DISCOVERY FRAMEWORK FOR LIVE CELL SYNAPTIC VESICLE RECYCLING ASSAYS

Hansang Cho[1], Samuel V. Alworth[1], Seho Oh[1], Yuki Cheng[1], Brooke G. Kelley[2], Greg Martin[2,3], Jane M. Sullivan[2] and James S.J. Lee[1]. [1]SVision LLC, 3633 136th Pl. SE, Suite 300, Bellevue, WA 98006. [2]Dept. of Physiology and Biophysics, and [3]Keck Imaging Center, University of Washington, Seattle, WA 98195.

Introduction

Automated quantitative analysis of live cell assays with complex endpoints, such as the vesicle recycling assay whose endpoints are revealed over time, could provide a powerful tool for basic research applications. However, it can be difficult to see subtle differences between assay test and control populations. The populations themselves may contain an ensemble of phenotypes that are obscured in the mixture.

We have previously developed robust image and data analysis[1,2] methods that can automatically assay the exocytotic recycling of synaptic vesicles labeled with fluorescent FM dyes in time-lapse microscopy images. In this study, we evaluated a novel discovery framework that could reveal the subtle spatial-temporal phenotypic differences, if any, between assay populations. Our results show that the tool can provide an effective way to uncover subpopulations containing distinctive characteristics in the image data.
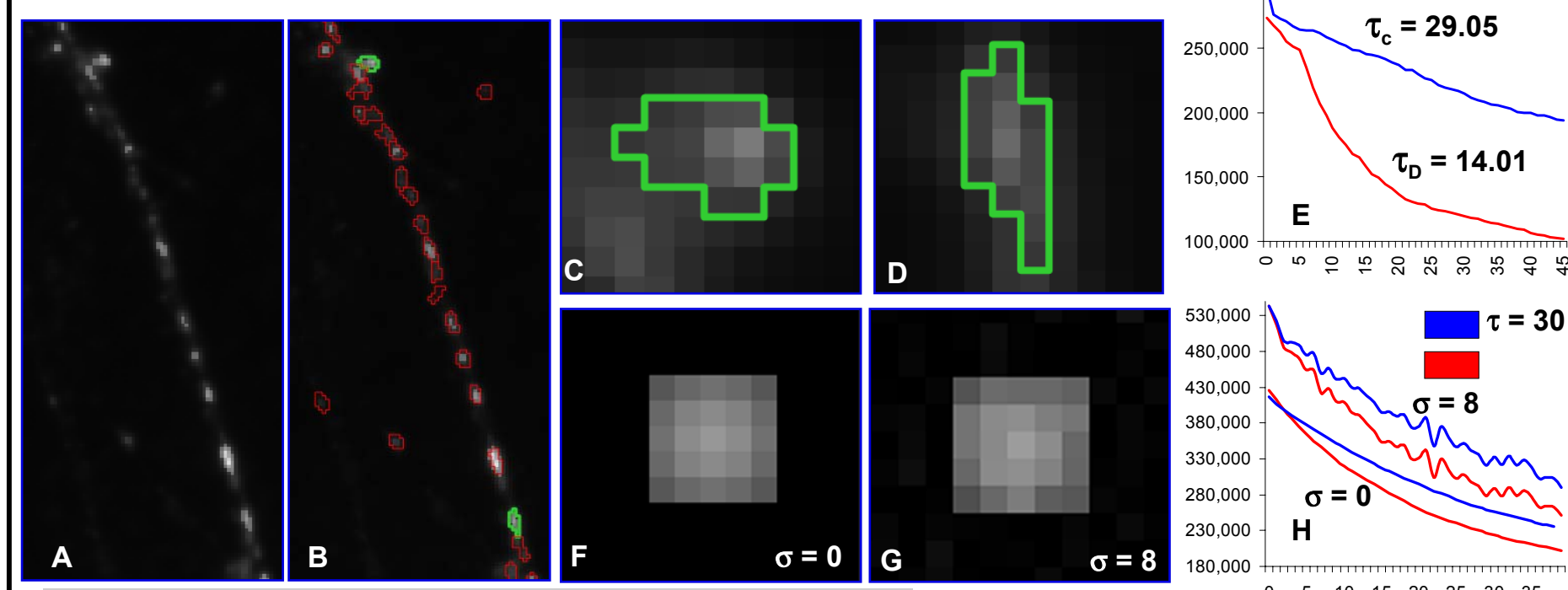
Methods

Fig. 1. Real and simulated data sets were used for this study

8 image sets from a synaptic vesicle recycling assay using fluorescent FM dye are used for this study. The time-lapse images show the destaining of individual synapses of a small number of rat hippocampal neurons in microisland culture. In 4 images the neurons are over-expressing the Amyloid Precursor Protein and GFP reporter (APP), and in the other 4 images the neurons are expressing GFP alone (GFP). Each image set had 65 images acquired at a rate of 1 Hz. In Fig. 1, (A) shows a representative image where the axon terminals are clearly labeled with FM dye along a single process, and (B) shows the same image with synaptic boundary definitions from our robust image analysis methods overlain. (C) and (D) show individual synapses on this same neuronal process that, as shown in (E), have very different destaining phenotypes; characterized here with the exponential dissociation (decay) model parameter τ. Additionally, two groups of simulated images with synthetic destaining phenotypes were created. Each simulated time-lapse image has 40 frames. The first group contains 512 synthetic synapses that destain according to the model with a τ value equal to 25. The second group contains 512 synthetic synapses with an equal mixture of τ equal to 20 and 30. Gaussian noise of $\sigma = 8$ is added to the simulated images prior to τ measurement. (F), (G) shows a representative synthetic synapse before and after noise addition, and (H) shows representative destaining individual object destaining curves prior to and after the addition of Gaussian noise.
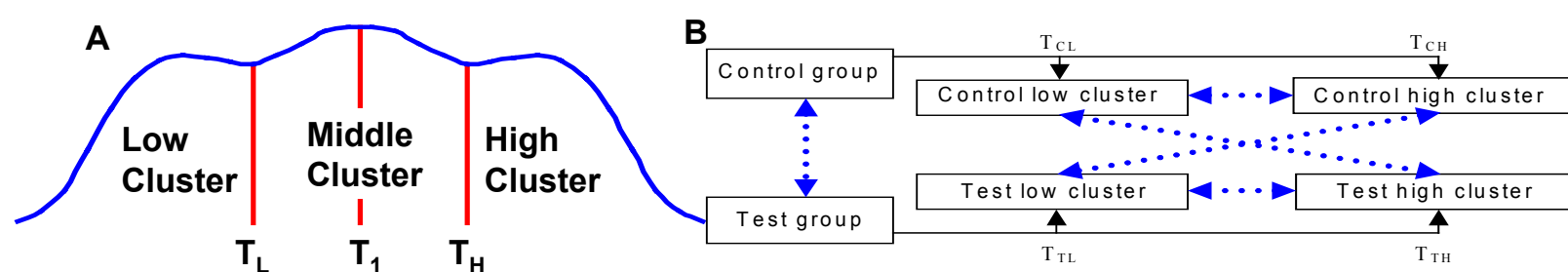

Fig. 2. A Discovery Framework was Created to Reveal Latent Populations

Automatic detection and fitting enables measurements of a large number of objects. The large sample size could reveal latent populations leading to new discoveries. We developed a discovery framework that automatically generates multiple sub-population clusters in each group of data. This allows the analysis of the cross cluster relations using assay quality measurements to reveal latent phenotypes.

Fig.2 (A) shows an example distribution of a measurement such as τ. The discovery framework uses an automatic population division algorithm to divide the measurement distribution into sub-population clusters. The algorithm divides the distribution into two, and repeats the process for each subsequent division. The algorithm automatically determines the threshold that maximizes the value: $(N_L \times m_L^2) + (N_H \times m_H^2)$, where N_L and N_H are the sample counts of the low and high sides of the threshold, and m_L^2, m_H^2 are the second order moments on the left and right sides of the threshold. Fig.2 (A) shows the example distribution is divided into four subpopulations using three thresholds; T_1, T_L, and T_H, the initial, low and high thresholds. We combined two subpopulations to form middle cluster.

The current framework compares distributions from one test and one control group. It divides each group's measurement distribution into three clusters, and then uses the low and high cluster from each distribution to compare across experimental groups as shown in Fig.2 (B). The comparison is made using the assay quality metric, "Z factor"[3-5]: $Z = 1 - \frac{(3\sigma_s + 3\sigma_c)}{|\mu_s - \mu_c|}$, where 'c' indicates control and 's' test samples. It provides a simple measurement of assay quality by characterizing the assay measurement dynamic range between test and control conditions, and the data variation associated with the measurements. A higher score is better (1 max).

Results

Simulated Image Results

Fig. 3 shows the 20/30 τ mixture image set having two peaks in the τ fitting result distribution, while 25 τ image set has single peak in τ fitting result distribution. The results after applying assay quality measurements for the whole sets and sub-clusters are listed in Table 1. Note that the Z factor between the two groups show no population separation. Yet the Z factor show distinctive characteristics between the inter population separation τ=20/30L vs τ =25H, and the intra population separations τ=20/30L vs τ=20/30H. The difference also exists for the inter population separation τ=20/30H vs τ = 25L. The simulated data confirms our approach.
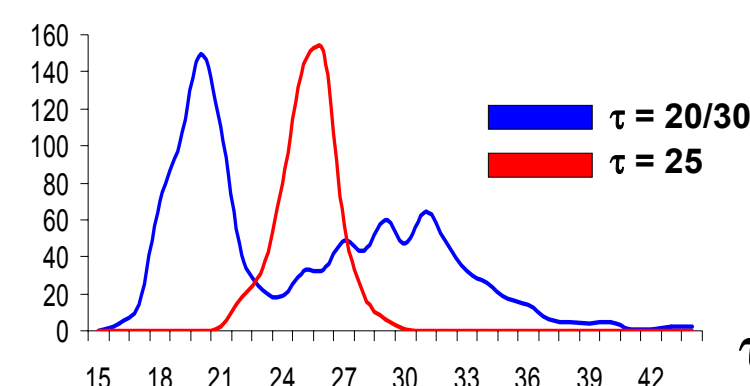

Fig. 3. Histogram of individual object τ for two simulated populations

Table 1. Z factors based cluster comparison using τ for two simulated populations

	20/30 vs 25	20/30(L) vs 20/30(H)	20/30(L) vs 25(H)	25(L) vs 25(H)	20/30(H) vs 25(L)
Z factor	-53.24	0.6889	0.4709	-5.0299	0.0892

Real Image Results

The 8 images consist of 4 pairs, each pair has an APP and a GFP image. The Z factors across different groups of populations are calculated for the 4 pairs combined, and each pair individually are listed in table 2. The Z factors show no distinctive difference across groups for the combined populations, the 121305a-c pair and the 121505c-a pair. The Z factors show possible characteristic differences for the 031506a-g pair and the 031906a-e pair. Fig. 4 shows the empirical τ distribution for the combined pair data and the 031506a-g pair. Fig. 4(A) shows that the APP and GFP distributions are highly overlapped for the combined pair. Fig. 4(B) shows that there is a sub-group of GFP objects with significantly lower τ values. The empirical data confirms the results from the discovery framework.

Table 2. Z factor results for the APP vs. GFP total set and image pairs

Z factor	4 pairs combined	121305a-c pair	121505c-a pair	031506a-g pair	031906a-e pair
APP_all/ GFP_all	-40.8105	-29.4589	-43.1911	-16.0978	-11.2379
APP_low/ APP_high	0.2746	0.4034	0.2411	0.3104	0.2087
APP_low/ GFP_high	0.2456	0.1507	0.2347	0.0444	0.4266
GFP_low/ GFP_high	0.2801	0.1865	0.2696	0.2703	0.2006
APP_high/ GFP_low	0.3069	0.4410	0.2764	0.4332	-0.2514

Fig. 4. τ distributions for (A) all GFP vs. APP images, and (B) for the 031506a-g image pair

We examined the representative objects from the low τ cluster, middle τ cluster, and high τ cluster for both APP and GFP images of the 031506a-g pair. The time plots of object average Intensity are shown in Fig. 5 along with images of the objects themselves (insets). The images of the representative objects are also shown. 5 (A) shows that the representative lower τ object of 031506g (GFP) image has very steep intensity dropping at early time course and then exhibits relative flat intensity change over time. This sub-population is unique as compared to the representative lower τ object of 031506a (APP) image as shown in Fig.5 (B). We believe this sub-population warrants further investigation for potential new discovery.
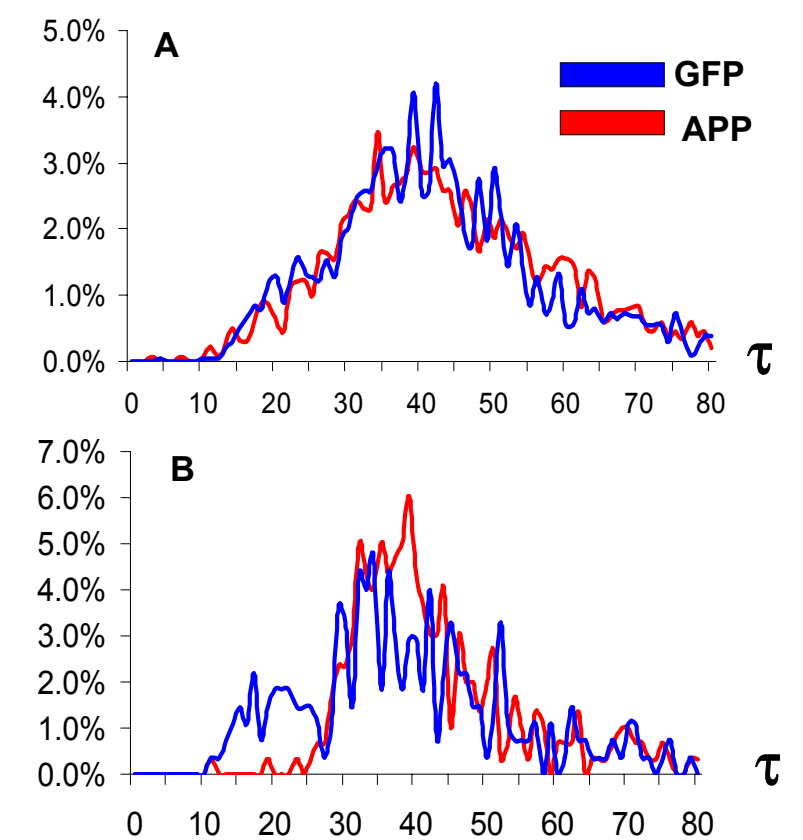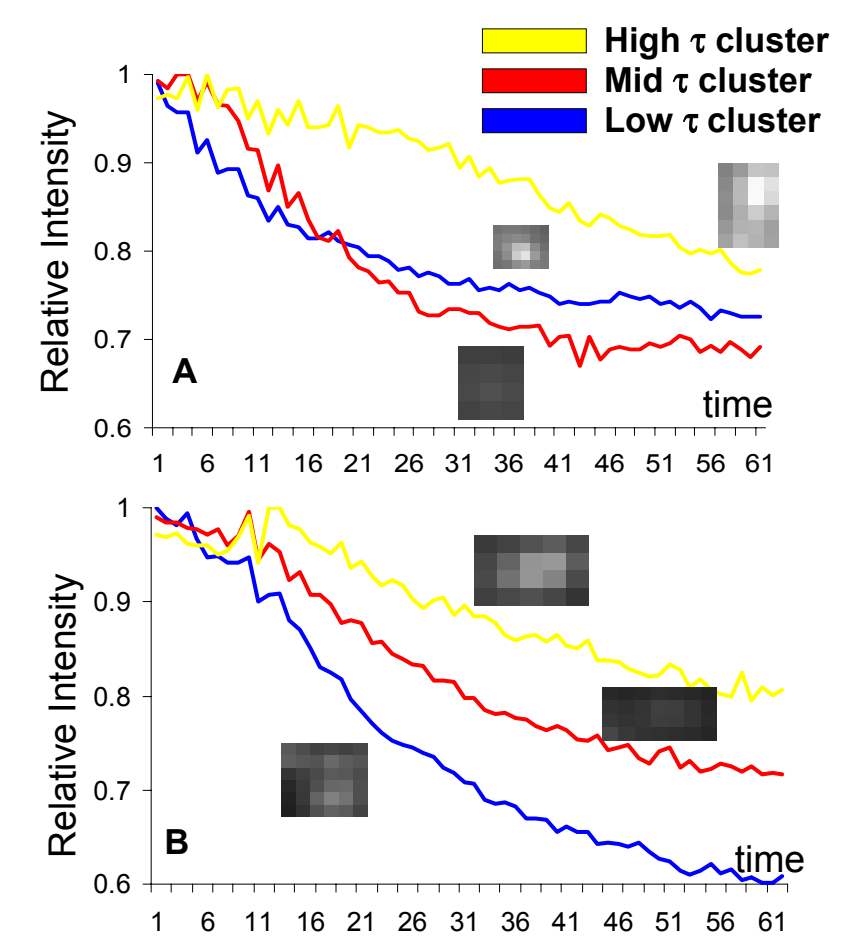

Fig. 5. Representative time plots for individual objects from different clusters of the (A) GFP and (B) APP experimental conditions

Conclusion

We applied our Discovery Framework to simulated data sets with mixed subpopulations, as well as a real image study comparing neuron cells expressing APP-GFP and GFP alone. We have shown that it could uncover latent sub-populations within experimental groups, and then provide an estimate of assay signal between the groups' sub-populations. This enables valuable insight into the experimental result which would otherwise be inconclusive. The framework could be a useful tool for exploring latent populations in a large body of image data. We expect that it is a particularly useful tool for basic research and phenotype discovery applications. Combined with our robust image and data analysis methods for the synaptic vesicle recycling assay, we believe that it could be useful for studying spatial temporal phenotypes using new approaches such as synaptoPHluorin and fluorescence quenchers. It provides a new tool for gaining insights into the mechanisms of endo/exocytosis, such as classical full collapse fusion or nonclassical "kiss and run" fusion retrieval. We will continue to characterize this tool for incorporation into our SVCell™ Microscopy Image Recognition Software platform.

1. **Alworth SV, Kelley BG, Oh S, Martin G, Sullivan M, Lee JSJ.** 2005. Robust analysis of subcellular, time-lapse microscopy assays. Poster presented at the IBC Life Sciences' Assays and Cellular Targets conference, Bellevue WA. October.
2. **Alworth SV, Cho H, Oh S, Cheng Y, Kelley BG, Martin G, Sullivan JM, Lee JSJ.** 2006. Robust modeling for the automated analysis of synaptic vesicle recycling assays. Poster presented at the American Society for Neurochemistry Conference. Portland, OR. March.
3. **Carpenter AE, Sabatini DM.** 2004. Systematic genome-wide screens of gene function. Nat Rev Genet. Jan;5(1):11-22.
4. **Zhang JH et al.** 1999. Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. J of Biom Scr 4(2):67-73
5. **Roquemore et al.** 2003. Is Z' factor the best assessment for the quality of cellular assays delivering higher content, poster of Society for Biomolecular Screening Conference, Portland, USA