode library. Performing a small-scale application experiment would show the viability of this approach and the impact that it can have on answering biological questions.

Aside from different applications of the barcodes with the same analysis method, future directions of this project could also look into expanding the ways in which the barcodes are measured and analyzed. For example, spectral imaging using microscopy can be used to perform spatial and/or temporal analysis on a cell or population of cells

expressing barcodes. By diversifying the ways in which barcodes can be analyzed, the possibilities of what can be studied using the barcodes becomes broader.

Acknowledgements

I would like to thank everybody in the Birtwistle lab for the support that they gave me over the last 2 years that I have been a part of the lab. I would like to especially thank Dr. Xiaoming Lu for mentoring me throughout my time in the lab and for working by my side for this entire project. Dr. Lu started working on this project before I joined the lab and developed most of the groundwork for where I began. She also contributed with performing experiments, analyzing results, writing the code needed to analyze nanopore sequencing results and fp unmixing results, and helped to write methods sections in this paper. I would also like to thank Megan Abravanel for her help in performing ongoing barcode identification experiments. Lastly, I would like to thank Dr. Marc Birtwistle for his continued support and aid in guiding me on my academic and research journey while I have been a part of his lab.

Methods

Cell Culture

HEK293T (ATCC CRL-3216) cells were grown in T-75 flask (Fisher Scientific, 07000665 in full growth media containing DMEM (Gibco, 10313021) supplemented with 10% (v/v) fetal bovine serum (Corning, MT35011CV) and 5% L-Glutamine (*Sigma-Aldrich*, G3126) under a 5% CO_2 atmosphere at 37^{o}C. Cells were sub-cultured every 2-3 days to prevent reaching confluence. This process involved lifting with 0.25% Trypsin (Corning,

MT25050CI), followed by centrifugation at 100*g* for 5 min at room temperature, and a final resuspension in full growth media.

Construction of probes

Sequences of fluorescent proteins were directly amplified from plasmids purchased from *Addgene* or synthesized to silently mutate recognition sites of type IIS restriction enzymes, including *BsaI*, *Esp3I,* and *BbsI.* Purified amplicons were then inserted into both *pR* and *pM* backbones through *BsaI-HF* sites (*NEB*, R3733L) by GoldenGate assembly with *T4* ligase (*NEB,* M0202M) according to the manufacturer's instructions to generate fluorescent probes. Those assembled products were directly transformed into DH5α chemically competent cells and colonies were grown overnight on Agar plates (containing chloramphenicol for pM or Kanamycin for pR at 37°C) Single colonies were then inoculated overnight (16 to 18 hrs) in 2.5 mL of LB Agar media (CAT) at 37°C. Using 0.2µL of the overnight culture colony PCR with *Tag* DNA polymerase (*NEB*, M0267L) was initiated at 95°C for 5min to release the DNA. A series of 10 touchdown cycles were used, consisting of 95°C for 30 sec, 65-53°C decreasing 1°C per cycle for 30 sec, and 68°C for 30 sec. Additional 25 PCR cycles with an annealing temperature at 52°C and a final extension at 68°C for 5 min were performed to ensure complete amplification. Candidates were selected by electrophoresis and further confirmed by double-digestion with proper restriction enzymes in the multiple cloning sites such as *NheI (NEB*, R3131L*), BamHI* (*NEB*, R3136L)*, EcoRI (NEB*, R3101S*), and ApaLi* (*NEB*, R0507S). Constructed pR and pM plasmids with each fluorecent protein gene were finally sequence verified (*Genewiz* or *Plasmidsaurus)*.

Construction of barcodes

For the barcode pool cloning, 0.45 pmol of all fluorescent proteins in both *pR* and *pM* were pooled to create a pR-fp library and a pM-fp library. Each of these libraries were then digested using BbsI-HF (*NEB*, R3539L). 10 units of BbsI (1µL) was added for each reaction and digestion was performed at 37°C for 22 hours The pR pool was then purified using the *Monarch*® PCR & DNA cleanup kits, while the digested pM pool was purified through gel extraction using the DNA gel extraction kits to keep the fragments of interest (fragment sizes between 1.4 and 1.7 kb). CIP treatment was then performed on the pR pool to prevent self-ligation using Quick CIP *(NEB*, M0525S*)*. Quick CIP was added, and the reaction was performed according to the manufacturers protocol. Ligation of the pR and pM pools was then performed at a 1:5 ratio by adding 1µL ofT4 Ligase *(NEB*, M0202T*)* and performing ligation overnight (~22hrs) at 16°C*.*

Flow Cytometry

6 x 10^4 HEK293T cells were seeded into 12-well plates (Fisher Scientific 08-772-29) 24h before transfection. Once cells reached 20% confluency, they were transfected with *pR* or *pM* fluorescent probes or barcode plasmids at concentrations ranging from 50ng to 300 ng using *Lipofectamine 3000* (*ThermoFisher Scientific,* L3000001) and the manufacturer's instructions. Forty-eight hours post-transfection, cells were trypsinized and centrifuged at 300*g* for 5 min at 4°C. Cell pellets were then washed with 1 mL of pre-chilled PBS containing 1% bovine serum albumin (Fisher Scientific, BP1600-100). Samples were resuspended in 0.5ml cold PBS/BSA buffer and then analyzed using the Cytek Aurora spectral flow cytometer.

Cytek Setting Analysis

Cytek Aurora settings are initially set for the use of beads for analysis, as such the settings are not optimal for HEK293T cells as these cells are larger than the beads. To account for this size difference the area scaling factors (ASFs) and gains needed to be adjusted to make sure the fluorescence intensity was accurate and on scale. To maintain the accuracy of the fluorescence intensity it was necessary that the height (H) and area (A) intensities were equivalent at the peak channel of each cell. For example, the peak channel of mRuby3 is YG1, therefore the YG area scaling factor was adjusted until the fluorescent intensities of YG1-A and YG1-H were roughly the same. We repeated this for each of the 18 fps adjusting the area scaling factors relating to the peak channel when needed. Finally, to keep the fluorescence on scale, the gains were adjusted until the brightest sample was fully visible on peak channel fluorescent intensity histograms in SpectroFlo.

Table 1: Laser Delay and Area Scaling Cytek Settings

	V	B	YG	R
Laser Delay	19.55	0	39.75	20.65
Area Scaling	0.4	0.55	1.4	1.66

Table 2: Gains for Cytek Settings

FCS	20

SSC		102		
SSC-B		94		
Channel Gains				
Channel #	V	B	YG	R
1	14	44	31	20
2	22	26	28	20
3	19	26	20	32
4	14	25	41	28
5	19	21	30	19
6	17	18	45	28
7	23	42	30	31
8	32	30	24	16
9	31	44	31	---
10	29	35	29	---
11	24	25	---	---
12	20	25	---	---
13	24	35	---	---

14	27	40	---	---
15	33	---	---	---
16	28	---	---	---

Nanopore Sequencing

The 400 ng of the barcode pool was double-digested using 0.4µL of *MfeI-HF* (*NEB,* R3589L) and 0.4 µL *NheI-HF* (*NEB*, R3131L) at 37°C overnight (~22 hrs) and purified by DNA gel extraction kits (*NEB*, T1020L). Purified DNA fragments (*NheI*-CMV-fp_m-tPT2A-fp_p-*MfeI*) were then prepared for nanopore sequencing using the Ligation Sequencing Kit V14 (Oxford Nanopore Technologies SQK-LSK114), NEBNext Ultra II End repair/dA-tailing Module (NEB E7546), and NEBNext Quick Ligation Module (NEB E6056) according to the manufacturer's instructions. Sequencing was then performed using the Oxford Nanopore Technologies MK1C.

Computational Methods

Nanopore Sequencing Analysis

All sequences contained in the pass package of nanopore sequencing results were decompressed and filtered to retain those only with the expected length. Custom python code (available as supplemental material) was used to analyze and align the sequence data. Briefly, the orientations of DNA sequences were adjusted to 5’-3’, starting from CMV promoter. The CMV common cassette within each read was aligned and scored

as a percentage of bases matching the CMV sequence. Reads exhibiting more than 80% alignment to the CMV sequence were selected for further analysis. This could minimize the global scoring errors during subsequent alignment of the full length of each read to the control library, which contains 324 potential MuSIC barcodes (324 is the total number of permutations while 154 is the number of barcode combinations). Thus, a score matrix of dimensions N x 324 score matrix was generated, where N represented the number of filtered reads. Within the matrix, the highest score among 324 scores for each read was assigned 1 point, indicating the most probable MuSIC barcode, while all the others were set to 0 points. Thus, this scoring matrix consists entirely of 1s and 0s. The points were summed along each column to generate a 1x324 vector.

pR Unmixing

FCS files are exported from SpectroFlo software and converted into CSV files using FCS Express or FlowJo. When exporting files from FlowJo we exported the scale values for each sample to keep the intensity values unmodified from the FCS files. CSV files were then analyzed using custom code utilizing non-negative least squares unmixing. Positive cells are gated manually, then the emission spectra for each positive cell is compared to the emission spectra of 19 references (autofluorescence and 18 fps). 19 comparison values are then outputted for each cell analyzed. The comparison values are a representation of how likely the reference spectra is to account for the cell’s emission spectra.

Determining Fp Threshold

In order to classify cells as positive or negative for a specific fp, we decided to find the comparison value (threshold) for each of the 18 fps that allowed for the most accurate classification. To do this we calculated the true positive rates (TPRs) and false positive rates (FPRs) at every comparison value ranging from 0 to 10 in intervals of 0.01. These values were then plotted in the form of an ROC curve and the threshold for each fp was determined to be the point closest to 100% TPR and 0% FPR.

References:

(1) Krutzik, P. O.; Clutter, M. R.; Trejo, A.; Nolan, G. P. Fluorescent Cell Barcoding for Multiplex Flow Cytometry. *Curr. Protoc. Cytom. Editor. Board J Paul Robinson Manag. Ed. Al* **2011**, *CHAPTER*, Unit-6.31. https://doi.org/10.1002/0471142956.cy0631s55.

(2) Walsh, C.; Cepko, C. L. Widespread Dispersion of Neuronal Clones Across Functional Regions of the Cerebral Cortex. *Science* **1992**, *255* (5043), 434–440. https://doi.org/10.1126/science.1734520.

(3) Krutzik, P. O.; Nolan, G. P. Fluorescent Cell Barcoding in Flow Cytometry Allows High-Throughput Drug Screening and Signaling Profiling. *Nat. Methods* **2006**, *3* (5), 361–368. https://doi.org/10.1038/nmeth872.

(4) Weissman, T. A.; Pan, Y. A. Brainbow: New Resources and Emerging Biological Applications for Multicolor Genetic Labeling and Analysis. *Genetics* **2015**, *199* (2), 293–306. https://doi.org/10.1534/genetics.114.172510.

(5) Kretzschmar, K.; Watt, F. M. Lineage Tracing. *Cell* **2012**, *148* (1–2), 33–45. https://doi.org/10.1016/j.cell.2012.01.002.

(6) Wu, S.-H. (Sam); Lee, J.-H.; Koo, B.-K. Lineage Tracing: Computational Reconstruction Goes Beyond the Limit of Imaging. *Mol. Cells* **2019**, *42* (2), 104–112. https://doi.org/10.14348/molcells.2019.0006.

(7) Boone, C.; Bussey, H.; Andrews, B. J. Exploring Genetic Interactions and Networks with Yeast. *Nat. Rev. Genet.* **2007**, *8* (6), 437–449. https://doi.org/10.1038/nrg2085.

(8) Adamson, B.; Norman, T. M.; Jost, M.; Cho, M. Y.; Nuñez, J. K.; Chen, Y.; Villalta, J. E.; Gilbert, L. A.; Horlbeck, M. A.; Hein, M. Y.; Pak, R. A.; Gray, A. N.; Gross, C. A.; Dixit, A.; Parnas, O.; Regev, A.; Weissman, J. S. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* **2016**, *167* (7), 1867-1882.e21. https://doi.org/10.1016/j.cell.2016.11.048.

(9) Bock et al. *High-content CRISPR screening | Nature Reviews Methods Primers*. https://www.nature.com/articles/s43586-021-00093-4 (accessed 2023-06-18).

(10) Macosko et al. *Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets - ScienceDirect*. https://www.sciencedirect.com/science/article/pii/S0092867415005498 (accessed 2023-06-29).

(11) Santinha, A. J.; Klingler, E.; Kuhn, M.; Farouni, R.; Lagler, S.; Kalamakis, G.; Lischetti, U.; Jabaudon, D.; Platt, R. J. Transcriptional Linkage Analysis with in Vivo AAV-Perturb-Seq. *Nature* **2023**, 1–9. https://doi.org/10.1038/s41586-023-06570-y.

(12) Kebschull, J. M.; Zador, A. M. Cellular Barcoding: Lineage Tracing, Screening and Beyond. *Nat. Methods* **2018**, *15* (11), 871–879. https://doi.org/10.1038/s41592-018-0185-x.

(13) Zheng, G. X. Y.; Terry, J. M.; Belgrader, P.; Ryvkin, P.; Bent, Z. W.; Wilson, R.; Ziraldo, S. B.; Wheeler, T. D.; McDermott, G. P.; Zhu, J.; Gregory, M. T.; Shuga, J.; Montesclaros, L.; Underwood, J. G.; Masquelier, D. A.; Nishimura, S. Y.; Schnall-Levin, M.; Wyatt, P. W.; Hindson, C. M.; Bharadwaj, R.; Wong, A.; Ness, K. D.; Beppu, L. W.; Deeg, H. J.; McFarland, C.; Loeb, K. R.; Valente, W. J.; Ericson, N. G.; Stevens, E. A.; Radich, J. P.; Mikkelsen, T. S.; Hindson, B. J.; Bielas, J. H.

Massively Parallel Digital Transcriptional Profiling of Single Cells. *Nat. Commun.* **2017**, *8* (1), 14049. https://doi.org/10.1038/ncomms14049.
(14) Replogle, J. M.; Saunders, R. A.; Pogson, A. N.; Hussmann, J. A.; Lenail, A.; Guna, A.; Mascibroda, L.; Wagner, E. J.; Adelman, K.; Lithwick-Yanai, G.; Iremadze, N.; Oberstrass, F.; Lipson, D.; Bonnar, J. L.; Jost, M.; Norman, T. M.; Weissman, J. S. Mapping Information-Rich Genotype-Phenotype Landscapes with Genome-Scale Perturb-Seq. *Cell* **2022**, *185* (14), 2559-2575.e28. https://doi.org/10.1016/j.cell.2022.05.013.
(15) McCarthy, M. E.; Anglin, C. M.; Peer, H. A.; Boleman, S. A.; Klaubert, S. R.; Birtwistle, M. R. Protocol for Creating Antibodies with Complex Fluorescence Spectra. *Bioconjug. Chem.* **2021**, *32* (6), 1156–1166. https://doi.org/10.1021/acs.bioconjchem.1c00220.
(16) Du, D.; Roguev, A.; Gordon, D. E.; Chen, M.; Chen, S.-H.; Shales, M.; Shen, J. P.; Ideker, T.; Mali, P.; Qi, L. S.; Krogan, N. J. Genetic Interaction Mapping in Mammalian Cells Using CRISPR Interference. *Nat. Methods* **2017**, *14* (6), 577–580. https://doi.org/10.1038/nmeth.4286.
(17) Replogle, J. M.; Norman, T. M.; Xu, A.; Hussmann, J. A.; Chen, J.; Cogan, J. Z.; Meer, E. J.; Terry, J. M.; Riordan, D. P.; Srinivas, N.; Fiddes, I. T.; Arthur, J. G.; Alvarado, L. J.; Pfeiffer, K. A.; Mikkelsen, T. S.; Weissman, J. S.; Adamson, B. Combinatorial Single-Cell CRISPR Screens by Direct Guide RNA Capture and Targeted Sequencing. *Nat. Biotechnol.* **2020**, *38* (8), 954–961. https://doi.org/10.1038/s41587-020-0470-y.
(18) Jaitin et al. *Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq - ScienceDirect*. https://www.sciencedirect.com/science/article/pii/S0092867416316117 (accessed 2023-06-29).
(19) Xie, S.; Duan, J.; Li, B.; Zhou, P.; Hon, G. C. Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells. *Mol. Cell* **2017**, *66* (2), 285-299.e5. https://doi.org/10.1016/j.molcel.2017.03.007.
(20) Hill, A. J.; McFaline-Figueroa, J. L.; Starita, L. M.; Gasperini, M. J.; Matreyek, K. A.; Packer, J.; Jackson, D.; Shendure, J.; Trapnell, C. On the Design of CRISPR-Based Single Cell Molecular Screens. *Nat. Methods* **2018**, *15* (4), 271–274. https://doi.org/10.1038/nmeth.4604.
(21) Dixit, A.; Parnas, O.; Li, B.; Chen, J.; Fulco, C. P.; Jerby-Arnon, L.; Marjanovic, N. D.; Dionne, D.; Burks, T.; Raychowdhury, R.; Adamson, B.; Norman, T. M.; Lander, E. S.; Weissman, J. S.; Friedman, N.; Regev, A. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **2016**, *167* (7), 1853-1866.e17. https://doi.org/10.1016/j.cell.2016.11.038.
(22) Datlinger, P.; Rendeiro, A. F.; Schmidl, C.; Krausgruber, T.; Traxler, P.; Klughammer, J.; Schuster, L. C.; Kuchler, A.; Alpar, D.; Bock, C. Pooled CRISPR Screening with Single-Cell Transcriptome Readout. *Nat. Methods* **2017**, *14* (3), 297–301. https://doi.org/10.1038/nmeth.4177.
(23) Schraivogel, D.; Gschwind, A. R.; Milbank, J. H.; Leonce, D. R.; Jakob, P.; Mathur, L.; Korbel, J. O.; Merten, C. A.; Velten, L.; Steinmetz, L. M. Targeted

Perturb-Seq Enables Genome-Scale Genetic Screens in Single Cells. *Nat. Methods* **2020**, *17* (6), 629–635. https://doi.org/10.1038/s41592-020-0837-5.
(24) Horlbeck, M. A.; Xu, A.; Wang, M.; Bennett, N. K.; Park, C. Y.; Bogdanoff, D.; Adamson, B.; Chow, E. D.; Kampmann, M.; Peterson, T. R.; Nakamura, K.; Fischbach, M. A.; Weissman, J. S.; Gilbert, L. A. Mapping the Genetic Landscape of Human Cells. *Cell* **2018**, *174* (4), 953-967.e22. https://doi.org/10.1016/j.cell.2018.06.010.
(25) McCarthy, M. E.; Dodd, W. B.; Lu, X.; Pritko, D. J.; Patel, N. D.; Haskell, C. V.; Sanabria, H.; Blenner, M. A.; Birtwistle, M. R. Theory for High-Throughput Genetic Interaction Screening. *ACS Synth. Biol.* **2023**, *12* (8), 2290–2300. https://doi.org/10.1021/acssynbio.2c00627.
(26) Holzapfel, H. Y.; Stern, A. D.; Bouhaddou, M.; Anglin, C. M.; Putur, D.; Comer, S.; Birtwistle, M. R. Fluorescence Multiplexing with Spectral Imaging and Combinatorics. *ACS Comb. Sci.* **2018**, *20* (11), 653–659. https://doi.org/10.1021/acscombsci.8b00101.
(27) McCarthy, M. E.; Lu, X.; Ogunleye, O.; Latham, D. R.; Abravanel, M.; Pritko, D.; Huggins, J. R.; Haskell, C. V.; Patel, N. D.; Pittman, Z. A.; Sanabria, H.; Birtwistle, M. R. Increasing Signal Intensity of Fluorescent Oligo-Labeled Antibodies to Enable Combination Multiplexing. bioRxiv May 30, 2024, p 2023.07.06.547965. https://doi.org/10.1101/2023.07.06.547965.
(28) Hellenkamp, B.; Schmid, S.; Doroshenko, O.; Opanasyuk, O.; Kühnemuth, R.; Rezaei Adariani, S.; Ambrose, B.; Aznauryan, M.; Barth, A.; Birkedal, V.; Bowen, M. E.; Chen, H.; Cordes, T.; Eilert, T.; Fijen, C.; Gebhardt, C.; Götz, M.; Gouridis, G.; Gratton, E.; Ha, T.; Hao, P.; Hanke, C. A.; Hartmann, A.; Hendrix, J.; Hildebrandt, L. L.; Hirschfeld, V.; Hohlbein, J.; Hua, B.; Hübner, C. G.; Kallis, E.; Kapanidis, A. N.; Kim, J.-Y.; Krainer, G.; Lamb, D. C.; Lee, N. K.; Lemke, E. A.; Levesque, B.; Levitus, M.; McCann, J. J.; Naredi-Rainer, N.; Nettels, D.; Ngo, T.; Qiu, R.; Robb, N. C.; Röcker, C.; Sanabria, H.; Schlierf, M.; Schröder, T.; Schuler, B.; Seidel, H.; Streit, L.; Thurn, J.; Tinnefeld, P.; Tyagi, S.; Vandenberk, N.; Vera, A. M.; Weninger, K. R.; Wünsch, B.; Yanez-Orozco, I. S.; Michaelis, J.; Seidel, C. A. M.; Craggs, T. D.; Hugel, T. Precision and Accuracy of Single-Molecule FRET Measurements—a Multi-Laboratory Benchmark Study. *Nat. Methods* **2018**, *15* (9), 669–676. https://doi.org/10.1038/s41592-018-0085-0.
(29) Ai, H.; Shaner, N. C.; Cheng, Z.; Tsien, R. Y.; Campbell, R. E. Exploration of New Chromophore Structures Leads to the Identification of Improved Blue Fluorescent Proteins. *Biochemistry* **2007**, *46* (20), 5904–5910. https://doi.org/10.1021/bi700199g.
(30) Subach, O. M.; Cranfill, P. J.; Davidson, M. W.; Verkhusha, V. V. An Enhanced Monomeric Blue Fluorescent Protein with the High Chemical Stability of the Chromophore. *PLOS ONE* **2011**, *6* (12), e28674. https://doi.org/10.1371/journal.pone.0028674.
(31) Ai, H.; Hazelwood, K. L.; Davidson, M. W.; Campbell, R. E. Fluorescent Protein FRET Pairs for Ratiometric Imaging of Dual Biosensors. *Nat. Methods* **2008**, *5* (5), 401–403. https://doi.org/10.1038/nmeth.1207.
(32) Markwardt, M. L.; Kremers, G.-J.; Kraft, C. A.; Ray, K.; Cranfill, P. J. C.; Wilson, K. A.; Day, R. N.; Wachter, R. M.; Davidson, M. W.; Rizzo, M. A. An Improved

Cerulean Fluorescent Protein with Enhanced Brightness and Reduced Reversible Photoswitching. *PLOS ONE* **2011**, *6* (3), e17896. https://doi.org/10.1371/journal.pone.0017896.
(33) Shcherbakova, D. M.; Hink, M. A.; Joosen, L.; Gadella, T. W. J.; Verkhusha, V. V. An Orange Fluorescent Protein with a Large Stokes Shift for Single-Excitation Multicolor FCCS and FRET Imaging. *J. Am. Chem. Soc.* **2012**, *134* (18), 7913–7923. https://doi.org/10.1021/ja3018972.
(34) Yang, J.; Wang, L.; Yang, F.; Luo, H.; Xu, L.; Lu, J.; Zeng, S.; Zhang, Z. mBeRFP, an Improved Large Stokes Shift Red Fluorescent Protein. *PLOS ONE* **2013**, *8* (6), e64849. https://doi.org/10.1371/journal.pone.0064849.
(35) Ai, H.; Henderson, J. N.; Remington, S. J.; Campbell, R. E. Directed Evolution of a Monomeric, Bright and Photostable Version of Clavularia Cyan Fluorescent Protein: Structural Characterization and Applications in Fluorescence Imaging. *Biochem. J.* **2006**, *400* (3), 531–540. https://doi.org/10.1042/BJ20060874.
(36) Cormack, B. P.; Valdivia, R. H.; Falkow, S. FACS-Optimized Mutants of the Green Fluorescent Protein (GFP). *Gene* **1996**, *173* (1), 33–38. https://doi.org/10.1016/0378-1119(95)00685-0.
(37) Chu, J.; Oh, Y.; Sens, A.; Ataie, N.; Dana, H.; Macklin, J. J.; Laviv, T.; Welf, E. S.; Dean, K. M.; Zhang, F.; Kim, B. B.; Tang, C. T.; Hu, M.; Baird, M. A.; Davidson, M. W.; Kay, M. A.; Fiolka, R.; Yasuda, R.; Kim, D. S.; Ng, H.-L.; Lin, M. Z. A Bright Cyan-Excitable Orange Fluorescent Protein Facilitates Dual-Emission Microscopy and Enhances Bioluminescence Imaging in Vivo. *Nat. Biotechnol.* **2016**, *34* (7), 760–767. https://doi.org/10.1038/nbt.3550.
(38) Bajar, B. T.; Wang, E. S.; Lam, A. J.; Kim, B. B.; Jacobs, C. L.; Howe, E. S.; Davidson, M. W.; Lin, M. Z.; Chu, J. Improving Brightness and Photostability of Green and Red Fluorescent Proteins for Live Cell Imaging and FRET Reporting. *Sci. Rep.* **2016**, *6* (1), 20889. https://doi.org/10.1038/srep20889.
(39) Kremers, G.-J.; Goedhart, J.; van Munster, E. B.; Gadella, T. W. J. Cyan and Yellow Super Fluorescent Proteins with Improved Brightness, Protein Folding, and FRET Förster Radius,. *Biochemistry* **2006**, *45* (21), 6570–6580. https://doi.org/10.1021/bi0516273.
(40) Hoi, H.; Howe, E. S.; Ding, Y.; Zhang, W.; Baird, M. A.; Sell, B. R.; Allen, J. R.; Davidson, M. W.; Campbell, R. E. An Engineered Monomeric *Zoanthus Sp.* Yellow Fluorescent Protein. *Chem. Biol.* **2013**, *20* (10), 1296–1304. https://doi.org/10.1016/j.chembiol.2013.08.008.
(41) Shaner, N. C.; Lin, M. Z.; McKeown, M. R.; Steinbach, P. A.; Hazelwood, K. L.; Davidson, M. W.; Tsien, R. Y. Improving the Photostability of Bright Monomeric Orange and Red Fluorescent Proteins. *Nat. Methods* **2008**, *5* (6), 545–551. https://doi.org/10.1038/nmeth.1209.
(42) Shcherbo, D.; Murphy, C. S.; Ermakova, G. V.; Solovieva, E. A.; Chepurnykh, T. V.; Shcheglov, A. S.; Verkhusha, V. V.; Pletnev, V. Z.; Hazelwood, K. L.; Roche, P. M.; Lukyanov, S.; Zaraisky, A. G.; Davidson, M. W.; Chudakov, D. M. Far-Red Fluorescent Tags for Protein Imaging in Living Tissues. *Biochem. J.* **2009**, *418* (3), 567–574. https://doi.org/10.1042/BJ20081949.
(43) Chu, J.; Haynes, R. D.; Corbel, S. Y.; Li, P.; González-González, E.; Burg, J. S.; Ataie, N. J.; Lam, A. J.; Cranfill, P. J.; Baird, M. A.; Davidson, M. W.; Ng, H.-L.;

Garcia, K. C.; Contag, C. H.; Shen, K.; Blau, H. M.; Lin, M. Z. Non-Invasive Intravital Imaging of Cellular Differentiation with a Bright Red-Excitable Fluorescent Protein. *Nat. Methods* **2014**, *11* (5), 572–578. https://doi.org/10.1038/nmeth.2888.

(44) Shcherbakova, D. M.; Baloban, M.; Emelyanov, A. V.; Brenowitz, M.; Guo, P.; Verkhusha, V. V. Bright Monomeric Near-Infrared Fluorescent Proteins as Tags and Biosensors for Multiscale Imaging. *Nat. Commun.* **2016**, *7* (1), 12405. https://doi.org/10.1038/ncomms12405.

(45) Sack, L. M.; Davoli, T.; Xu, Q.; Li, M. Z.; Elledge, S. J. Sources of Error in Mammalian Genetic Screens. *G3 GenesGenomesGenetics* **2016**, *6* (9), 2781–2790. https://doi.org/10.1534/g3.116.030973.

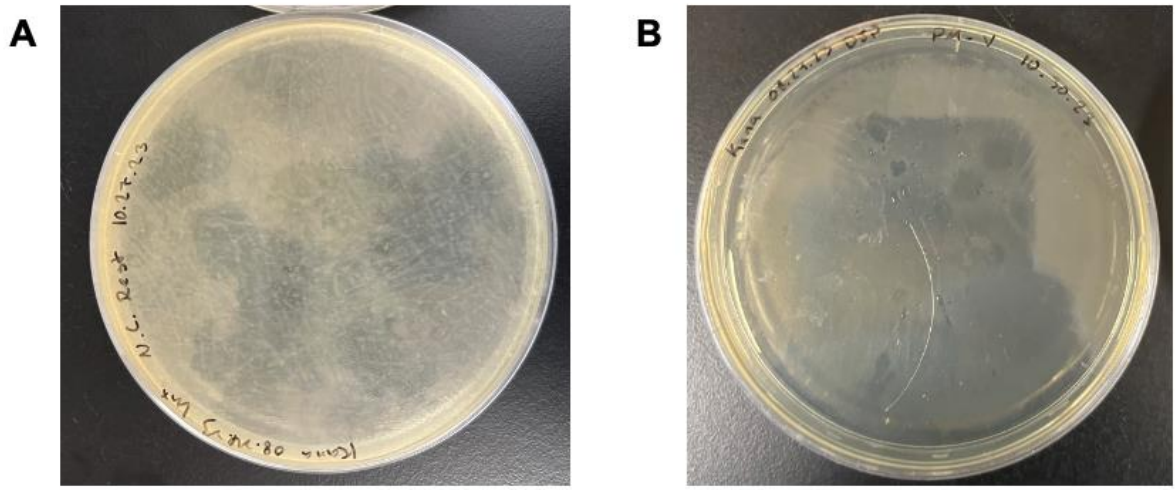

Figure S1: Chemical Transformation Results. Results from the chemical transformation of the negative control of the barcode library are shown here. For these experiments the negative control was made by digestion of the pR plasmids with BbsI and no addition of any pM plasmids. (**A**) After digestion of the pR plasmids, a ligation experiment was performed on the pool and the products were chemically transformed.. (**B**) After digestion of the pR plasmids, no ligation experiment was performed and the products were chemcially transformed.

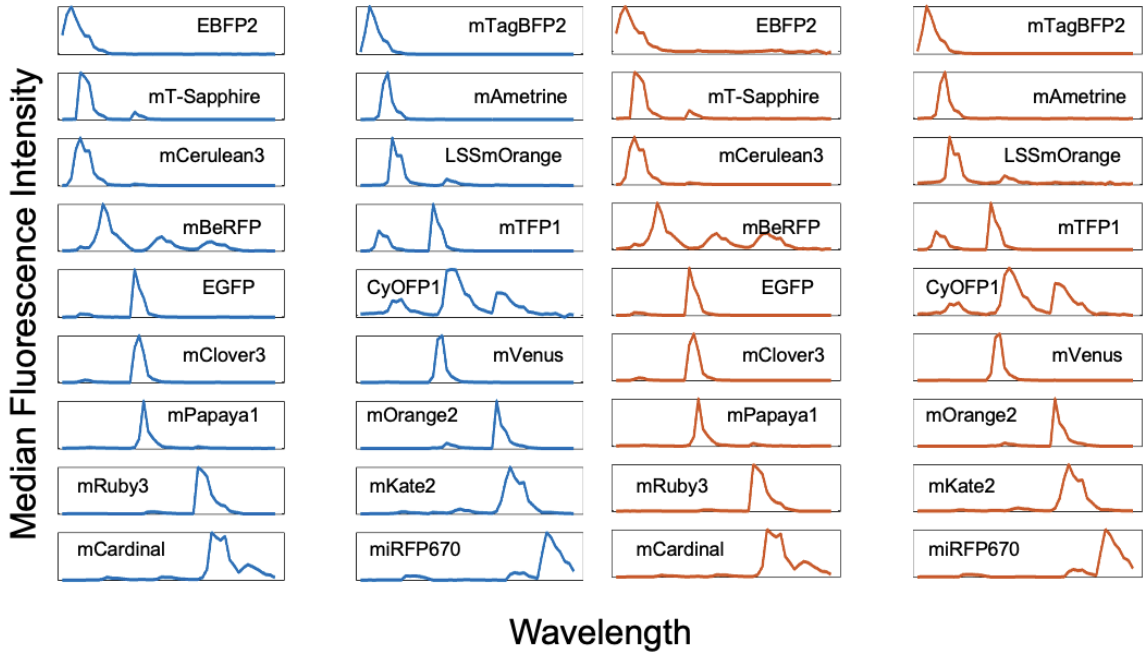

Figure S2. Comparison of Fluorescent Protein Emission Spectra. All 18 fluorescent proteins are analyzed to determine the spectral shape for each fp in each channel remains consistent throughout the analysis. The two experiments were performed a year apart from one another.

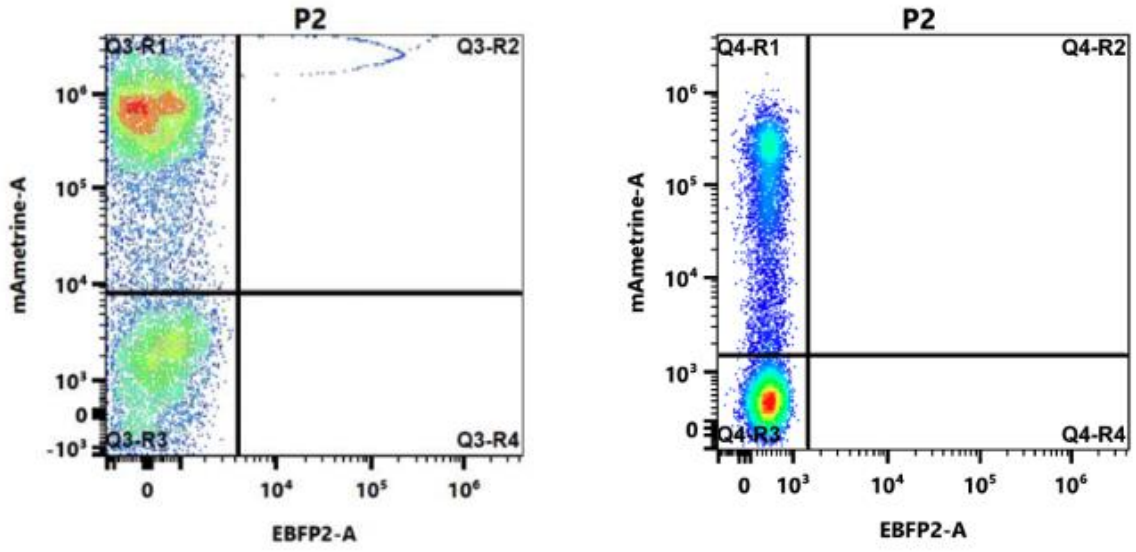

Figure S3: On-Scale vs. Off-Scale Fluorescence Intensities. For both of these results, a plasmid expressing mAmetrine was transfected into a population of cells. The transfected cells were then analyzed using a spectral flow cytometer. The samples were then unmixed to show the intensities of each fluorescent protein expressed within each cell. Results on the left came from an experiment before optimization of experimental and mechanical settings, whereas the results on the right are from an experiment performed after the settings optimization.

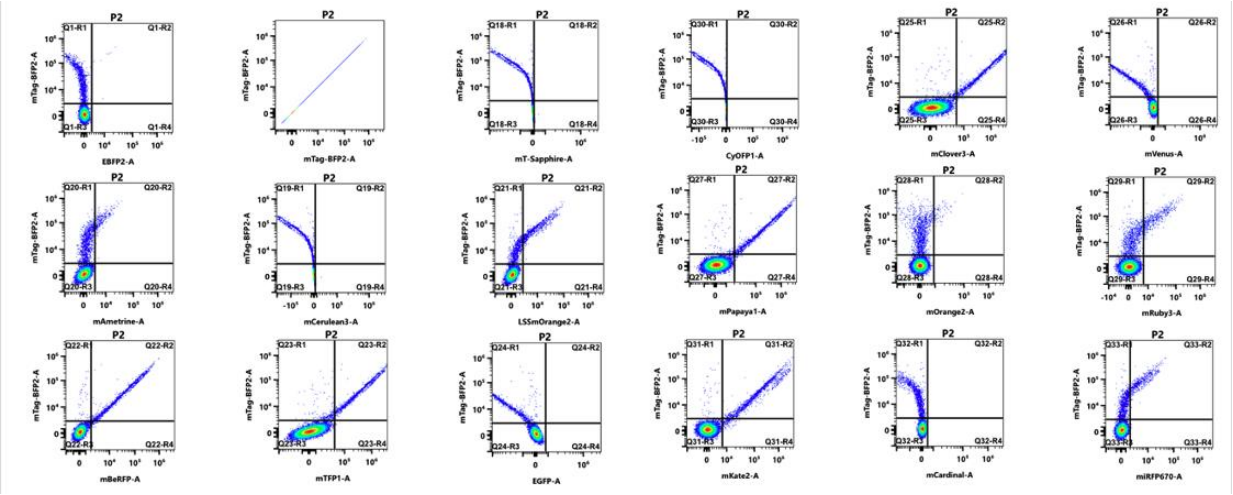

Figure S4: Unmixing of Barcode Pool with Flow Cytometry Software. A fluorescent barcode containing fluorescent proteins mT-BFP2 and mKate2 was transfected into a population of cells. Unmixing of the cells analyzed was performed using flow cytometry software to determine which fluorescent proteins were found in each cell.

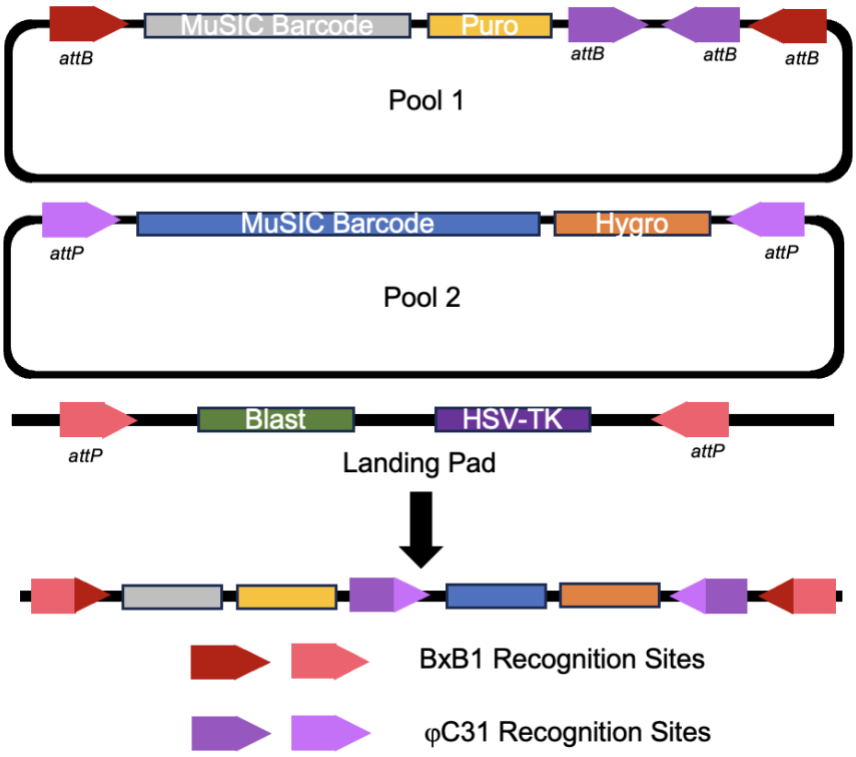

Figure S5: Workflow for Stable Integration of MuSIC Barcodes. Workflow showing how recombinases can be utilized to stably integrate two MuSIC barcodes into a single cell.