Engineering Subcellular Fluorescence for High-Throughput Microscopy Screening in Microbial Biofuels Production

Austin T. Jones

ABSTRACT

Engineering microbes for biofuels production demands precise control over subcellular localization, cell morphology, motility, and other such phenotypes that are only observable via microscopy. However, screening libraries of subcellular phenotypes using microscopy is currently inefficient and impractical. To bridge the phenotype of interest with the genetic identity of each microscopy-screened library member, we engineered microscopy barcodes (MiCodes) by targeting three fluorescent proteins (RFP, CFP, and GFP) to four organelles (nucleus, cell periphery, vacuolar membrane, and actin) in unique combinations within *Saccharomyces cerevisiae*. The MiCodes were optimized by selection of organelle shape and fluorescent protein expression creating visually distinct and identifiable cell signatures. The MiCode assembly was completed by iterative Golden Gate cloning, allowing for the engineering of large constructs with interchangeable parts. MiCodes were used to analyze interaction strength of a 36 member leucine zipper library using a bait and prey targeting assay to determine pairwise affinities and find orthogonal protein pairs. By coupling a unique fluorescent signature to each genotype present in a library, MiCodes enable high-throughput library screening using microscopy. MiCodes can be scaled to million member libraries with an automated screening method using script developed on two platforms: MatLab for image acquisition and CellProfiler for image analysis. Using photoactivatable fluorescent proteins and a confocal microscope increases the number of possible MiCodes and accuracy of MiCode identification, further propelling their application to large-scale libraries. MiCodes serve as a tool for synthetic biologists to increase the complexity and harness greater control of metabolic engineering for biofuels synthesis.

KEYWORDS

synthetic biology, leucine zipper, protein interaction, subcellular localization, library screening

INTRODUCTION

To minimize anthropogenic climate change, there has been a shift away from fossil fuels towards renewable energy, notably agriculture-based biofuels (Fargione et al. 2008, Huang et al. 2012). Currently, ethanol from corn and sugar cane comprise the majority of biofuels production; however, this form of alternative energy is plagued with inefficiencies and indirect social costs that hinder its ability to compete with fossil fuels (Fargione et al. 2008, Huang et al. 2012, Zhu and Zhuang 2012). Decomposing vegetation from the conversion of natural landscapes into farmland devoted to biofuels production releases CO_2, creating a carbon debt- a negative carbon dioxide payoff- that requires centuries of biofuels production to overcome (Fargione et al. 2008). Lifecycle analysis has shown zero net energy gain associated with agricultural biofuels given that the farm's electricity and fertilizer are derived from conventional fossil fuel sources (Zhu and Zhuang 2012). Current agricultural biofuels also impact food security: as farmland is diverted towards biofuels and away from food agriculture, food prices skyrocket worldwide as supply falls (Huang et al. 2012). Rising prices have resulted in indirect social costs with protests and civil unrest occurring in the world's underdeveloped and developing nations (Huang et al. 2012). With a world population estimated to reach 10 billion by the end of the century, we can no longer divert crops towards fuel and away from food (Lutz et al. 2001). These biological and social dynamics have rendered agricultural based biofuels as an unviable option to compete with fossil fuels.

To address these drawbacks of agricultural biofuels, bioengineering and synthetic biology have focused on diverting the fuel source away from plants and towards microorganisms (Keasling and Chou 2008, Hong and Nielsen 2012, Peralta-Yahya et al. 2012). By reengineering the metabolic pathways responsible for the production of ethanol, butanol, high-octane hydrocarbons, and other forms of fuel, microbial cells can be programmed as "cell factories" for the efficient production of alternative fuel sources (Hong and Nielsen 2012). Although some microorganisms naturally ferment ethanol, a higher volume of output can be achieved by redirecting pathways responsible for amino acid biosynthesis towards the production of the desired product, enabling the organism to produce fuel in an aerobic environment (Keasling and Chou 2008). However, as biochemical pathways become repurposed and expanded within the cell, undesirable interactions with metabolic side paths hinder efficiency and reduce the fuel

output (Lee et al. 2012). Protein scaffolds are seen as an emerging solution to limit competing reactions in a metabolic pathway, tethering the enzymes that comprise the biochemical pathway into a confined protein complex. Protein scaffolds ultimately localize reaction machinery thereby minimizing toxicity of metabolic intermediates, diffusion outside of the reaction complex, allosteric negative feedback, and cross talk with other metabolic pathways (Hong and Nielsen 2012). Designing microbes for biofuels production using protein scaffolds demands precise control over subcellular localization in distinct membrane compartments, cell morphology, and other such phenotypes that are only observable via microscopy.

This engineering of complex biological systems is difficult due to the limited knowledge regarding the synergistic effect of parameters being varied within a system. To achieve the highest optimization, combinatorial libraries are constructed with each member (cells, proteins, etc.) possessing a distinct and random combination of possible variables (Cohen et al. 1998, Pepperkok and Ellenberg 2006). Such a combinatorial approach often exceeds one million variants, a size that microscopy- the only means to view subcellular properties- is unable to screen in a feasible manner (Pepperkok and Ellenberg 2006).

Microscopy barcodes (MiCodes) are a system of fluorescent barcodes that will enable the use of fluorescence microscopy as a high throughput screening technique by linking a library member's phenotype of interest (morphology, subcellular localization, etc.) to its precise genotype variation (Dueber, unpublished data). The barcode is created by targeting an array of fluorescent proteins to organelles within the cell creating a set of recognizable and distinct patterns represented by the function (Figure 1). Upon finding a phenotype of interest, the barcode can be revealed under the fluorescent microscope channels to link the genetic features that brought about the organism's observed phenotype, thereby avoiding individual cell isolation and sequencing. Engineering these properties and screening for successful mutants of a desired phenotype is challenging due to the inherent low throughput (the rate of outputs in the screening process) imposed by microscopy. By linking a library member's phenotype of interest with the respective genotype, MiCodes expands the potential of fluorescence microscopy as a screening technique for subcellular and morphological properties.

$$Number\ of\ MiCodes = 2^{number\ of\ fluorescent\ proteins^{number\ of\ targetted\ organelles}}$$

Figure 1. Theoretical Number of MiCodes. The number of theoretical MiCodes as a function of the number of fluorescent proteins and organelles to which they are localized.

The objective of this study is to enable high-throughput library screening with microscopy by coupling a unique fluorescent signature to each genotype present in a library. These MiCodes were generated by targeting combinations of fluorescent proteins to organelles within *Saccharomyces cerevisiae*, baker's yeast (Hansen, 1883). We foresee the ability to scale MiCodes to library sizes of 10^6 or more, ensuring that this technology could be practical for common library sizes used in synthetic biology. We examine whether the screening and analysis system could be automated by writing novel script using existing image processing software, enabling computer detection and identification of MiCodes. As a proof of concept, we apply MiCodes to the problem of finding unique pairs of interacting proteins for a synthetic biology toolkit. Such protein interactions are termed orthogonal proteins: ones that specifically interact with a designated partner and do not exhibit affinity for other proteins within the cell. These protein interactions are of interest in synthetic biology and microbial biofuels production to prevent exchanges across engineered pathways, yielding maximum fuel output. MiCodes will provide a mechanism for high-throughput microscopy screening, an essential component for engineering orthogonal protein interactions.

METHODS AND RESULTS

MiCode design

The model organisms *Saccharomyces cerevisiae* and *Escherichia coli* were used to construct and test the MiCode system on a library of leucine zippers of varying affinities, analyzing their interaction strength through novel software script. These model organisms were used because of the extensive knowledge of their systems. *E. coli* was the site of plasmid construction while *S. cerevisiae* was chosen as a simple eukaryote containing the organelles necessary to create the MiCode.

Organelle targeting proteins and signal sequences were fused to various fluorescent proteins through Golden Gate cloning, a highly efficient, one-pot (having reagents mixed in one chamber without the need for intermediate purification steps) reaction to assemble combinatorial MiCode components. Using protein localization studies in budding yeast from the Yeast GFP Database produced by the O'Shea and Weissman labs at UCSF, we searched for organelles that met the following criteria: existing targeting proteins or signal sequences in the literature and visible distinction amongst other fluorescently tagged organelles. We PCRed eligible proteins from the genome, fused them to fluorescent proteins, and narrowed our choices aiming to find consistent morphology and geometry.

The MiCodes were optimized by organelle shape and the level of fluorescent protein expression creating visually distinct and identifiable cell signatures. The initial twelve candidates of organelle-localizing proteins were narrowed to four, selected based on visual distinction on size and shape (Figure 2). These included actin, nucleus, vacuolar membrane, and cell periphery. The cell periphery acts as a large ring encapsulating the smaller ring of the vacuolar membrane while the large circular shape of the nucleus can be distinguished from the many smaller dots of actin. Targeting proteins were used for the nucleus, vacuolar membrane, and actin while targeting sequences for the peroxisome and cell periphery were fused to the C terminus of the flurophores in lieu of targeting proteins (Table 1). The peroxisome functioned in the bait and prey assay to determine protein interaction strength and was not a component of the MiCoding system. Using these four localizing tags ensured that a MiCode could be identified readily in the case that it encoded for all four organelles in the same fluorescent channel.

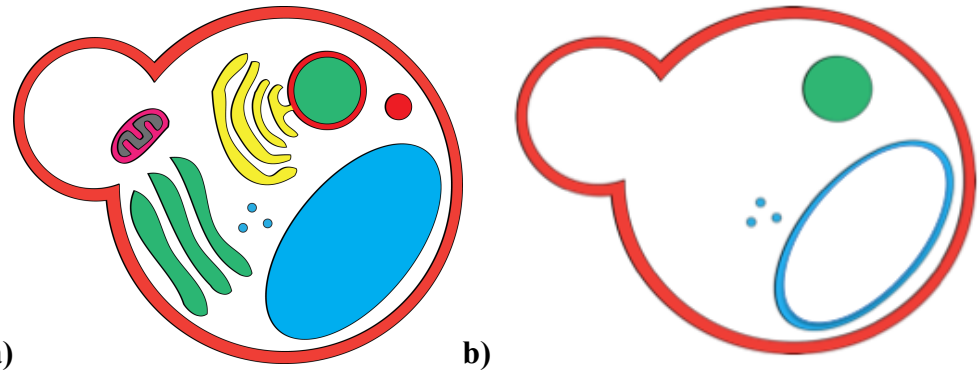


Figure 2. Yeast and MiCode organelles a) Representation of yeast organelles and b) the four locations used for MiCoding: cell periphery, vacuolar membrane, actin, and nucleus (right).

Table 1. Proteins and signal sequences used for the localization of the fluorescent proteins. The nucleus, vacuolar membrane, actin, and cell periphery comprised the MiCodes while the peroxisome was used in the bait and prey protein affinity assay.

Organelle	Gene	Protein Function	C terminal Signal Sequence
Nucleus	H2A	Histone subunit	-
Vacuolar Membrane	ZRC1	Vacuolar Membrane zinc transporter	-
Actin	ABP1	Actin cytoskeleton	-
Peroxisome	-	-	Ser-Lys-Leu
Cell Periphery	-	-	Cys- Ile- Ile- Cys

Fluorescent proteins previously synthesized from the Dueber lab were targeted to the five organelles (Table 2). Fluorophores were equalized in expression by varying promoter strength to ensure all fluorescent proteins were visible simultaneously. Five different promoters across five levels of magnitude were selected and characterized (Figure 3). Actin, vacuolar membrane, and cell periphery were placed under the ADH1 promoter while the nucleus—showing naturally stronger fluorescence—was placed under the control of a weaker promoter, CYC1. Equalizing the expression of fluorescent proteins ensured that one component of the MiCode did not mask any other fluorescent protein, further enabling accurate detection and identification (Figure 4).

Table 2. Fluorescent proteins used in the MiCoding system.

Name	Class of Fluorophore
Venus	Green (yellow) fluorescent protein
mCherry	Red fluorescent protein
PAmCherry	Photoactivatable red fluorescent protein
PAGFP	Photoactivatable green fluorescent protein
CFP	Cyan fluorescent protein

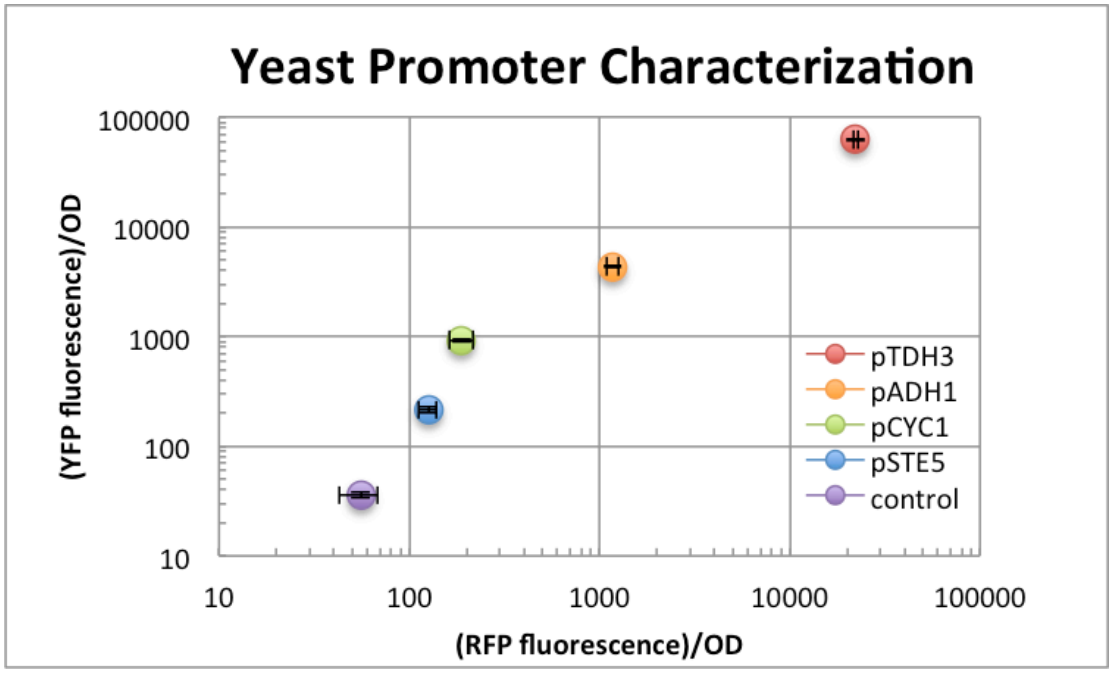


Figure 3: Characterization of different promoter strengths across five orders of magnitude. Actin, vacuolar membrane, and cell periphery were placed under the ADH1 promoter (orange) while the nucleus—showing naturally stronger fluorescence—was placed under the control of a weaker promoter, CYC1 (green). Error bars are shown in both axes.

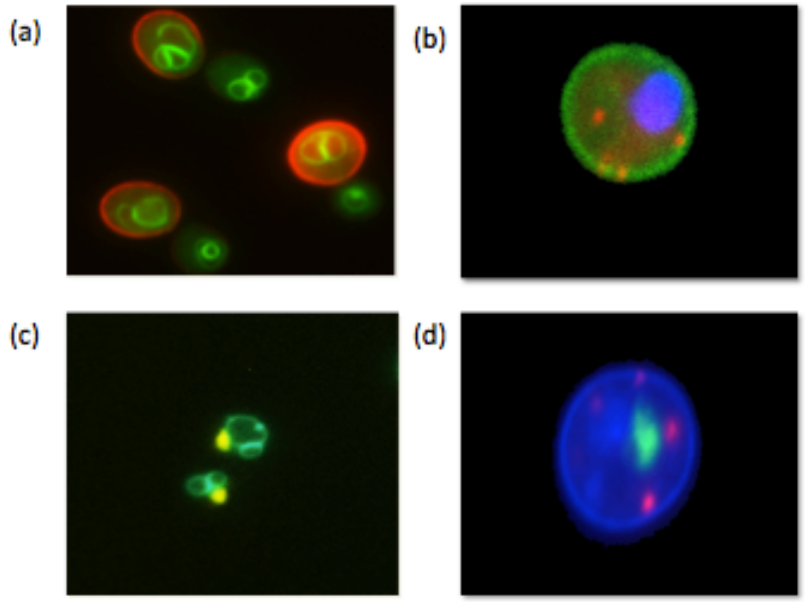


Figure 4. Promoter optimized MiCodes displaying even fluorescence across mutliple channels. Shown are: (a) red cell periphery and green vaculoar membrane; (b) green cell periphery, red actin, blue nucelus; (c) blue vacuolar membrane and green and blue nucleus; (d) blue cell periphery, blue vacuolar membrane, red actin, and green nucleus.

The MiCode assembly was completed by iterative Golden Gate cloning reactions using type II restriction endonucleases, allowing the formation of large constructs with emphasis on the interchangeability of parts (Engler et al. 2008). This method of assembly is powered by type IIs restriction enzymes which cut distal to their six base pair recognition site, allowing full control over the four base pair overhangs essential for control of plasmid assembly (Figure 5). This technique allowed us to create libraries of multiple parts in a one pot cloning step.

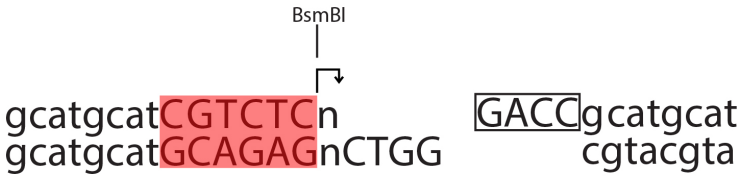


Figure 5. Type II restriction endonuclease, BsmBI, used in Golden Gate cloning. The restriction enzyme cleaves distal to the six base pair recognition site (highlighted in red) creating a GACC overhang. Control over the order in which plasmids assemble in a one pot reaction is created by altering the four nucleotide overhangs for complementarity between adjacent parts. Figure adopted from Engler et al. 2008.

The fundamental unit is the parts plasmid, containing a backbone with replication origin, selection marker, 5' upstream region, promoter, part, terminator, and 3' downstream region (Figure 6).

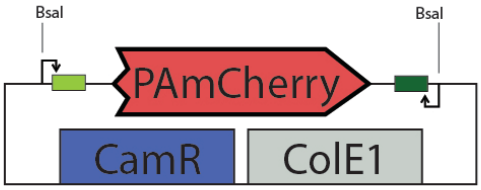


Figure 6. Parts plasmid. The plasmid used in the initial round of construction containing a backbone with origin and marker, 5' upstream region, promoter, part, terminator, and 3' downstream region. The 5' upstream and 3' downstream regions are the sites where the type II endonuclease cleaves allowing the part (PAmCherry in this case) to assemble to other parts in the one pot reaction.

Each parts plasmid can be joined together to make a cassette plasmid in a one-pot reaction. The overhangs designed within the parts plasmid dictate the correct order in which they are assembled in the one tube reaction to create the cassette (Figure 7).

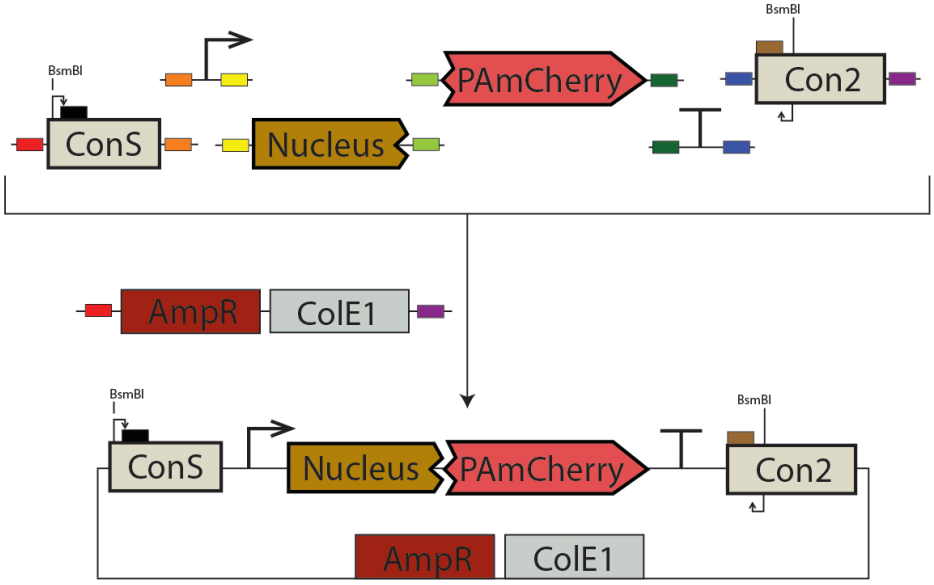


Figure 7. Cassette plasmid. A product of six individual parts plasmids fused by Golden Gate cloning. Each is comprised of a targeting protein fused to a fluorescent protein, in this case encoding a red photoactivtable nucleus.

Cassettes encode an organelle targeting protein fused with a fluorescent protein encoding localized fluorescence to the specific organelle. Because of this DNA physical linkage, the data for the MiCode will always be retained.

Case study: leucine zipper assay

Leucine zippers were used to test the MiCodes because of their vast application to many biological functions. Protein-protein interactions are used in structure, motility, and signal transduction and understanding their interaction affinity is essential in the engineering of such phenotypes. Leucine zippers are forms of coiled coils made of two alpha helices each with a heptad repeating pattern $(abcdefg)_n$ (Thompson et al. 2012). These coils interact with each other through the hydrophobicity of leucine residues occurring at the *a* and *d* residues while complementary electrostatic forces exist between the *e* and *g* positions. The leucine zipper sequences were computationally designed by the Keating Lab at MIT and biased towards orthogonality. The Keating Lab has created powerful software to predict protein interactions de novo, applying the technology towards designing these leucine zippers to specifically interact

only with its partner. For example, zipper 1A was predicted to specifically bind to zipper 1B acting as an orthogonally interacting pair, not interacting with itself as a homodimeric complex or any other member of the library.

The implementation of the leucine zippers into the MiCoding system involved combining the gene cassettes to form multigene cassettes that encode multiple organelle-targeted fluorescent proteins along with one leucine zipper of each pairwise interaction (Figure 8). Up to this point, all plasmid formation was deliberate and each construct's identity was known. By mixing the multi-gene cassettes together in a one-pot reaction, a library of leucine zippers has been created of all combinatorial interaction pairs (Figure 9). Due to the physical linkage of the MiCode with the leucine zipper, we can deduce with absolute certainty the genetic identity of the leucine pair observed in each cell by observing its fluorescent phenotype.

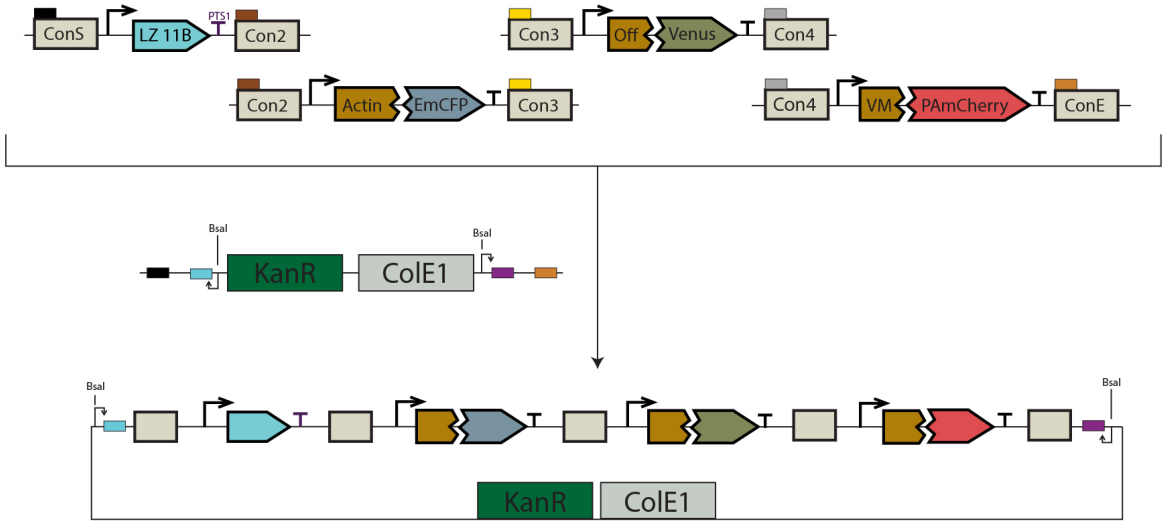


Figure 8. Multigene cassette. Product of fusing a MiCode with a leucine zippers. This multigene cassette encodes for a photoactivatable red nucleus, green cell periphery, an off blue channel, and prey leucine zipper 11A.

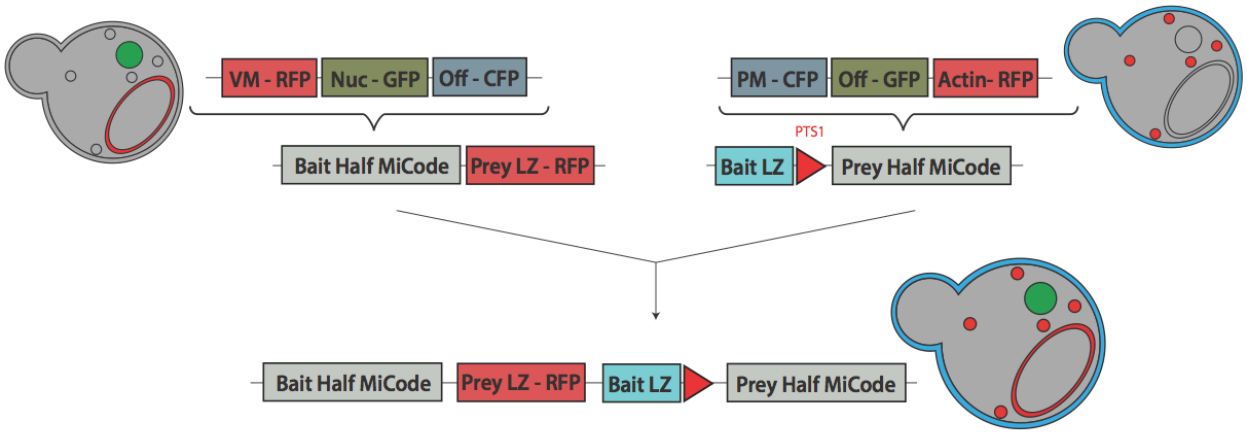


Figure 9. The complete MiCode assembly: combining the sets of 6 prey and 6 bait multigene cassettes in a one pot reaction to create a 36 member library

MiCodes were used to analyze a 36 member leucine zipper library of interaction strengths through a bait and prey targeting assay to determine their affinities and test for orthogonal interactions. We used a protein fragment complementation assay, a system of bait and prey leucine zippers (Ghosh et al 2000). The bait leucine zipper protein was tagged to a peroxisome targeting sequence (PTS1) and the prey protein was fused to red fluorescent protein (RFP). Our measurement of interaction strength was accomplished by the red pixel intensity of the peroxisome relative to that in the cytosol. The higher the pixel intensity localized at the organelle relative to the background fluorescence correlated to larger interaction strength because of the increased complementation between bait and prey zipper. Three leucine zippers pairs across a range of known affinity (1-10 nM for interaction, 1 uM for no interaction) were used to establish the dynamic range of the assay, acting as a control to compare those of unknown affinity.

The strong leucine zipper interaction pair displayed red punctate fluorescence in the peroxisomes of the yeast cells, the medium interaction pairs showed some concentration of RFP in the peroxisomes while also diffuse color in the cytosol, and the weak interaction pair only yielded diffuse red fluorescence in the cytosol (Figure 10).

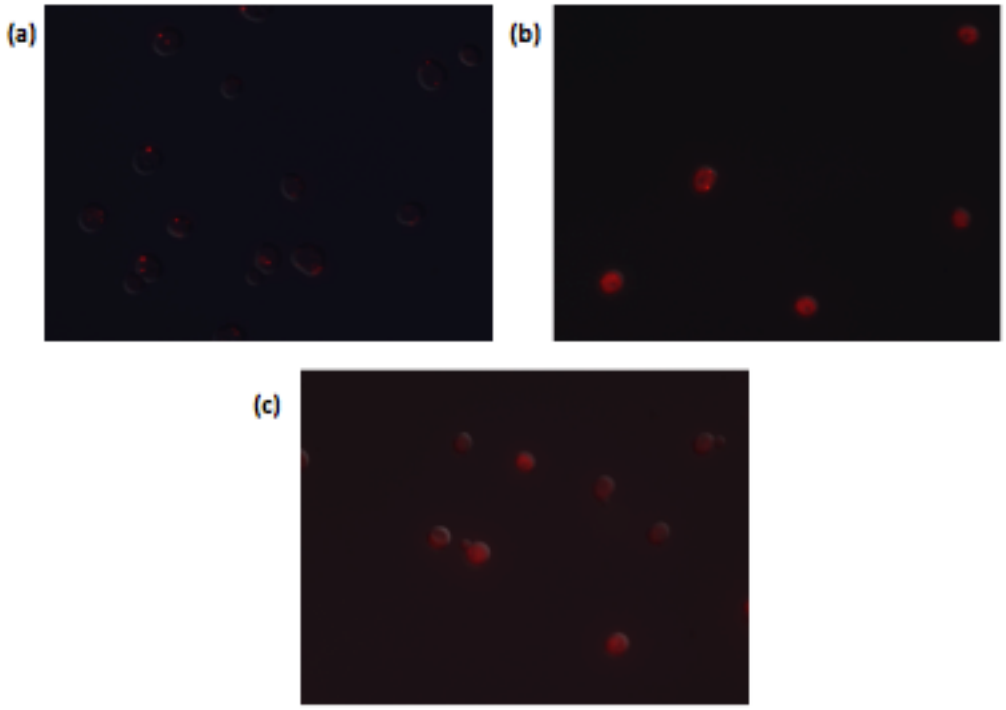


Figure 10. Bait and prey assay for strong, medium, and weak leucine zipper interactions of known affinity. (a) strong interacting showing punctate fluorescence in the peroxisome (b) medium strength and (c) weakly interacting leucine zippers shown by diffuse red fluorescence in the cytosol.

Testing the leucine zipper assay showed most significant differentiation between strong and weak as well as strong and medium interaction strength ($p<0.01$) (Figure 8). Medium and weak strength did not show significant distinctiveness ($p<0.08$). The functional range of the assay was also established to extend minimally to 500x brightness of the peroxisome relative to the cytosol to an upper limit of nearly 3500x pixel intensity.

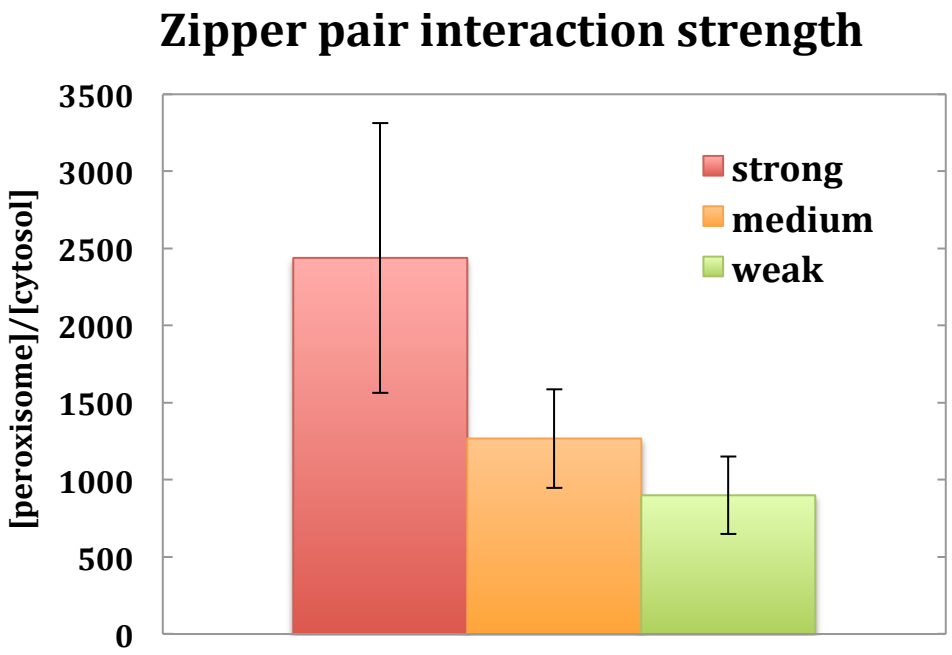


Figure 11. Fluorescence of peroxisome using bait-prey assay relative to the cytosol of known interaction strengths. Both the strong (n=10) and weak (n=11) interactions as well as strong and medium (n=19) interactions are statistically distinct ($p<0.01$) from one another. The medium and weak interactions are not significantly distinct ($p<0.08$).

Having established the functional range of our assay with previously characterized data, we applied the same system to form 6x6 matrix of 36 unknown leucine zipper interactions amongst the Keating leucine zippers 1A, 1B, 11A, 11B, 16A, and 16B (Figure 12). Three teams of two biologists completed MiCode analysis. Each member was educated with characteristics of each organelle targeted with the three fluorescent proteins to minimize variability between operators. MiCoded cells were removed from the data analysis where the negative control failed (when the PAGFP on the bait zipper showed the peroxisome out of focus in the imaging plane) and/or the MiCode identity was not agreed upon by pair of biologists. Those MiCodes that passed such criteria were included in the interaction matrix. The 6x6 heat map was arranged detailing the interaction affinities for every combination of zipper pairs (Figure 13). Each box provides the number of cells with a visibly strong interaction over the number of total cells observed. Boxes in grey are interactions are ones not seen in the set or which failed to fulfill the controls mentioned above.

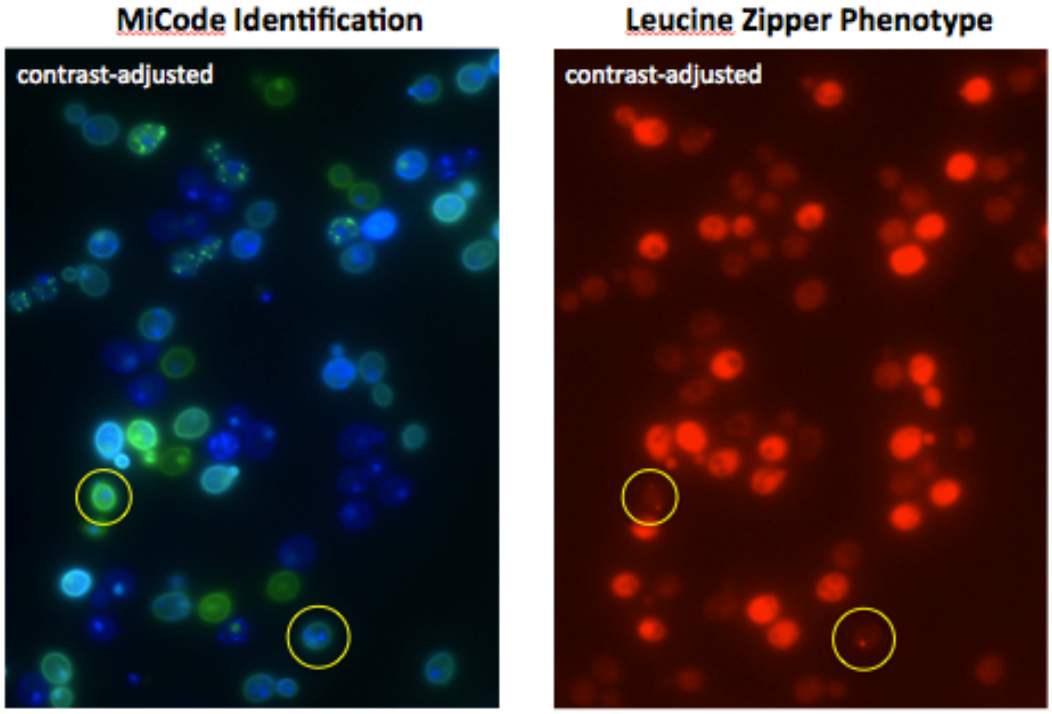


Figure 12. 36 member leucine zipper library with MiCodes. The cells encoding various combinations of leucine zipper interactions in the 36 member library. The green and blue fluorescent channel (left) reveals the MiCode while the red fluorescent channel shows the protein interaction strength as shown through the bait and prey assay.

Prey \ Bait	1A	1B	11A	11B	16A	16B
1A		4/4	0/4	0/1	3/12	0/6
1B	3/4		0/1		0/8	
11A	0/1		0/8		1/8	
11B				1/1	0/4	
16A	0/5		0/7	0/4	0/7	0/2
16B			0/4	0/2	0/1	0/1

Figure 13. The 6x6 interaction matrix of the 36 combinatorial interactions of leucine zippers. The cells account for cells with visible interaction / cells observed.

Automation

To enable MiCodes as a high throughput technology, Matlab and CellProfiler script were developed to make MiCode identification and analysis automatable. MatLab (developed by MathWorks) was utilized for image acquisition, the partitioning of individual MiCoded cells into separate images, while script in CellProfilier (a project based at the Broad Institute Imaging Platform) was created for image analysis, to detect each of the organelles and its associated fluorescent color. In conjunction, the two platforms form a pipeline for automated MiCode identification (Figure 14). All images fed into the software programs were taken using a basic fluorescence microscope.

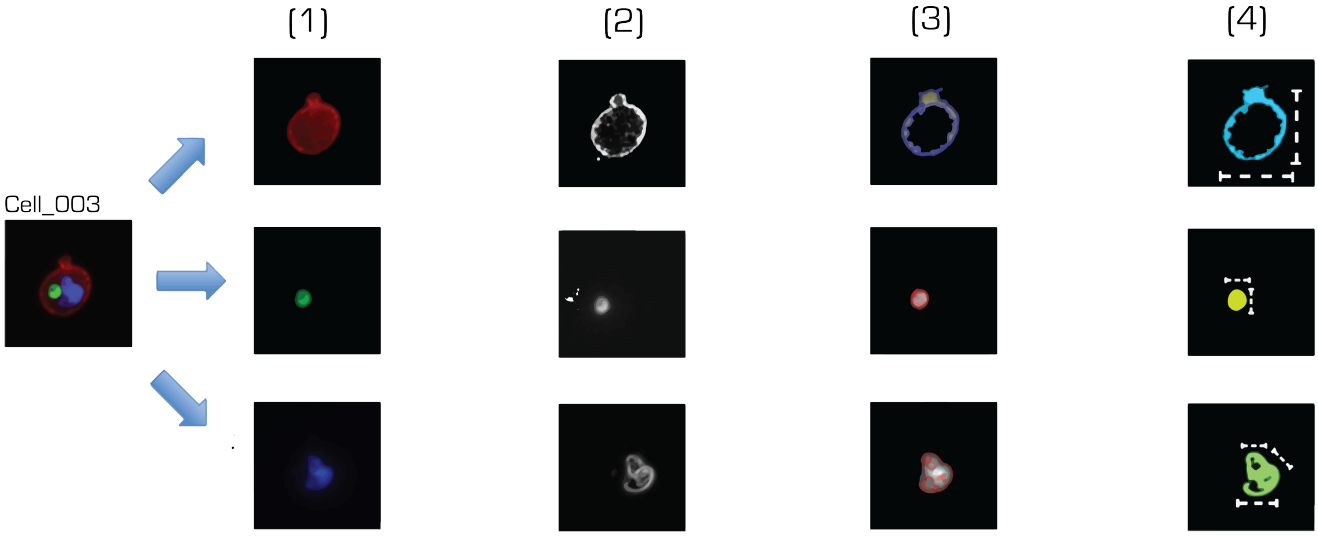


Figure 14: Automation pipeline procedures: [1] Cell Profiler acquires single cell images of each fluorescent channel. [2] Cells are converted to gray scale and a thresholding algorithm is applied to remove background pixels while retaining ones of interest [3] Distinct regions of high intensity pixels are determined [4] Geometric and intensity measurements for these objects are determined.

Developing the MiCode automation is essential to scale this for its application to million-member libraries. The image acquisition script written in MatLab involves cell segmentation, the recognition of individual cells from their background. We used the gradient approximations from the Sobel operator and many filtering options to estimate the outlines of individual yeast cells. Next, the background pixels were cleared using consecutive erosion and dilation steps. A refining algorithm was created using data of 400 yeast cell images to create geometric criteria to

detect cells with greater precision. Using images with minimal clumping, the software can segment 96% of cells accurately (Figure 15).

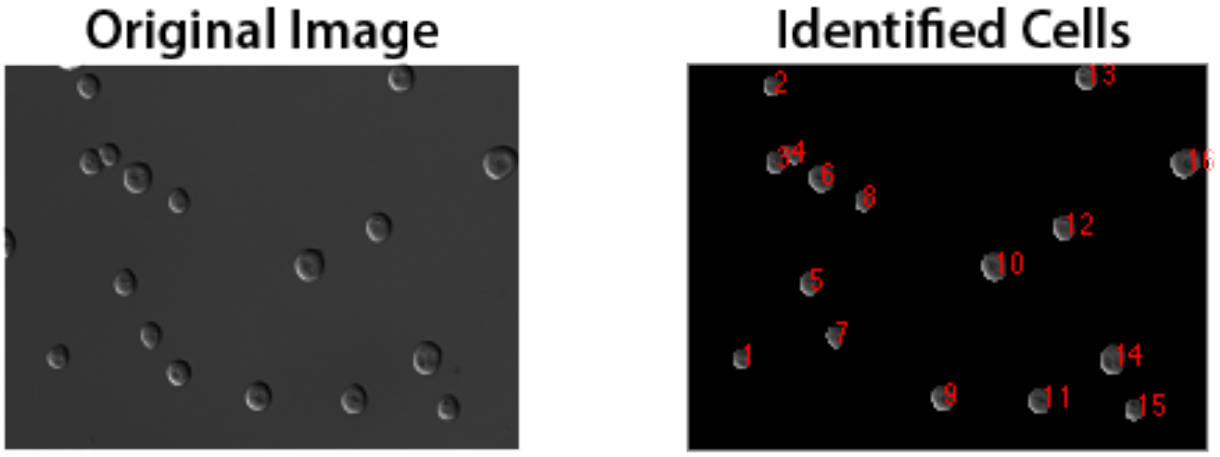
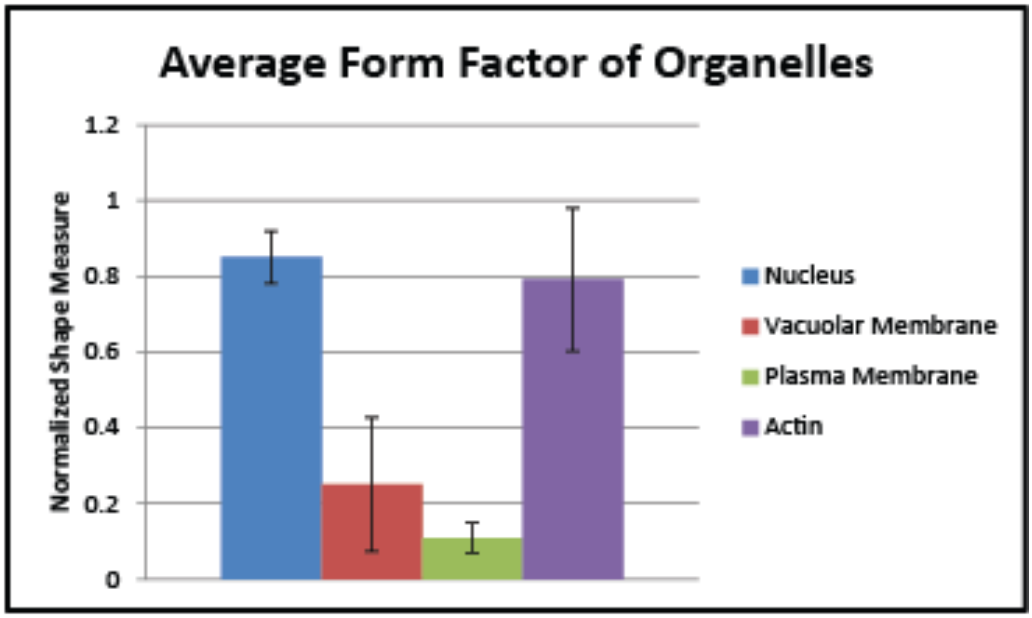


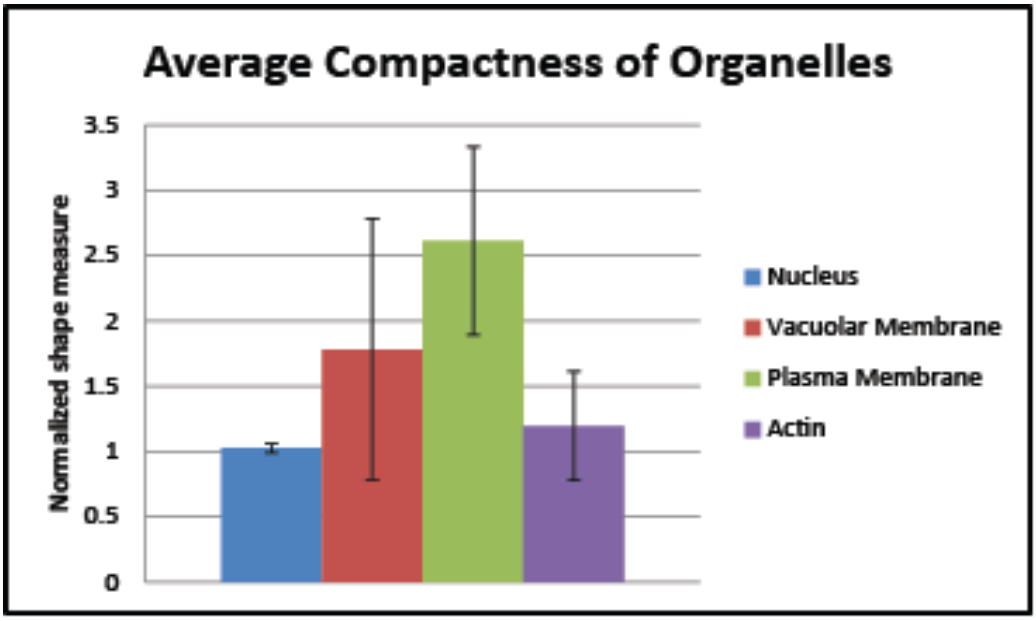
Figure 15. Image Acquisition. Applying the MatLab cell acquisition script identifies each of the yeast cells and separates them into individual images for use in CellProfiler.

As part of the analysis, organelle detection script in CellProfiler was written for computer identification of the MiCode organelles based upon the geometric configurations of each organelle, providing quantitative analysis of the fluorescent phenotype. The four organelles were explicitly chosen for the ease in computer detection based upon size and shape: nucleus (large circle), actin (small circle), cell periphery (large ring), and vacuolar membrane (small ring). The identities of organelles in each image were identified using the parameters from quantitative differences of size and shape unique to the organelle (Figure 16). Setting cutoffs in each of these parameters provided robust characterization of the organelles for MiCode automation.

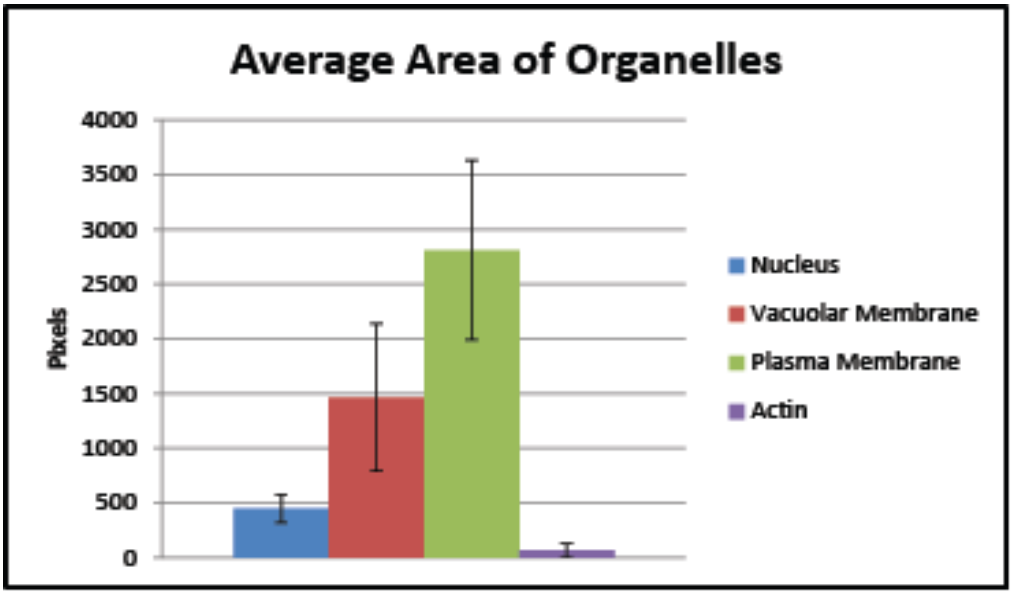
(a)


Average Form Factor of Organelles
Normalized Shape Measure
1.2
1
0.8
0.6
0.4
0.2
0
Nucleus
Vacuolar Membrane
Plasma Membrane
Actin

(b)


Average Compactness of Organelles
Normalized shape measure
3.5
3
2.5
2
1.5
1
0.5
0
Nucleus
Vacuolar Membrane
Plasma Membrane
Actin

(c)



(d)

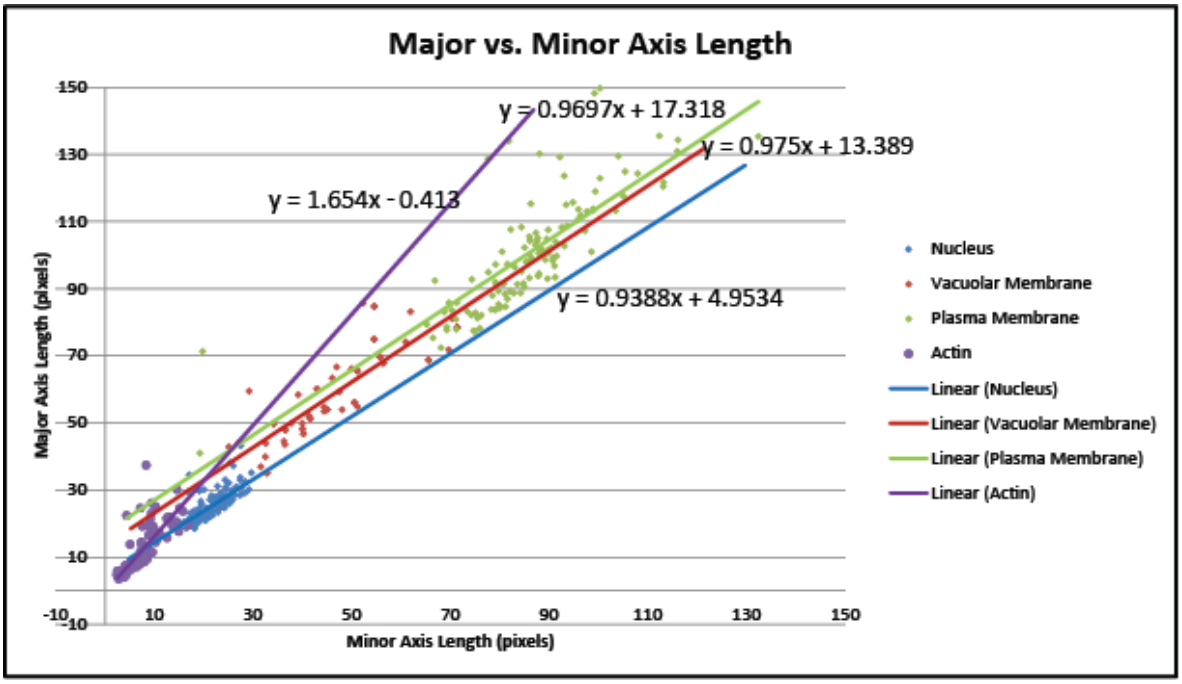


Figure 16. The geometric data of organelle size and shape used in MiCode automatable detection and analysis.

To determine the accuracy of computer analysis, a mixed population of 250 cells comprising three distinct MiCodes were imaged and analyzed using the automation pipeline in conjunction with biologist identification (Figure 17). The images were run through MatLab cell segmentation and then CellProfiler for MiCode construction. The results were compared with that of human eye detection to determine the accuracy of the script. Comparing MiCode

identification by eye and by computer showed 90% accuracy of automated nucleus detection in the green/YFP channel. However, in the red channel, the computer detected 40 nuclei in comparison to zero identified by the biologists, thereby showing false negative detection by the automation software.

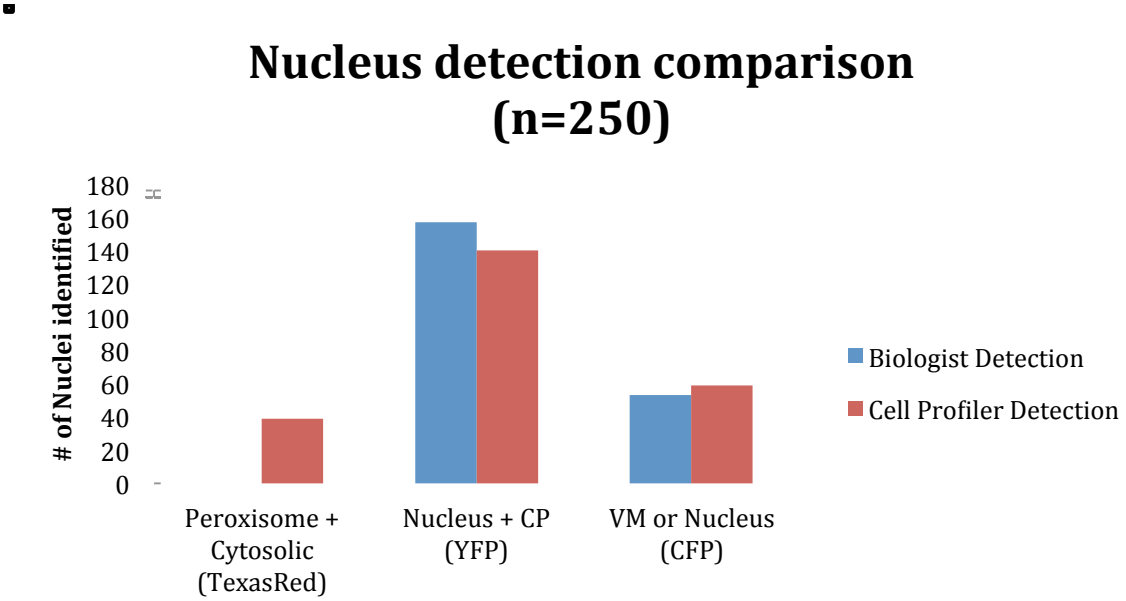


Figure 17. Human and computer nucleus identification. Data gathered from one trial of a mixed population of 250 cells across three MiCodes.

DISCUSSION

MiCodes

MiCodes successfully link the genotype of a library member with an observable phenotype, conferring microscopy as a practical tool to screen libraries of subcellular phenotypes. This experiment displayed the successful creation of a MiCode set that was applied to determining orthogonal pairings of a 36 matrix of leucine zippers. Computer script developed to automate the MiCoding process proves the feasibility of MiCodes as a high throughput microscopy screening technology.

Using Golden Gate cloning enables one to scale up the MiCoding system to variable library sizes of interest; however, the number of possible MiCodes is limited by the accuracy of organelle identification. The current theoretical library size is represented by the function 2^{xy} : *2*

representing the number of states of the fluorescent protein (on or off), x as the number of distinct fluorescent proteins, and y representing the number of organelles to which the fluorescent proteins are targeted (Figure 1). In our experiment, the use of three fluorescent proteins and four organelles yielded 5,096 possible MiCode combinations. However, we were limited in the accuracy of human organelle detection (62%-76% range depending on the biologist team) using a standard fluorescence microscope (Zeiss AxioCam MRm).

To undergo such large scale cloning, the Golden Gate assembly scheme was necessary to construct the MiCodes on a feasible timescale. Rather than completing separate and sequential cloning reactions for each MiCode, Golden Gate assembly occurs in a simultaneous, one pot reaction (Engler et al. 2008). Since the MiCodes and the leucine zippers are linked on the DNA level and created are in a known fashion when making the cassette (Figure 9), their identity can be retraced after mixing them in a random combination through the Golden Gate assembly. Using this cloning scheme is critical for expansion beyond the limited set of MiCodes used in this experiment because it enables the feasible construction of upwards of 10^6 unique plasmids.

Leucine zipper assay

Applying MiCodes to a set of leucine zippers of unknown affinity allowed high throughput detection of both potential orthogonal sets as well as non-orthogonal/non-exclusive binders using fluorescence microscopy. We predicted shading between each reciprocal leucine zipper pair (1A-1B, 11A-11B, 16A-16B) of the interaction matrix if the leucine zippers indeed displayed orthogonal interactions. As shown in Figure 13, leucine zippers 1A and 1B appear to interact strongly as the Keating Lab had predicted. However, their potential to orthogonal binders is reduced by noting zipper 1A's interaction with 16A. The numerous non-orthogonal pairs (1A-16A and 11A-16A) prove to be weak binding pairs that are variable in their interaction affinity and non-exclusive in their binding interactions. Interaction of 11B with itself was not predicted; the leucine zippers were created to bind as heterodimers with their partner and unable to act as homodimers with another identical protein. Gray boxes are interactions not seen in this round of data collection either due to the fact that the MiCode identity was not agreed upon by the team of biologists or failure of the negative control (PAGFP localized in the peroxisome was out of focus therefore accuracy could not be guaranteed from the image). The single round of

data collection does not provide a means for statistical analysis, but after multiple round of collection with larger sample sizes, zipper interactions can be tested for the significance of orthogonal binding. These sets will be sent back to the Keating Lab for further development using their computational design software thereby optimizing the orthogonal leucine zipper set. After multiple iterative rounds of designing and testing, we aim to produce an extensive set of orthogonal proteins, acting as a tool kit for biofuels production in synthetic biology.

Automation

Automation of the MiCoding system enables this technology to be practical with library sizes upwards of one million members without losing accuracy. Showing a 90% accuracy of computer detection for nuclei, we believe that script optimized for other organelles can be equally successful. To address the high false negatives, we chose to loosen the thresholding parameters in organelle identification. MiCodes can be constructed by a human easily with Golden Gate cloning while the more time intensive analysis can be automated by a computer. We estimate that with an average of twenty five cells per image and a library of 10^5 cells, the 4,000 cell microscopy images could be taken in three hours. With the automation processing speed averaging 30 seconds per image, the analysis of 4,000 images would require 33 hours. In total, automated MiCode processing would take an estimated 36 hours for a set of 10^5 cells.

Advantages over current methods

MiCodes enable one to screen for subcellular localization, cell motility, and cell morphology as well as intermediate phenotypes that cannot be categorized in a binary system, two abilities that no other competing screening technology can provide. FACS (fluorescence activated cell sorting) is a competing technology that is able to sort fluorescent cells depending on set ranges of fluorescence (Cormack et al. 1996). However, this technology is limited by its binary selection process that can only measure whole-cell fluorescence. MiCodes enable the use of many subcellular fluorescence measurements on a dynamic range rather than a binary system. Yeast Two Hybrid systems have also been designed for similar purposes (Walhout and Vidal 2001), however these measurement systems are a function of the transcriptional and translational

output which can influenced by many additional factors in the cell. MiCodes provide a direct measurement of fluorescence in this case as a function of protein binding strength.

Limitations and future directions

The MiCoding system has a theoretical limit of the number of organelles one can accurately discern within the cell and is prone to error with recombination between multiple MiCode plasmids sharing similar homology. The four organelles in the experiment were chosen for their uniqueness in size and shape. The ability of the human or computer observer to accurately identify the fluorescent organelles in the cell places a theoretical limit on the number of MiCodes that one can produce. With more numerous cassettes comes the duplication of promoters, terminators, and fluorescent protein sequences, posing the threat of recombination with high degrees of similarity. Because the known DNA link at the cassette level is essential to backtrack from fluorescent barcode variable DNA sequence of interest (in our case leucine zipper sequence), recombination and mixing of these sequences would destroy identification reliability and the ability to know the retrace back to the cell of interest's identity. While recombination does pose a threat, it can be overcome with codon variability (alternative three nucleotide sequences that code for identical amino acids) amongst the cassettes especially in promoter sequences. Using varying promoters for the fluorescent proteins reduces homology within MiCode sequences thereby suppressing genetic recombination and ensuring that all MiCodes remain accurately associable with their leucine zipper sequence.

To scale up the MiCodes technology, improving upon the discernibility of fluorescent organelles and decreasing analysis time is essential. With the current expansion of synthetic biology and use of libraries entering ranges of 10^6 and beyond, we foresee avenues to further expand the theoretical library sizes of MiCodes. Use of photoactivatable fluorophores creates an additional third state beyond simply "on" or "off" forming a set MiCodes that is represented by the function 3^{xy}. Since the distinguishability of the organelles acts as the ceiling on the number of MiCodes that can be created, using a confocal microscope (whereby images are taken in multiple planes throughout the depth of the cell) can create a more accurate image to analyze. Organelle staining can also be used to discern different organelles from one another to increase reliability when expanding to a larger set of organelles.

Speeding up the analysis is also an area that can be targeted to scale up the MiCodes technology. Use of a motorized stage in microscopy can eliminate the human labor from working the microscope. Using machine learning for the automation process also holds promise. Instead of providing the computer with human created script, the machine would be provided with known sets of MiCodes and their identities from which it can draw its own rules in order to categorize and identify MiCodes of unknown identity.

Broader implications

MiCodes present many applications from areas of medicine to alternative energy; most significantly, they provide a means to more efficiently engineer synthetic metabolic pathways in microbes for biofuels production. Tumors are composed of heterogeneous sets of cells with many different genotypes. MiCoding cells in a tumor could identify cells of distinct genotypes within the mass, providing information about their generation and growth for accurate medical therapeutics. MiCodes can also be applied to gene identification within cells containing unknown functions of open reading frames. Mutagenesis in MiCoded cells and screening for their phenotypes under exposure of different growth conditions (pH, temperature, nutrients, etc) could provide insight to the function of such genes. The most exciting application of MiCodes is related to engineered microbial biofuels (Dueber et al 2009, Lee et al 2012). If one wanted to create the largest organelle possible in which to house the biofuels factory, they could place the genes in the organelle biosynthesis pathway under different combinations of promoters. This combinatorial pattern works well with MiCodes when the researcher does not want to work out each individual combination separately. As addressed in our experiment, MiCodes show promise for finding orthogonal protein interactions, an essential tool for synthetic biology as more complex genetic circuits are being designed. Microbial biofuels production requires precision and command over engineering the complexity of biology, an ability that MiCodes enables for biological researchers.

Conclusion

As a high throughput fluorescence microscopy technique that covers the limitations of other competing technologies, MiCodes is an efficient tool when studying visible phenotypes including subcellular localization, motility, and morphology that are essential for understanding and engineering biological properties. MiCodes act as a high throughput library screening technology that can be automated and scaled for libraries of 10^6. Applying MiCodes to improve microbial biofuels production will move alternative energy production away from corn-based ethanol and other agricultural biofuels. MiCodes augment synthetic biology's role in the production of microbial biofuels: a more sustainable, efficient, and environmentally conscious source of energy.

ACKNOWLEDGEMENTS

Thank you Team ES 196, especially Tina and Anne, for their continued support, insight, and direction throughout the past year. Your willingness to meet in office hours, provide a non-engineering perspective, and continuously grade drafts of the thesis have been invaluable. Thank you Mom and Dad for fostering an inquisitive mind and love of science from an early age that has culminated in my pursuit in the field. I would like to thank the Dueber Lab for hosting myself and the rest of the iGEM team and especially Dr. John Dueber, Will DeLoache, and Terry Johnson for their advising and insight throughout this project. Thank you to the Keating Lab for providing the leucine zippers and willingness to collaborate. Thank you Agilent Technologies for the majority of the financial support as well as the National Science Foundation, Integrated DNA Technologies, and the Synthetic Biology Engineering Research Center.

REFERENCES

Chalfie, M., Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher. 1994. Green fluorescent protein as a marker for gene expression. Science 263:802–805.

Cohen, B.A., P. Colas, and R. Brents. 1998. An artificial cell-cycle inhibitor isolated from a combinatorial library. PNAS 95: 14272-14277.

Cormack, B. P., R. H. Valdivia, and S. Falkow. 1996. FACS-optimized mutants of the green fluorescent protein (GFP). Gene 173:33–38.

Dueber, J. E., G. C. Wu, G. R. Malmirchegini, T. S. Moon, C. J. Petzold, A. V. Ullal, K. L. J. Prather, and J. D. Keasling. 2009. Synthetic protein scaffolds provide modular control over metabolic flux. Nature Biotechnology 27:753–759.

Dugar, D., and G. Stephanopoulos. 2011. Relative potential of biosynthetic pathways for biofuels and bio-based products. Nature biotechnology 29:1074–1078.

Engler, C., R. Gruetzner, R. Kandzia, and S. Marillonnet. 2009. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. PLoS ONE 4:e5553.

Fargione, J., J. Hill, D. Tilman, S. Polasky, and P. Hawthorne. 2008. Land Clearing and the Biofuel Carbon Debt. Science 319:1235–1238.

Ghosh I, A.D. Hamilton, L. Regan. 2000. Antiparallel leucine zipper- directed protein reassembly: application to the green fluorescent protein. J Am Chem Soc 122: 5658–5659.

Haynes, K. A., and P. A. Silver. 2009. Eukaryotic systems broaden the scope of synthetic biology. The Journal of Cell Biology 187:589–596.

Hong, K.-K., and J. Nielsen. 2012. Metabolic engineering of Saccharomyces cerevisiae: a key cell factory platform for future biorefineries. Cellular and Molecular Life Sciences 69:2671–2690.

Huang, J., J. Yang, S. Msangi, S. Rozelle, and A. Weersink. 2012. Biofuels and the poor: Global impact pathways of biofuels on agricultural markets. Food Policy 37:439–451.

Thompson, K.E., C.J. Bashor, W.A. Lim, and A.E. Keating. 2012. SYNZIP protein interaction toolbox: in vitro and in vivo specifications of heterospecific coiled-coil interaction domains. ACS Synth Biol 1(4): 118–129.

Keasling, J. D., and H. Chou. 2008. Metabolic engineering delivers next-generation biofuels. Nature biotechnology 26:298–299.

Lee, H., W. C. DeLoache, and J. E. Dueber. 2012. Spatial organization of enzymes for metabolic engineering. Metabolic Engineering 14:242–251.

Lutz, W., W. Sanderson, and S. Scherbov. 2001. The end of world population growth. Nature 412: 543-545.

Pepperkok, R. & Ellenberg, J. High-throughput fluorescence microscopy for systems biology. *Nature Reviews Molecular Cell Biology* **7**, 690–696 (2006).

Peralta-Yahya, P. P., F. Zhang, S. B. del Cardayre, and J. D. Keasling. 2012. Microbial engineering for the production of advanced biofuels. Nature 488:320–328.

Reinke, A. W., R. A. Grant, and A. E. Keating. 2010. A Synthetic Coiled-Coil Interactome Provides Heterospecific Modules for Molecular Engineering. Journal of the American Chemical Society 132:6025–6031.

Rodríguez, A., H. A. Cristóbal, and V. L. Colin. 2011. The Role of Synthetic Biology in the Design of Microbial Cell Factories for Biofuel Production. Journal of Biomedicine and Biotechnology 2011:1–9.

Thompson, K. E., C. J. Bashor, W. A. Lim, and A. E. Keating. 2012. SYNZIP Protein Interaction Toolbox: in Vitro and in Vivo Specifications of Heterospecific Coiled-Coil Interaction Domains. ACS Synthetic Biology 1:118–129.

Walhout, A. J. M., and M. Vidal. 2001. High-Throughput Yeast Two-Hybrid Assays for Large-Scale Protein Interaction Mapping. Methods 24:297–306.

Zanta, M. A., P. Belguise-Valladier, and J.-P. Behr. 1999. Gene delivery: A single nuclear localization signal peptide is sufficient to carry DNA to the cell nucleus. Proceedings of the National Academy of Sciences 96:91–96.

Zhu, J. Y., and X. S. Zhuang. 2012. Conceptual net energy output for biofuel production from lignocellulosic biomass through biorefining. Progress in Energy and Combustion Science 38:583–598.

APPENDIX A: Golden Gate Cloning

Contents

(8-0.5*number of plasmids) uL ddH20
1uL T4 DNA ligase buffer
0.5 uL T4 DNA ligase
0.5 uL Bsa1 Restriction enzyme
0.5uL of each plasmid part

Thermocycler incubation protocol

2 minutes at 37°C
5 minutes at 16°C
repeat steps 25 times
5 minutes at 50°C (final digestion)
5 minutes at 80°C (heat inactivation).